COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 113 (2014) 383–395

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A usability evaluation of medical software at an expert conference setting

*Raymond Robert Bond**, *Dewar D. Finlay, Chris D. Nugent, George Moore, Daniel Guldenring*

University of Ulster, Northern Ireland, UK

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form
2 October 2013
Accepted 5 October 2013

Keywords:
Usability
Interfaces
ECG
Medical software
Interaction design
HCI

ABSTRACT

Introduction: A usability test was employed to evaluate two medical software applications at an expert conference setting. One software application is a medical diagnostic tool (electrocardiogram [ECG] viewer) and the other is a medical research tool (electrode misplacement simulator [EMS]). These novel applications have yet to be adopted by the healthcare domain, thus, (1) we wanted to determine the potential user acceptance of these applications and (2) we wanted to determine the feasibility of evaluating medical diagnostic and medical research software at a conference setting as opposed to the conventional laboratory setting.
Methods: The medical diagnostic tool (ECG viewer) was evaluated using seven delegates and the medical research tool (EMS) was evaluated using 17 delegates that were recruited at the 2010 International Conference on Computing in Cardiology. Each delegate/participant was required to use the software and undertake a set of predefined tasks during the session breaks at the conference. User interactions with the software were recorded using screen-recording software. The 'think-aloud' protocol was also used to elicit verbal feedback from the participants whilst they attempted the pre-defined tasks. Before and after each session, participants completed a pre-test and a post-test questionnaire respectively.
Results: The average duration of a usability session at the conference was 34.69 min (SD = 10.28). However, taking into account that 10 min was dedicated to the pre-test and post-test questionnaires, the average time dedication to user interaction of the medical software was 24.69 min (SD = 10.28). Given we have shown that usability data can be collected at conferences, this paper details the advantages of conference-based usability studies over the laboratory-based approach. For example, given delegates gather at one geographical location, a conference-based usability evaluation facilitates recruitment of a convenient sample of international subject experts. This would otherwise be very expensive to arrange. A conference-based approach also allows for data to be collected over a few days as opposed to months by avoiding administration duties normally involved in laboratory based approach, e.g. mailing invitation letters as part of a recruitment campaign.

Following analysis of the user video recordings, 41 (previously unknown) use errors were identified in the advanced ECG viewer and 29 were identified in the EMS application. All use errors were given a consensus severity rating from two independent usability experts. Out of a rating scale of 4 (where 1 = cosmetic and 4 = critical), the average severity rating for

* Corresponding author. Tel.: +44 7762768678.
E-mail address: rb.bond@ulster.ac.uk (R.R. Bond).

http://dx.doi.org/10.1016/j.cmpb.2013.10.006

the ECG viewer was 2.24 (SD = 1.09) and the average severity rating for the EMS application was 2.34 (SD = 0.97). We were also able to extract task completion rates and times from the video recordings to determine the effectiveness of the software applications. For example, six out of seven tasks were completed by all participants when using both applications. This statistic alone suggests both applications already have a high degree of usability. As well as extracting data from the video recordings, we were also able to extract data from the questionnaires. Using a semantic differential scale (where 1 = poor and 5 = excellent), delegates highly rated the 'responsiveness', 'usefulness', 'learnability' and the 'look and feel' of both applications.

Conclusion: This study has shown the potential user acceptance and user-friendliness of the novel EMS and the ECG viewer applications within the healthcare domain. It has also shown that both medical diagnostic software and medical research software can be evaluated for their usability at an expert conference setting. The primary advantage of a conference-based usability evaluation over a laboratory-based evaluation is the high concentration of experts at one location, which is convenient, less time consuming and less expensive.

1. Introduction

In the US alone, iatrogenic or medical errors are responsible for 100,000 deaths and 1,000,000 injuries every year [1]. The total cost of all medical errors in the US has accumulated to approximately $50 billion. As a result, the 'science of error prevention' has been deemed a new, yet important topic in medicine [2]. Within this topic, the 'usability' of medical devices has recently become a growing interest [3]. Usability as defined by Nielsen [4] is "a quality attribute that assesses how easy user interfaces are to use".

Although all medical errors are not caused by badly designed devices, many errors are. These errors, however, may be reduced by evaluating the usability of medical devices prior to their commercial availability [5]. A usability evaluation has been coined a 'usability engineering process' [6]. This process can be used to uncover foreseeable misuse of the device, also known as the 'use errors' before they are used to detrimental effect in real-life clinical practice. Once these foreseeable 'use errors' have been identified, these can be 'designed-out' of the system and usability can be 'designed-into' the system. As defined by Nielsen [4], an error is "any action that does not accomplish the desired goal". Nielsen makes reference to the term 'use error', as opposed to the more traditional term, 'user error'. The reason being, errors committed whilst using the device may actually be due to badly designed interfaces, as opposed to an error solely induced by the user [7]. This shifts the responsibility from the user to the actual system and the manufacturer. Nevertheless, 'design-induced errors' can be found prior to the launch of the device with the aid of a usability engineering process. In recognition of this, the Food and Drug Administration (FDA) presented the Good Manufacturing Practices regulation in 1997 [8] which requires medical device manufactures to document user needs and to verify that these needs have been met. This regulation was later followed by an international standard in 2007, namely 'ISO/IEC 62366: Medical Devices-Application of Usability Engineering to Medical Devices' [6] which requires manufactures to actually carry out a usability engineering process. According to the Medical Device Directive [9], medical devices not only include the physical hardware but also the software. Therefore, a usability engineering process must be in place to legally introduce, not only medical hardware but also medical software into routine healthcare.

With the emergence of Medicine 2.0, a number of Web-based medical software applications have been developed by researchers to improve patient care, interoperability and remote access to tools and applications irrespective of the user's geographical location [10–13]. Medicine 2.0 (synonymous with Health 2.0) is the concept to avail of already established Web 2.0 concepts and technologies [10]. The authors of this paper have developed two applications that, to our knowledge, are the first of their kind [14,15]. Both applications have been developed using the Adobe Flash technology [16]. Although the applications have been developed, they have not been formally evaluated for their usability [14]. Therefore this study evaluates the usability of these applications and identifies any existing usability issues. Both applications parse and visualise Electrocardiogram (ECG) data. An ECG is a recording of the electrical activity of the heart and is one of the most commonly used diagnostic tools in clinical practice [17].

1.1. *The ECG electrode misplacement simulator*

The first software application that has been evaluated in this study is a medical research tool called the ECG electrode misplacement simulator (EMS), which can be viewed in Fig. 1. An ECG is recorded using 10 electrodes placed at precise anatomical landmarks. These landmarks have been in use for more than 70 years and were formally endorsed in 1938 [18]. It has been reported that up to 4% of ECGs are recorded using incorrect electrode positions [19]. Electrodes placed incorrectly can result in a misdiagnosis and therefore yield inappropriate medical treatment. This is a well known but ill-addressed problem and on-going studies continue to investigate the effects of electrode misplacement [19–23]. The EMS is Web-based software that allows clinical researchers to interactively misplace electrodes whilst observing the effects this has on the ECG waveforms/signals. The software can also be used as part of a training programme for ECG technicians given it has been reported that effective ECG training can reduce

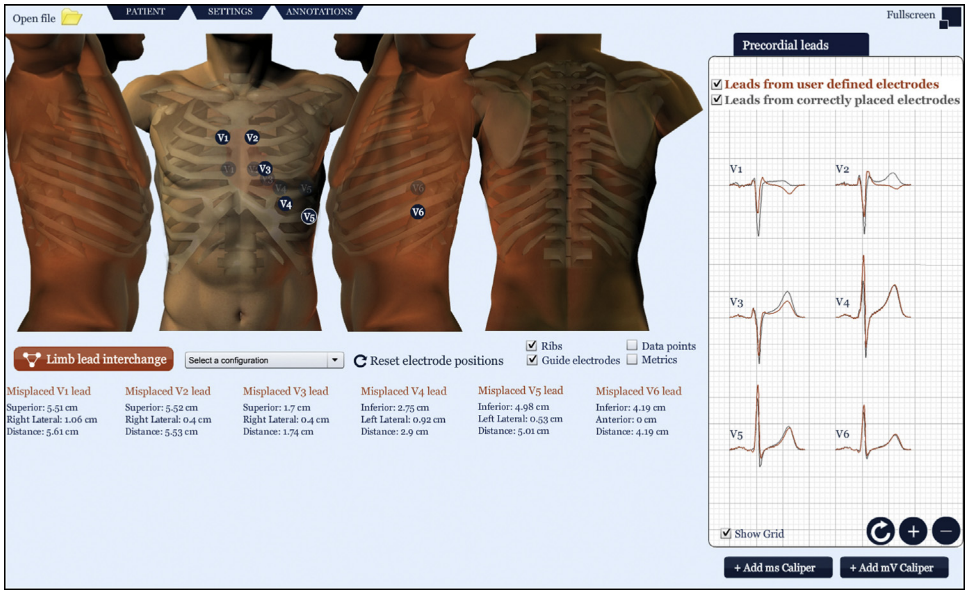


Fig. 1 – A screenshot of the electrode misplacement simulator application.

the frequency of recording errors by 75% [19]. This usability evaluation will go some way in showing the potential user acceptance and friendliness of this application. It is important that software similar to this is adopted in the future for researching the effects of electrode misplacement and for training ECG technicians. If adopted, this software may also aid clinical researchers in the discovery of criteria that could be integrated into ECG machines for the automatic detection of electrode placement errors. The EMS software is freely available on the Web [24].

1.2. *Advanced ECG viewer*

The second software application is a medical diagnostic tool. This is an advanced Web-based ECG viewer. A screen shot of this application is presented in Fig. 2. The ECG viewer not only facilitates visualisation of data recorded from the standard ECG acquisition methods (i.e. the 12-lead ECG), it facilitates visualisation of data from advanced ECG acquisition methods (i.e. the Body Surface Potential Map or BSPM). A BSPM records and displays the entire cardiac activity of the heart whereas the standard 12-lead ECG represents only a sample of the cardiac activity [25]. Although the BSPM has a higher sensitivity in detecting cardiac abnormalities when compared to the standard 12-lead ECG [25,26], it has not yet been adopted for routine clinical practice. This is partly due to the fact, that no 'clinician-friendly' advanced ECG viewer has been developed [14]. This ECG (or BSPM) viewer is the first viewer of its kind and it is important that its potential user acceptance and user friendliness is assessed through a usability study. If such a viewer is adopted for routine clinical practice, it has the potential to increase the accuracy of diagnoses in patients with cardiac related diseases. As a result, more patients will receive the appropriate medical treatment (e.g. reperfusion therapy) in a timely manner. There will therefore be a reduction in cardiac related morbidity and mortality. The viewer itself is freely available on the Web [27].

1.3. *Assessing usability*

These two medical software applications were developed by a small team of engineers, computer scientists, multimedia designers and clinicians. Although the actual design of these applications was partly influenced by appropriate feedback obtained from potential users, the authors thought it important to adopt a systematic approach that would test and improve the usability of these applications. There are, however, a number of methodologies available to systematically evaluate usability [28]. A heuristic evaluation is a usability inspection method where a small number of usability experts are employed to assess a system [4]. These experts analyse whether the system adheres to a set of predefined heuristics otherwise known as usability principles. This is a cheap method often referred to as usability discount engineering [4]. This method has, however, been successfully used to evaluate medical applications [29,30]. A higher fidelity method, yet more expensive, is the 'usability test'. A usability test is where real users are recruited to attempt a number of representative tasks using the software whilst an observer records the 'use errors' or the usability problems each user encounters. This method has already been widely used to successfully evaluate medical software and has been extensively cited as the preferred methodology [30,31]. For the reasons given, we adopted the usability test methodology in this study. A usability test is however normally carried out in a laboratory setting. And given the two software applications are intended to be used by expert clinicians and researchers, we have taken the novel approach of evaluating the applications in an expert conference setting. We also wanted to determine the feasibility of evaluating both medical diagnostic and medical research software at a conference setting as opposed to the conventional laboratory setting. This is due to the fact that a conference-based setting offers some obvious benefits (high concentration of experts at one location, which is both convenient and economic).

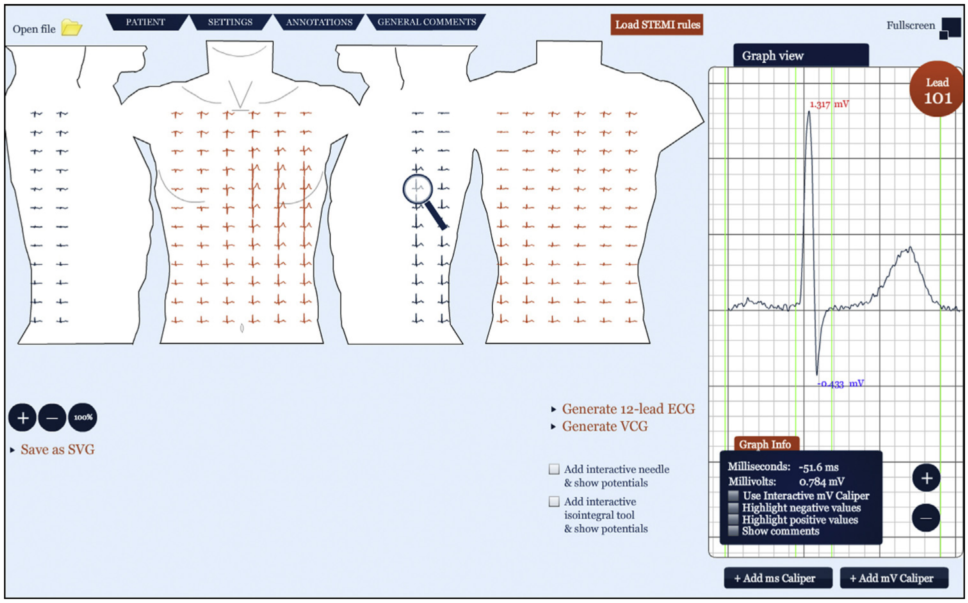


Fig. 2 – A screenshot of the advanced ECG viewer application.

2. Methods

2.1. *Conference-based usability test*

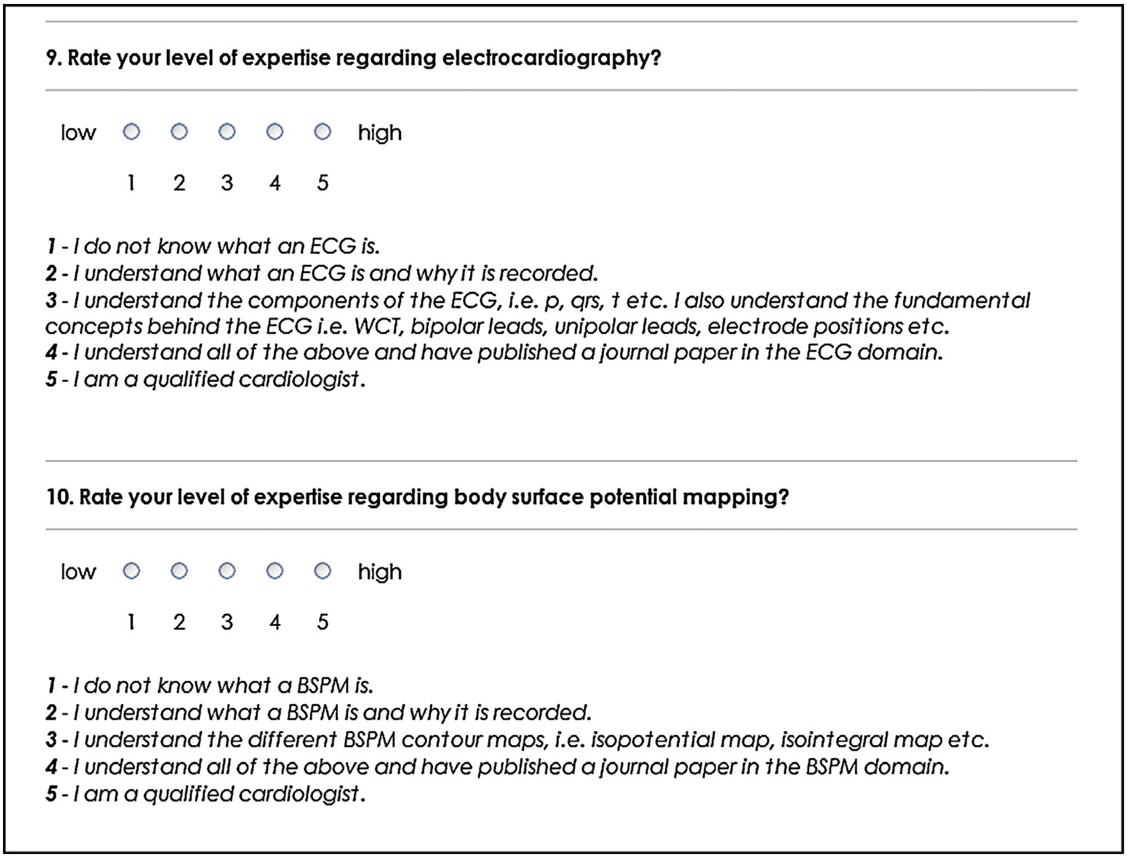
The Annual International Conference on Computing in Cardiology [32] plays a significant role in presenting research on the topic of computerised electrocardiography and as a result, attracts world leading experts in this domain. This conference was thus seen as a suitable venue for engaging with expert users.

2.2. *Recruiting delegates/participants*

An information sheet that invited participation in the usability evaluation was inserted into the delegate packs and distributed to all delegates ($n \approx 300$) at the 2010 International Conference on Computing in Cardiology. The information sheet requested those willing to participate to contact a member of the local organising committee. From this, 17 delegates either responded or were solicited to take part in the evaluation of the medical research tool (EMS) and seven delegates participated in the evaluation of the diagnostic tool (ECG viewer). Given usability specialists recommend five participants to achieve the maximum cost/benefit ratio [33], the number of participants recruited to evaluate each of the applications in this study was deemed to be sufficient. It has been reported that just five participants can identify approximately 80% of all usability problems in a software application [34]. To avoid potential bias, incentives for taking part in the study were not offered. For example, if there is a monetary incentive, participants may feel obliged to give a positive review of the software.

Table 1 – Demographics of all participants involved in the evaluation of the EMS application and the ECG viewer application.

	Participants involved in the evaluation of the EMS application ($n = 17$)	Participants involved in the evaluation of the ECG viewer application ($n = 7$)
Gender	16 males, 1 female	7 males
Age	Mean 42 ± 12 years of age	Mean 39 ± 11 years of age
Occupation	15 academics, 1 industry. 1 participant is based in both industry and academia	6 academics, 1 participant is based in both industry and academia
Primary expertise	8 electrical engineers, 2 computer scientists, 6 medical doctors and 1 medical physicist	2 electrical engineers, 2 computer scientists, 2 medical doctors and 1 medical physicist
First language:	9 English, 3 Dutch, 2 German, 2 Spanish, 1 Hungarian	4 English, 2 German, 1 Dutch
Average level of computer literacy	Mean 4.3 ± 0.8 Mode: 5	Mean 4.7 ± 0.5 Mode: 5
Average computer usage per week	Mean 28 ± 11 h	Mean 34 ± 3 h
Average Internet usage per week	Mean 17 ± 12 h	Mean 19 ± 13 h
Average level of ECG expertise	Mean 4.12 ± 0.6 Mode: 4	Mean 4 ± 0.4 Mode: 4
Average level of BSPM expertise	Mean 2.88 ± 1.1 Mode: 2	Mean 3.6 ± 1 Mode: 4



9. Rate your level of expertise regarding electrocardiography?

low ○ ○ ○ ○ ○ high

1 2 3 4 5

1 - I do not know what an ECG is.
2 - I understand what an ECG is and why it is recorded.
3 - I understand the components of the ECG, i.e. p, qrs, t etc. I also understand the fundamental concepts behind the ECG i.e. WCT, bipolar leads, unipolar leads, electrode positions etc.
4 - I understand all of the above and have published a journal paper in the ECG domain.
5 - I am a qualified cardiologist.

10. Rate your level of expertise regarding body surface potential mapping?

low ○ ○ ○ ○ ○ high

1 2 3 4 5

1 - I do not know what a BSPM is.
2 - I understand what a BSPM is and why it is recorded.
3 - I understand the different BSPM contour maps, i.e. isopotential map, isointegral map etc.
4 - I understand all of the above and have published a journal paper in the BSPM domain.
5 - I am a qualified cardiologist.

Fig. 3 – Excerpt of the Web-based pre-test questionnaire. The two questions shown here were used to obtain participant expertise of the ECG and the BSPM.

2.3. Pre-test questionnaire

Prior to the usability testing, all participants signed a voluntary consent form allowing for screen and audio recordings. Each participant completed a pre-test questionnaire (available online [35]). This questionnaire consists of ten questions that were used to profile the participants (i.e. age, gender, first language, occupation, primary expertise and their computer literacy). All demographics obtained from this questionnaire has been summarised in Table 1.

One particular demographic of interest is information regarding participant expertise of the ECG and the BSPM. The expertise of the participant was systematically obtained from two questions that incorporated a five-point 'semantic differential scale'. The semantic differential scale used throughout the questionnaire attributed rating one to be the negative connotation and rating five to be the positive connotation. In this case, the expertise scale rated from one meaning 'low' expertise to five meaning 'high' expertise. The five options on the scale included a description that was used to aid participants and to improve the validity of the answers. These descriptions can be found in Fig. 3 and participant expertise can be found in Table 2. The reason delegates were more suitable to evaluate the EMS is that there were fewer advanced ECG (i.e. BSPM) experts when compared to the number of experts on the standard 12-lead ECG. Although this conference attracts international ECG researchers, the lack of BSPM experts at level 5 that were recruited may indicate the diminishing number of researchers in this domain.

2.4. Representative tasks

Subsequent to completing the pre-test questionnaire, each participant was asked to attempt representative tasks using the software. A task is considered 'representative' when the task is fundamental to the system and is a common task in the system's normal operational use. Like any representative task, these tasks (Table 3), were determined by a consensus from the developer and potential users. It was considered important to ensure the actual wording of the tasks did not provide clues

Table 2 – The ECG and BSPM expertise levels from all participants as obtained from the pre-test questionnaire. This knowledge was used to infer which participants were suited to evaluate which application or indeed both applications.

Rating	General ECG expertise (%)	BSPM expertise
1	0	6
2	0	35
3	12	29
4	65	24
5	23	6

Table 3 – Details of the tasks attempted by participants in both ECG applications respectively.

Tasks	Tasks associated with the EMS application ($n = 6$)	Tasks associated with the ECG viewer application ($n = 7$)
Small task 1	Find a textual description of the patient's diagnosis	Find the patient's name
Small task 2	Hide the ribs used to depict the electrode positions	Calculate the width of the QRS?
Small task 3	Hide the data points	Calculate the width and height of the T wave in lead 75?
Elaborate task 1	Use the software to visualise the difference in waveform morphology in V1 and V2 if they were placed in the third intercostal space as opposed to the fourth	Find the lead with the highest R wave
Elaborate task 2	Use this software to find out which limb leads are not affected when the left arm and right arm electrodes are switched	You suspect a myocardial infarction; use the viewer to find which lead has the highest elevation at ST60 (J point + 60 milliseconds)
Elaborate task 3	For a scientific study you want to find the QRST similarity coefficient of leads V3 and V4 when they are switched	Visualise the entire ST segment and export this image to a file to send in an email
Elaborate task 4	None was given	You notice that the T wave in Lead 78 is flat; leave a comment over the T wave for a colleague to read and verify

as to which tools were required to complete the task. The participants involved in evaluating the EMS software were given three small tasks and three more demanding tasks. As for the participants involved in evaluating the ECG viewer, these participants were given three small tasks and four elaborate tasks. The labels 'small' and 'elaborate' were associated with the tasks as opposed to the labels 'simple' and 'difficult'. The reason being, participants may find the supposedly 'simple' tasks difficult and the supposedly 'difficult' tasks simple. In this respect, it is important participants are not made to feel incapable; if in the case they cannot complete a so-called 'simple task'. All elaborate tasks were printed in a large font size on A4 paper and placed in front of the participant. Each task was also read out to the participant. It was printed because it is important to allow participants to re-read the task, especially if the task involves multiple steps.

2.5. *Hardware and software requirements*

Relatively inexpensive apparatus (i.e. desktop computer, digital microphone and screen recording software) were used to evaluate the usability of the two medical software applications. The computer was placed in a private area to avoid pre-conditioning other potential participants and to avoid the subject being distracted. This was reinforced by a 'please do not disturb' sign indicating that 'a usability test is in progress'. Whilst the participant attempted each of the tasks, a digital microphone and screen recording software were used to record verbal feedback and user interactions with the software. The Apple Quicktime version X [36] was used to record the participant's interactions on the screen. There are a number of benefits for recording the screen and the audio:

- The observer does not need to take notes.
- And therefore note taking can no longer distract the user.
- The researcher does not need to rely on written notes or his/her memory.
- The observer can spend more time watching and concentrating on the user's interactions.

2.6. *Think-aloud protocol*

It was important to record the participant's voice because they were asked to 'think-aloud' whilst attempting the tasks. As recommended by Jaspers [28], all participants were given time to practice thinking aloud before attempting any of the elaborate tasks. The exact method used in this study has been called the 'concurrent think-aloud protocol' as opposed to the alternative 'retrospective think-aloud protocol' where the participant is asked to think-aloud after the task has been completed. It has been reported, however, that the 'concurrent' approach detects more use errors when compared to the retrospective approach [28]. The concurrent think-aloud protocol has been successfully used in the past [30,31,37] in the laboratory setting as an effective methodology for highlighting usability problems and for obtaining qualitative feedback about the software in question. The 'think-aloud' protocol is an effective tool that helps gain an insight into the user's working memory and to the way they actually solve problems. It originated in the field of cognitive psychology as a tool to elicit verbal reports from human subjects [28]. It was later used by Lewis [38] in 1982 for evaluating computer interfaces.

2.7. *Post-test questionnaire*

After each participant attempted all tasks, they then completed a post-test questionnaire. This questionnaire was developed for presentation on the Web. There are a number of advantages for using Web-based questionnaires as opposed to paper-based questionnaires. These include the fact that typed answers avoid ambiguous hand writing, questions can be designed in a way to force an answer, the answers themselves can be automatically conditioned for direct use with statistical software and the fact Web-based questionnaires are environmentally friendly. It was, however, primarily Web-based because it kept the participant in close proximity to the digital microphone. As a result, additional verbal feedback from the participant was recorded whilst they completed the questionnaire. Post-test questionnaires used to evaluate

Table 4 – Advantages and disadvantages of conducting a usability test in an expert conference setting.

Advantages	Disadvantages
• High concentration of subject experts at the same geographical location	• Conferences have busy schedules
Effective advertisement of the usability study, i.e. access to delegate packs	• Limited amount of time with each participant. Session breaks last no more than 30 min to 1 h
• Potential participants can be better informed (face-to-face queries)	• Only a small number of tasks can be attempted due to this time constraint
• A conference facilitates recruitment of representative users for evaluating research-orientated applications	• A constraint on time and the number of tasks may affect the overall conclusion on the software's usability
• International representation of potential users	• May be difficult to find a private (noise-free) area to carry out the study
• Unique expert feedback can be obtained from a specialised conference	• A conference setting may involve recruiting a convenient sample of subjects who are inclined to experiment and may not be representative of all practitioners
• Data from participants can be collected over a period of a few days as opposed to weeks or months	
• The location and setting is more convenient for participants. They can attend the conference as planned and take part in the study all at the same location.	
• Less expensive in terms of human resources and no additional travel costs for participants.	

each of the medical software applications can be found online [39,40].

2.8. *The script*

The entire usability engineering process in this study was carried out using a carefully crafted script. A script was used to ensure participants were informed in a systematic fashion. For example, when using the think-aloud protocol as a method to assess usability, it is important to convey this very concept to users before they interact with the software. Therefore, at the start of each session, participants were asked to verbalise their thoughts, i.e. what they were thinking, what they were trying to accomplish and what they were going to do next. The researcher on site was also scripted to say that, "we are testing the software and not you (the participant)". Although this statement is true, it also helps relax the participant. The script also included a series of milestones to remind the researcher when to carry out important administrative duties, e.g. when to hand out questionnaires and when to start the audio and screen recording software etc.

3. Results

As a result of this study, a number of advantages and disadvantages for carrying out a usability evaluation of medical software at an expert conference setting as opposed to a laboratory setting have been identified and presented in Table 4.

One of the main advantages of a conference-based usability evaluation compared to a laboratory-based approach is the high concentration of experts at one geographical location. In addition, these experts also represent the international stage and under normal circumstances it would be very expensive to arrange international experts to partake in a face-to-face usability evaluation of a specialised medical software application. Therefore, the conference-based approach facilitates recruitment of a convenient sample of international subject experts. Furthermore, given the high concentration of experts, a usability evaluation can be carried out over a few days as opposed to weeks or months. This is contrary to a laboratory-based usability evaluation where weeks or sometimes months are dedicated to mailing invitation letters and sending endless emails as part of the initial recruitment campaign. A conference-based usability evaluation also allows effective advertisement of the study through dissemination of information sheets in the delegate packs. It also allows for face-to-face discussion regarding any queries a potential subject may have, Moreover, although subjects can volunteer to participate in the study, a conference setting also facilitates the active solicitation of potential subjects to partake in a usability evaluation. One of the most pertinent advantages is the fact that a conference-based approach does not require subjects to take additional time and money for travel.

The authors believe the advantages out-weigh any disadvantages associated with undertaking a usability test in a conference setting. The most significant disadvantage is that less time is spent with each participant given the busy schedule of a typical conference. Within this study, the average duration of a usability session was 34.69 min (SD = 10.28). However, taking into account that 10 min was dedicated to the pre-test and post-test questionnaires, the average time dedicated to actual user interaction of the medical software was 24.69 min (SD = 10.28). Given session breaks at conferences are usually 30 min to 1 h, this data shows the feasibility of integrating usability studies at a conference setting. We did find that four tasks performed within a 30-min time span is about the maximum number of tasks that could be given at conference setting, however we did find that this was adequate for evaluating either diagnostic software or medical research software.

From inspection of the video recordings use errors were transcribed into a structured table. The frequency of use errors

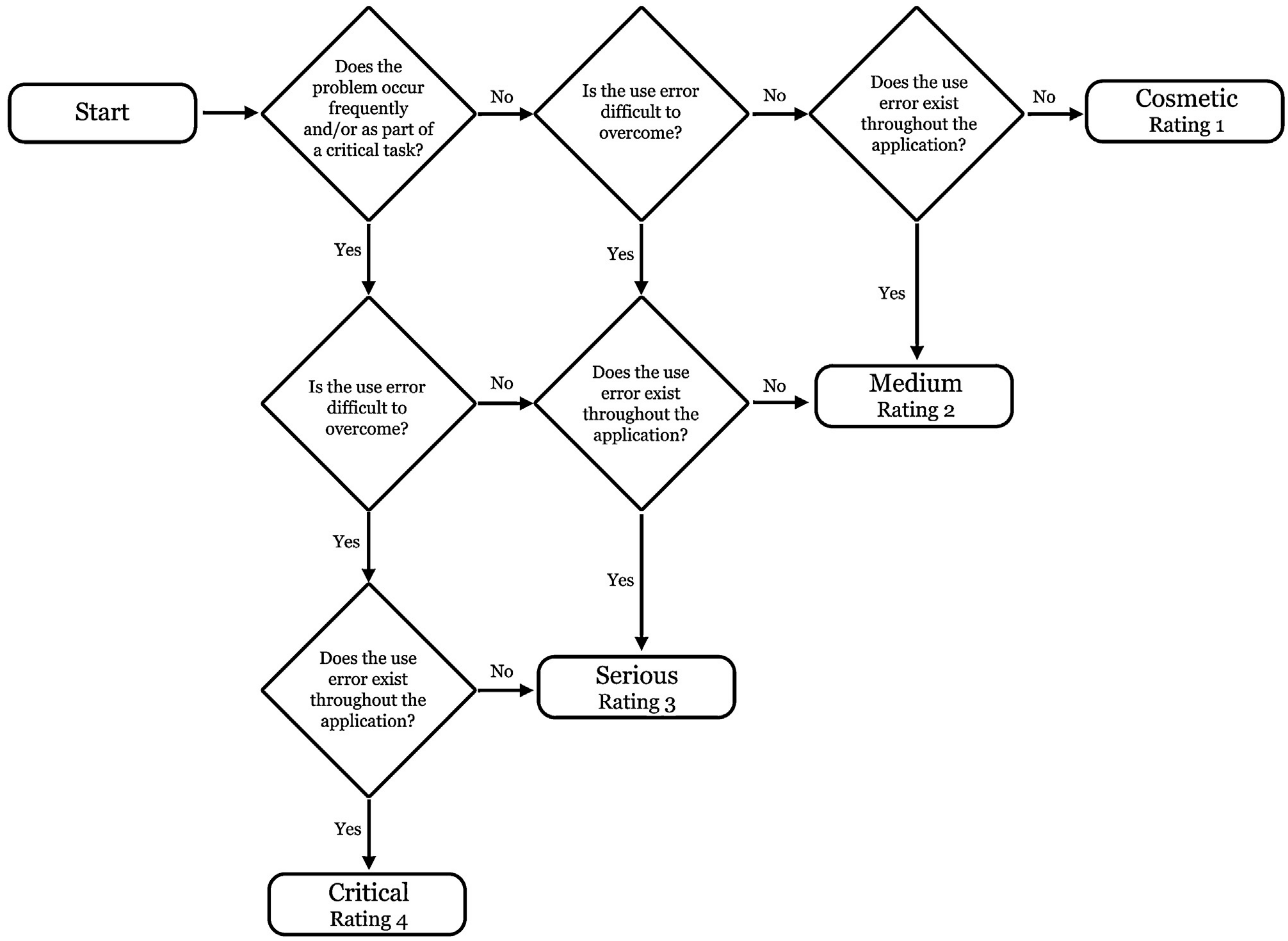


Fig. 4 – This flowchart diagram was used to systematically assign a severity rating to each of the use errors found in both medical applications. The design of the flowchart was influenced by a combination of two existing severity classification methods [4,41].

was also recorded during this process. Each use error was classified by severity using criteria specified in Fig. 4. A use error can have one of four severity ratings, i.e. cosmetic, medium, serious or critical. To improve the accuracy of this process, a severity rating was determined by two independent usability experts. Amongst the use errors detected in the ECG viewer, 78% were given the same severity rating by both experts. Moreover, from the use errors detected in the EMS application, 59% were given the same severity rating by both experts. This illustrates the subjectivity of a severity rating process. The remaining disputed severity ratings were re-evaluated by both usability experts until a consensus was made. It is important to note that rating the severity rating of use errors is a vital process. It can be used to prioritise use errors that require urgent attention from those that are merely cosmetic.

Based on this usability study at a conference setting, 41 use errors (average severity = 2.24 [SD = 1.09]) were discovered in the ECG viewer and 29 use errors (average severity = 2.34 [SD = 0.97]) were discovered in the EMS application. The severity of these use errors are presented in Fig. 5.

An example set of use errors that were identified in the EMS application can be found in Table 5. This table exhibits a description of the use error, the number of times the use error occurred (frequency), the severity rating and a recommended solution. It also displays the usability principle violated. These usability principles were articulated by Nielsen [4] and have been used by a large number of researchers [28] as heuristics for evaluating the usability of computer interfaces.

Alongside the discovery of unknown use errors, we also found that it was possible to evaluate the efficiency, usefulness, learnability and user satisfaction of using the medical software applications [4].

3.1. *Efficiency*

Efficiency can be measured by evaluating the task completion rates and task completion times. These can be extracted from time-stamps in the actual video recordings. The completion times of the elaborate tasks for both applications can be viewed in Tables 6 and 7. The completion rates for the first two elaborate tasks carried out in the EMS were 100% indicating that all users completed these tasks without any assistance. The third elaborate task did, however, have a completion rate of 82%. The reason for this is that the task involved identifying abbreviated information in the interface and as a result, users found this information difficult to locate (i.e. interface showed

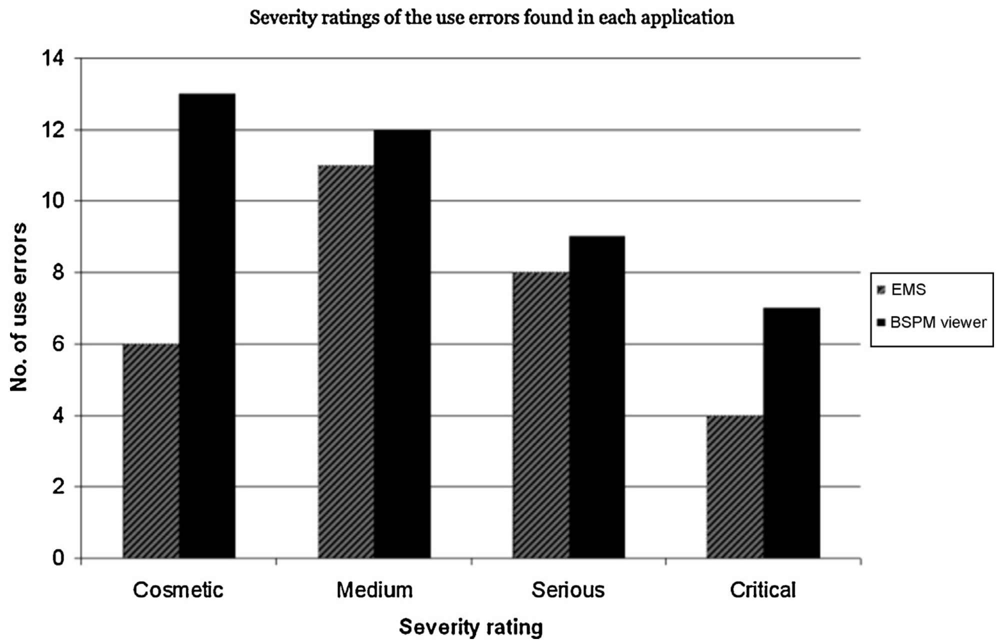


Fig. 5 – The severity classification of the use errors identified in both ECG applications, respectively.

Table 5 – A sample of use errors identified in the EMS application.

Use error	Severity rating	Usability principle violated	Frequency	Recommended solution
User assumed the orange leads to be the correctly placed leads	4 – critical	1 – visibility of system status	Identified by 6 participants	Change the key to 'user defined leads' as opposed to 'misplaced leads'
User tried to drag the limb electrodes since the precordial electrodes can be dragged	3 – serious	4 – consistency and standards	Identified by 13 participants	Implement a limb electrode drag feature
When dragging the precordial electrodes, they are not brought to the front of the interface, i.e. you can drag electrodes under other electrodes	2 – medium	4 – consistency and standards	Identified by 2 participants	Change depth of the electrode when the left button of the mouse is pressed
User tried to zoom into a lead by double clicking	1 – cosmetic	3 – user control and freedom	Identified by 1 participant	Add this zoom feature

'CC' instead of 'Correlation Coefficient' or '*r*'). The completion times of the tasks carried out in the EMS application was considered satisfactory. However, it is important to point out that like most statistics, those derived from a usability test are open for subjective interpretation.

With regards to the evaluation of the ECG viewer, all four elaborate tasks were completed by all participants without any assistance. However, the average completion time for task one is excessive. This excessive completion time was due to a lack of 'visual hierarchy', i.e. the tool required to complete the task had no visual prominence. Elaborate tasks three and four were considered satisfactory.

From the post-test questionnaire, participants rated the 'responsiveness' of the EMS application between four and five

Table 6 – The completion rates and completion times of the elaborate tasks carried out using the EMS application.

Elaborate task	Completion rate (%)	Average completion time (min)
1	100	1.12
2	100	1.08
3	82	1.12

Table 7 – The completion rates and completion times of the elaborate tasks carried out using the ECG viewer application.

Elaborate task	Completion rate (%)	Average completion time (min)
1	100	2.59
2	100	0.54
3	100	1.10
4	100	1.35

on a five-point semantic differential scale (mode: 5, mean: 4.64). They also rated the 'responsiveness' of the ECG viewer between three and five (mode: 4/5, mean: 4.28). When combining these metrics, the efficiency of these two medical applications was assessed from a conference-based usability evaluation.

3.2. *Usefulness*

The 'usefulness' of both applications was evaluated qualitatively using the post-test questionnaire. Participants rated the 'usefulness' of the EMS (mode: 5, mean: 4.52) and the ECG viewer (mode: 4, mean: 4) in the clinical ECG domain. The usefulness was further assessed by a number of polar questions, which have been summarised in Figs. 6 and 7 for both medical software applications respectively.

3.3. *Learnability*

Learnability was also assessed by a number of questions in the post-test questionnaire. Participants were asked to rate how easy they thought the applications were to learn. Participants rated the 'learnability' of the EMS (mode: 4, mean: 4.17) and the ECG viewer (mode: 3/4, mean: 3.7). The learnability was also assessed from several polar questions, i.e. 94% of participants said they could remember how to use the EMS application if viewed for a second time, 88% said that they could learn the application without formal training and that they could learn all of the features within a period of hours. Regarding the ECG viewer, 71% said they could remember how to use the software if seen for a second time, 57% indicated that they could learn the application without formal training and 71% said they could learn all of the features in the ECG viewer within a period of hours. The investigators found this information useful, however, to quantify and assess the 'learnability' of software at a high level, a longitudinal study would be required to assess the actual 'learning curve'.

3.4. *User satisfaction and aesthetics*

User satisfaction and views on the aesthetics of the software applications was assessed in the post-test questionnaire. Regardless of its subjective nature, degrees of user satisfaction has been associated with 'high control' of a software application [4]. Participants were asked to rate the 'look and feel' of both applications. Delegates rated the 'look and feel' of the EMS application (mode:5, mean:4.29) and also the 'look and feel' of the ECG viewer (mode: 5, mean: 3.85). 76% said the colour scheme of the EMS was appropriate and 71% also said that the layout was intuitive. Similarly, 57% of the relevant participants said the colour scheme of the ECG viewer was appropriate and 86% said the layout of the ECG viewer was intuitive. The reason why the colour scheme of the ECG viewer application did not receive significant user satisfaction was due to inadequate contrast, i.e. white text was presented on top of a pale yellow background (such text is difficult to read).

4. Discussion

4.1. *Medicine 2.0*

Traditional healthcare applications are restricted for use at specific geographical locations; however, with the advent of Web 2.0, and the emergence of Medicine 2.0, some healthcare applications can now be accessed anywhere in the world. In this study, the post-test questionnaire asked participants whether they preferred the desktop or the Web as a platform for using medical software. Under normal circumstances, this question may not be fully appreciated by all healthcare professionals, however, it was asked after the participant had interacted with a Web-based application. The questionnaire itself also provided a clear description of the Web as a platform. As a result, 76% of participants said they prefer the Web as a platform to access medical software as opposed to the

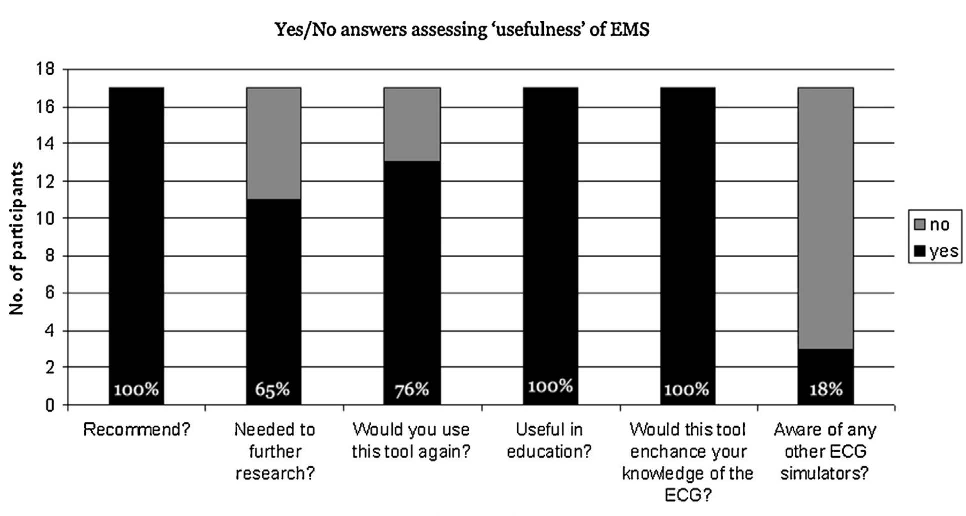


Fig. 6 – Results from a series of polar questions from the post-test questionnaire, which assessed the usefulness of the EMS application. Percentages represent the group of participants that answered yes.

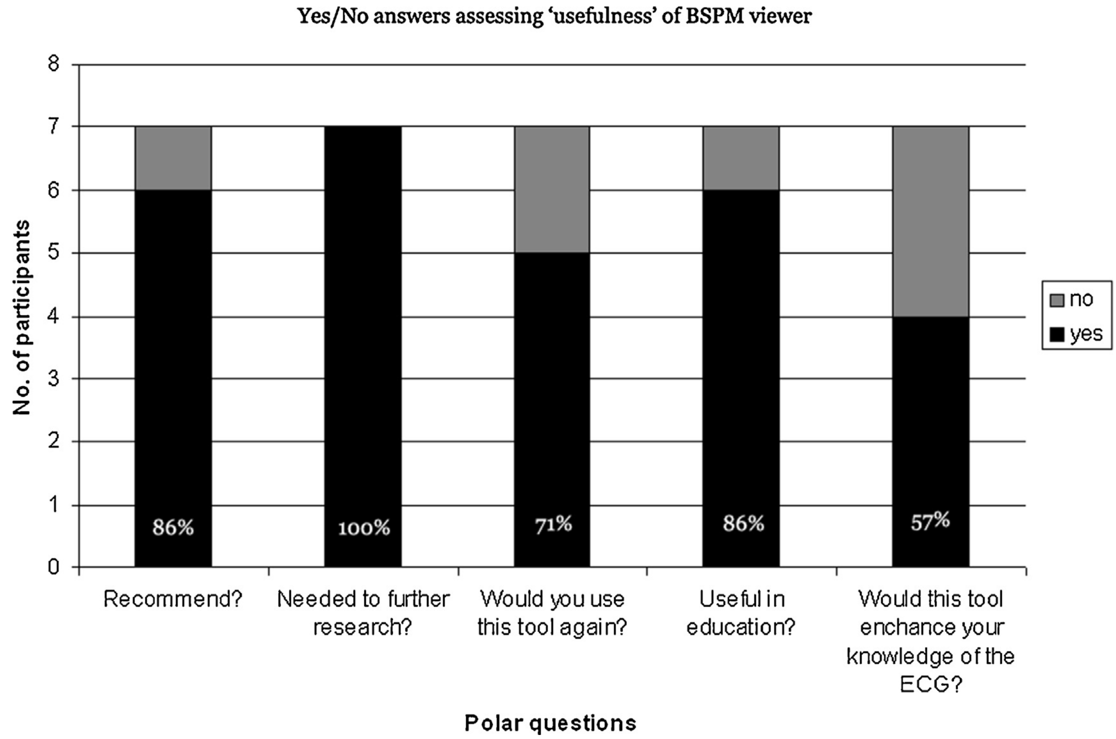


Fig. 7 – Results from a series of polar questions from the post-test questionnaire, which assessed the usefulness of the ECG (BSPM) viewer application. The percentages represent the group of participants that answered yes.

desktop. This indicates some acceptance and recognition of the Medicine 2.0 initiative amongst healthcare experts. However, users gave many valid reasons as to why Internet based software applications are not preferable, e.g. "with a desktop application I can almost guarantee its availability – its not subject to Internet access like most Web applications".

4.2. *Profiling participants*

The subjects within this study were profiled as 'expert ECG researchers'. ECG researchers do, however, normally have their own primary specialisation, for example, 'engineering' or 'medicine'. With this information obtained from the pre-test questionnaire, we calculated the average number of errors (mean = 5, SD = 3.7) identified through delegates that were primarily engineers and the average number of errors (mean = 3, SD = 2) identified through delegates who were medical professionals. Although slightly more usability problems were identified through users with an engineering background, an unpaired two-sample *t*-test test does not indicate any statistical significance (p-value = 0.310) to support the hypothesis that medical professionals encountered more errors when compared to engineers or vice versa. This may indicate that overly strict user profiles do not necessarily result in the detection of more 'use errors'. This would subscribe to the idea of "recruiting loosely" as pointed out by a usability expert, Krug [34]. This means that to assess the usability of an interface, participants do not necessarily need to meet a detailed set of criteria.

5. Conclusion

A conference-based usability engineering methodology has been used to evaluate the usability of two medical software applications. One application is a medical diagnostic tool (ECG viewer) and the other is a medical research tool (EMS). Both applications are novel and have only been recently introduced into the research domain through journal publications and conference proceedings. The ECG viewer is a potential diagnostic tool and the EMS is a potential medical research and training tool. However, given these applications have yet to be adopted by the healthcare domain, the aforementioned usability and survey results go some way to demonstrating the potential user acceptance and user-friendliness of both applications.

The usability evaluation process involved a 'usability test' that to our knowledge is the first to be performed at an expert conference setting as opposed to a conventional laboratory. It is important to present these novel usability engineering methodologies given a perfected usability engineering protocol has yet to be invented. As of 2010, the National Institute for Standards and Technology (NIST) in the US are looking to develop a standard framework for assessing the usability of health IT systems but have yet to confer a recommended protocol [42]. This study has shown that both medical diagnostic software and medical research software can be evaluated to an extent for their usability at a conference setting. However, a similar study would need to be carried out to investigate the feasibility of assessing physical 'medical devices' at a conference setting (as opposed to medical software). Given medical device companies would be interested in this kind of study, the authors plan to carry this out. Furthermore, to fully validate and endorse this conference-based approach, a controlled trial would also need to be executed to directly compare and contrast this approach to the laboratory approach. As such, the authors also plan to design an effective trial within this context. We have also identified some limitations with the 'usability test' methodology. One limitation is regarding the subjective nature of interpreting completion times. Therefore for future studies we aim to determine baseline completion times as determined from expert users. This would facilitate a comparative analysis. Another limitation is that a conference-based approach by itself is not sufficient to gain regulatory approval and thus should not replace dedicated clinical validation studies and clinical trials. Finally, a conference-based approach is limited to recruiting delegates with a particular 'expert' profile. And such 'expect' profiles may not represent intended 'novice' users of a proposed medical device.

REFERENCES

[1] L.T. Kohn, J. Corrigan, M.S. Donaldson, To Err is Human: Building a Safer Health System, 1st ed., National Academies Press, Washington, DC, 2000.

[2] P. Croskerry, D. Sinclair, Emergency medicine: a practice prone to error, Can. J. Emerg. Med. 3 (2001) 271–276.

[3] S. Mchome, S. Sachdeva, S. Bhalla, A brief survey: usability in healthcare, International Conference On Electronics and Information Engineering 1 (2010) 463–467.

[4] J. Nielsen, Usability Engineering, 1st ed., Academic Press, London, 1993.

[5] R. Koppel, J.P. Metlay, A. Cohen, B. Abaluck, A.R. Localio, S.E. Kimmel, et al., Role of computerized physician order entry systems in facilitating medication errors, J. Am. Med. Assoc. 293 (2005) 1197.

[6] ISO/IEC, IEC 62366: 2007 Medical Devices – Application of Usability Engineering to Medical Devices, 2010. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=3859 (01.11.10).

[7] T.R. Johnson, X. Tang, M.J. Graham, J. Brixey, J.P. Turley, J. Zhang, et al., Attitudes toward medical device use errors and the prevention of adverse events, Jt. Comm. J. Qual. Patient Saf. 33 (2007) 689–694.

[8] FDA, Human Factors Implications of the New GMP Rule Overall Requirements of the New Quality System Regulation, 2010. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/HumanFactors/ucm119215.htm (01.11.10).

[9] Council of the European Communities, Council Directive 93/42/EEC of 14 June 1993 Concerning Medical Devices, 2010. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993L0042:EN:HTML (01.11.10).

[10] T.H. Van De Belt, L. Engelen, S. Berben, L. Schoonhoven, Definition of health 2.0 and medicine 2.0: a systematic review, J. Med. Internet Res. 12 (2010) e18.

[11] G. Djukic, L. Mezzasalma, L. Serasini, S. Ghione, PAOLINA (PAziente on LINe, Ambulatoriale) as a Web application for facilitating the storage and the management of self-measured blood pressure data, Comput. Cardiol. (2008) 889–892.

[12] M. Oefinger, W. Zong, M. Krieger, R.G. Mark, An interactive Web-based tool for multi-scale physiological data visualization, Comput. Cardiol. (2004) 569–571.

[13] Y.S. Lim, D.D. Feng, T.W. Cai, A Web-based collaborative system for medical image analysis and diagnosis (2000) 93–95.

[14] R.R. Bond, D.D. Finlay, C.D. Nugent, G. Moore, A Web-based tool for processing and visualizing body surface potential maps, J. Electrocardiol. 43 (2010) 560–565.

[15] R.R. Bond, D.D. Finlay, C.D. Nugent, G. Moore, D. Guldenring, A simulation tool for visualizing and studying the effects of electrode misplacement on the 12-lead electrocardiogram, in: Proceedings from VIP workshop on visualization (Vol. 2), J. Electrocardiol. 44 (2011) 439–444.

[16] Adobe, Adobe Flash Player, 2010. Available: http://www.adobe.com/products/flashplayer/ (01.11.10).

[17] J.R. Hampton, The ECG Made Easy, 6th ed., Churchill Livingstone, Edinburgh, New York, 2003.

[18] Joint Recommendations of the American Heart Association and the Cardiac Society of Great Britain and Ireland. Standardization of Precordial Leads, Am Heart J. 15 (1938) 107–108.

[19] T. Thaler, V. Tempelmann, M. Maggiorini, A. Rudiger, The frequency of electrocardiographic errors due to electrode cable switches: a before and after study, J. Electrocardiol. 43 (2010) 676–681.

[20] A. Rudiger, L. Schob, F. Follath, Influence of electrode misplacement on the electrocardiographic signs of inferior myocardial ischemia, Am. J. Emerg. Med. 21 (2003) 574–577.

[21] A. Rudiger, J.P. Hellermann, R. Mukherjee, F. Follath, J. Turina, Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings, Am. J. Emerg. Med. 25 (2007) 174–178.

[22] R.A. Harrigan, Electrode misconnection, misplacement, and artifact, Emerg. Med. Clin. North Am. 24 (2006) 227–235.

[23] J. Garcia-Niebla, Morphologies suggestive of V1 and V2 lead misplacement, Rev. Esp. Cardiol. 61 (2008) 1109–1110.

[24] R.R. Bond, Electrode Misplacement Simulator, 2011, Available: http://scm.ulster.ac.uk/~scmresearch/bond/ems/ (01.11.11).

[25] J.P. Ornato, I.B.A. Menown, M.A. Peberdy, M.C. Kontos, J.W. Riddell, G.L. Higgins, et al., Body surface mapping vs 12-lead electrocardiography to detect ST-elevation myocardial infarction, Am. J. Emerg. Med. 27 (2009) 779–784.

[26] R.S. MacLeod, C.R. Johnson, in: Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Map3d; interactive scientific visualization for bioengineering data (1993) 30–31.

[27] R.R. Bond, Body Surface Potential Viewer, 2011, Available: http://scm.ulster.ac.uk/~scmresearch/bond/bspm/ (01.11.11).

[28] M.W.M. Jaspers, A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence, Int. J. Med. Inf. 78 (2009) 340–353.

[29] A. Joshi, M. Arora, L. Dai, K. Price, L. Vizer, A. Sears, Usability of a patient education and motivation tool using heuristic evaluation, J. Med. Internet Res. 11 (2009).

[30] J. Choi, S. Bakken, Web-based education for low-literate parents in Neonatal Intensive Care Unit: development of a website and heuristic evaluation and usability testing, Int. J. Med. Inf. 79 (2010) 565–575.

[31] J. Stinson, Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis, J. Med. Internet Res. 12 (2010) e30.

[32] CINC, Comput. Cardiol. (2011), Available: http://www.cinc.org (01.11.11).
[33] J. Nielsen, T.K. Landauer, A mathematical model of the finding of usability problems, in: Proceedings of the 1993 Conference on Human Factors in Computing Systems, 1993, pp. 206–213.
[34] S. Krug, Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems, 1st ed., New Riders Press, Berkeley, 2009.
[35] R.R. Bond, Pre-test Questionnaire, 2011, Available: http://scm.ulster.ac.uk/~scmresearch/bond/survey/pretest.html (01.11.11).
[36] Apple, Apple QuickTime, 2011, Available: http://www.apple.com/quicktime/ (01.11.11).
[37] J. Kjeldskov, M.B. Skov, J. Stage, A longitudinal study of usability in health care: does time heal? Int. J. Med. Inf. 79 (2010) e135–e143.
[38] C. Lewis, Using the "Thinking-Aloud" Method in Cognitive Interface Design, IBM TJ Watson Research Center, 1982.
[39] R.R. Bond, Post-test Questionnaire, 2011, Available: http://scm.ulster.ac.uk/~scmresearch/bond/survey/post_bspm%20viewer.html (01.11.11).
[40] R.R. Bond, Post-test Questionnaire, 2011, Available: http://scm.ulster.ac.uk/~scmresearch/bond/survey/post_simulator.html (01.11.11).
[41] D. Travis, How to Prioritise Usability Problems, 2010, Available: http://www.userfocus.co.uk/articles/prioritise.html (01.11.10).
[42] BiomedMe, National Institute for Standards and Technology (NIST) to Evaluate Ease of Use of Health IT Systems, 2011, Available: http://biomed-me.com/general/national-institute-for-standards-and-technology-nist-to-evaluate-ease-of-use-of-health-it-systems_6213.html (01.11.11).