Journal of Clinical Monitoring and Computing (2004) 18: 365–372 

# USABILITY EVALUATION OF A GUI PROTOTYPE FOR A VENTILATOR MACHINE

*Yuanhua Liu, Lic. Tech. and Anna-Lisa Osvalder, PhD*

Liu Y, Osvalder A-L. Usability evaluation of a GUI prototype for a ventilator machine.
J Clin Monit 2004; 18: 365–372

**ABSTRACT. Objective.** Information presentation on the monitor screen of ventilator machines affects nurses' response and decision-making during ventilation treatment. The purpose of this study is to evaluate whether a new GUI (graphical user interface) prototype, the so-called *circular display* prototype can make deviations easy to detect. **Method.** A numerical display prototype was made and used as a reference display in the evaluation. Six task scenarios that involved parameter changes were selected to simulate a real situation under volume control (VC) mode during ventilation treatment. Usability tests with the two display designs were carried out in a usability laboratory. Twenty medical nursing students participated as test subjects in the usability tests. **Results.** The objective results showed that the graphical circular display had an advantage over the numerical display in interpreting parameter changes, but not in reducing the error rates for detecting the number of parameter changes or for forming an overall picture of the patient's situation. Furthermore, the circular display prototype did not improve the detection time. **Conclusions.** Although the majority of the test subjects preferred the graphical circular display, the results implied that several aspects of this prototype should be improved in a future development study.

**KEY WORDS.** Graphical circular display, numerical display, prototype, usability tests, deviation detection.

## INTRODUCTION

Due to the wide use of computer-based medical devices in hospitals, monitoring has become an essential and major task for clinicians during medical treatment. Visual displays of physiological information influence a clinician's response to changes in a patient's condition during administration of anaesthesia [1]. In the medical context, data displays that organize and present data in a format that matches the cognitive interpretation made by the clinician should enhance the speed and accuracy of decision-making [2]. Consequently, the data display format becomes an important factor in clinical monitoring and decision-making. In critical care, e.g. in anaesthesia work, very few studies have examined the effect of the format of clinical data presentation on the ability of anaesthesiologists to efficiently and accurately detect changes in the patient's condition [3].

By using a graphical circle diagram of the physiological state space, Siegel et al. [4] made a variety of studies of alterations in physiological, biochemical and immunological variables in order to analyze alterations in patients after trauma sepsis, surgical liver disease and other critical illnesses. For example, Siegel et al., used a circle diagram

From the Department of Product and Production Development, Division of Human Factors Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden.

Received 28 April 2004, and in revised form 2 December 2004. Accepted for publication 13 December, 2004.

Address correspondence to Y. Liu, Department of Product and Production Development, Division of Human Factors Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden.
E-mail: yuanhua@hfe.chalmers.se

to show physiological classification by which physicians or clinicians can compare the stress responses of critically injured or ill patients with patients who are in a normal state, considering the adequacy of their specific response pattern to meet their physiological needs. The purpose of this study was not only to devise and use the circle diagram to gain a better understanding of the nature of physiologic variables, but also to be able to quantify the patient's degree of improvement or deterioration in physiological state.

In a study by Gurushanthaiah et al. [3], the efficiencies of three display formats, i.e. numerical, histogram and polygon displays, were evaluated and compared in a partial-task laboratory simulation. The results showed that graphic displays might enhance the detection of acute changes in patient physiological status during anaesthesia administration. Michels et al. [5] compared an updated graphical anaesthesia display with the traditional waveform alpha-numerical anaesthesia display. They found that some of the simulated critical events were not only detected correctly but also identified faster on an integrated graphical display than on a traditional numerical one.

In order to reduce the mental workload of nurses and to improve the accuracy of monitoring work in the ICU, a new graphical *circular display* prototype of a ventilator machine was proposed by a large international medical company. The purpose of the present study was to evaluate whether the new graphical circular display prototype of a ventilator machine could make deviations easy to detect, and the results will be used in further development of the prototype.

## MATERIALS AND METHODS

The graphical circular display prototype proposed by the medical company was a computer prototype. The visual representation provided by prototype was expected to make it '*Easy to detect deviations*': (1) easy to detect a parameter that deviates; (2) easy to understand the meaning of the deviation; (3) easy to monitor the information at a distance; and (4) easy to get the overall picture of the patient's condition.

The circular form on the display prototype was divided into six equal parts, each of which represents a critical parameter for Volume Control (VC) mode during ventilation treatment (Figure 1): Ppeak (Airway Pressure), PEEP (Positive End Expiratory Pressure), MVe (Minute Volume), RR (Respiratory Rate), $FiO_2$ (Amount of Oxygen Delivered), $etCO_2$ (End Tidal Carbon Dioxide). Upper and lower alarm thresholds were illustrated with two circles; the coloured parts in-between represented the actual values of the parameters. When the actual values of the parameters were in a normal state, a perfect coloured circle was shown. According to the magnitude of the deviation, the deviation of a parameter value was shown in three steps, each of which was coded with one of three colours – green (normal state), blue (attention notice) and yellow (urgent alarm).

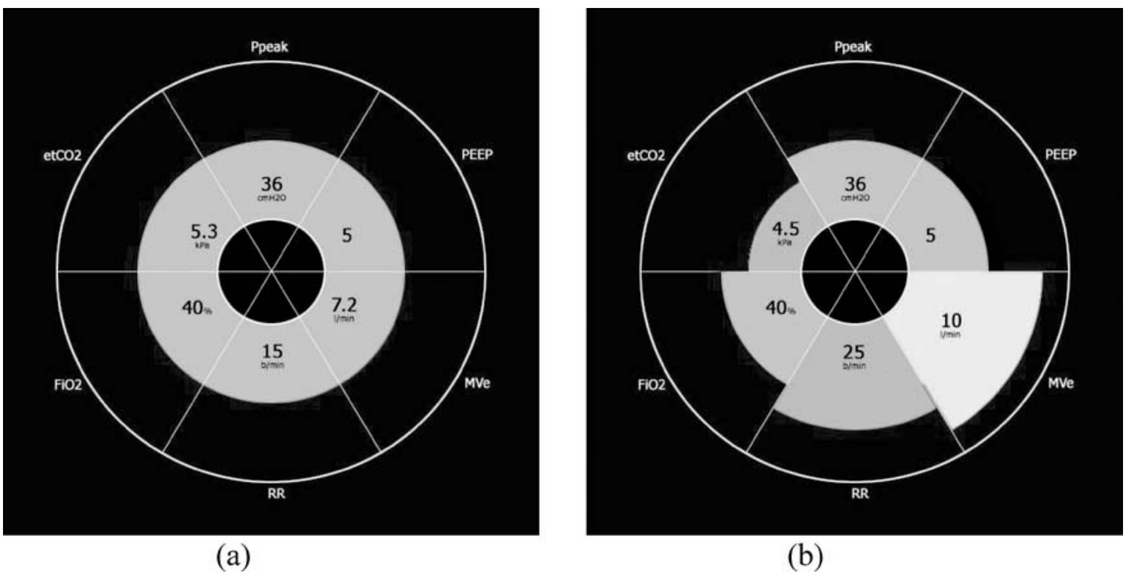



*Fig. 1. (a) Normal state, (b) a deviation state.*

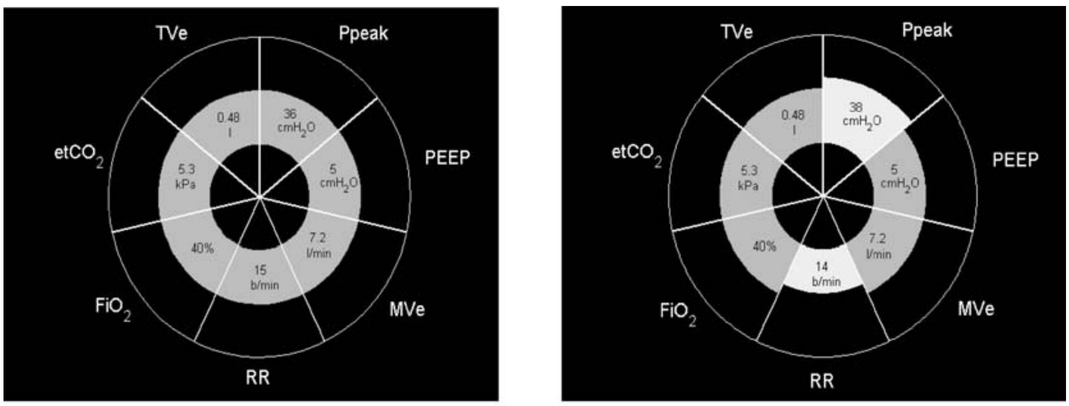
### *Interview study*

In the first part of this study, interviews were carried out with six expert nurses at the intensive care unit (ICU) at Sahlgrenska University Hospital in Göteborg, Sweden. Although an integrated graphical display that consists of numerical data and a traditional wave-form alphanumerical diagram is used on the monitor screen of ventilators in ICUs at many hospitals nowadays, the interview results showed that most nurses read only the numerical values when they looked at the monitor display during ventilation treatment. The reason given for this was that it is difficult to understand the meaning of the traditional wave-form alphanumerical diagram. Therefore, the integrated graphical display functioned only as a numerical display. From the interviews, it was found that TVe (Tidal Volume) also is a critical parameter in the Volume Control mode during ventilation treatment. Although alarms are important for ventilation work, interviews also made clear that alarm with fewer hierarchical levels were preferred, since alarms with too many hierarchical levels would distract and stress the nurses.

### *Modifications of the two display prototypes*

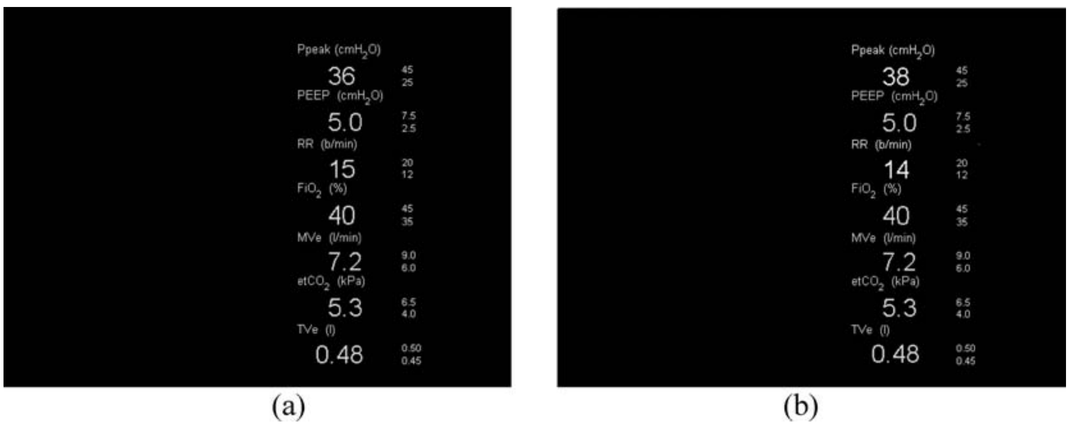
Based on the results of the interviews, the initial circular display prototype from the medical company was modified. Tidal Volume (TVe) was also added as a critical parameter in the display. Furthermore, only one alarm level was kept, which was indicated with yellow colour (Figure 2).

For comparison, a numerical display prototype was developed as a reference display. This prototype is a modification including no graphical curves which are normally included on ventilator monitor screens nowadays (Figure 3).




(a) (b)

*Fig. 2. The modified circular display in (a) normal state, (b) a deviation state.*




(a) (b)

*Fig. 3. Reference numerical display prototype in (a) normal state, (b) a deviation state.*

### *Usability test*

The usability test is an experimental method that is normally used for evaluation of user interfaces in many areas. In the present study, the usability tests were conducted according to a general testing procedure proposed by Nielsen [6] and McClelland [7]. The tests were made in the usability laboratory at Chalmers University of Technology, Sweden. Twenty medical nursing students from the nursing school in Göteborg participated as test subjects. Each test subject was requested to participate in two sessions, one with the numerical display prototype and one with the graphical circular display prototype. The sequence of the six task scenarios was randomized for the two test sessions, to avoid the learning effect. Prior to the usability tests, four pilot tests were carried out to check and adjust the test procedure.

With the help of expert nurses, six task scenarios were selected to represent deviation states that would simulate the actual situation of Volume Control (VC) mode during ventilation treatment in the usability test. Each scenario consisted of two pictures, where the state before deviation and the state after deviation were shown: (1) Normal state–$FiO_2$ is changed; (2) Normal state–RR, MVe and $etCO_2$ are changed; (3) RR, MVe and TVe are changed–RR, MVe, $etCO_2$ and TVe are changed; (4) Ppeak, RR and TVe are changed–RR, MVe and $etCO_2$ are changed; (5) Normal state–MVe, $etCO_2$, RR, Ppeak and TVe are changed; (6) MVe, Ppeak and $etCO_2$ are changed–Normal state.

During the evaluation, the test subjects were asked to detect and interpret visible deviations of any of the seven parameters, to judge the severity of the deviation by taking into account the overall situation, and to explain the reasons for their decision when they detected and confirmed any deviations as shown on the monitor screen.

At the end of the second test session, each test subject was requested to compare the two display prototypes in a discussion with the test instructors, so that all of the test subjects could give their opinions completely. Furthermore, the subjects were encouraged to comment freely on the circular display design, for instance, its colour coding and letter size.

For each task scenario, data on detection time and error rates during deviation detection were collected. Two kinds of data were gathered during the usability tests: objective and subjective. Objective data, such as detection time and error rates during deviation detection, were collected from the digital stopwatch and the video tapes. Subjective data, such as judgment of the severity of the deviations by taking into account the overall situation, selection of the display prototype that makes it easier to detect deviations and reasons for this, and selection of the display prototype that makes it easier to interpret the overall situation and reasons for this were collected from the video tapes and from the recorded discussions afterwards.

## RESULTS

### *Objective results*

The detection time needed for each task scenario reflects the response time and sensitivity to the data representation on the display. The hypothesis was that the display format had an effect on the detection time. $T$ test was conducted on the data collected from each scenario. Since $P$ values from all the six scenarios were larger than 0.05, i.e., $P_i$ ($i$ = 1, 2, 3, 4, 5, 6) > 0.05, the hypothesis was rejected (Table 1). This means that there was no significant difference between the numerical display format and the graphical circular display format in terms of detection time.

*Table 1. T-test results on detection time from the two display prototypes*

| | | | | Paired differences | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 95% confidence interval of the difference | | | | |
| | *N* | Correlation | Sig. | Mean | Std. deviation | Std. error mean | Lower | Upper | *t* | df | Sig. (2-tailed) |
| Pair 1 (S1G–S1N) | 20 | 0.192 | 0.418 | −5.60 | 27.580 | 6.167 | −18.51 | 7.31 | −0.908 | 19 | 0.375 |
| Pair 2 (S2G–S2N) | 20 | 0.084 | 0.724 | −7.35 | 37.294 | 8.339 | −24.80 | 10.10 | −0.881 | 19 | 0.389 |
| Pair 3 (S3G–S3N) | 20 | 0.455 | 0.044 | −11.30 | 25.446 | 5.690 | −23.21 | 0.61 | −1.986 | 19 | 0.062 |
| Pair 4 (S4G–S4N) | 20 | 0.339 | 0.144 | 6.10 | 55.284 | 12.362 | −19.77 | 31.97 | 0.493 | 19 | 0.627 |
| Pair 5 (S5G–S5N) | 20 | 0.557 | 0.011 | −4.45 | 17.437 | 3.899 | −12.61 | 3.71 | −1.141 | 19 | 0.268 |
| Pair 6 (S6G–S6N) | 20 | 0.108 | 0.649 | −7.35 | 38.453 | 8.598 | −25.35 | 10.65 | −0.855 | 19 | 0.403 |

*Note.* S: Scenario; G: graphical 'cake display' prototype; N: numerical display prototype.

*Table 2. T-test results on error rates during deviation detection from the two display prototypes*

| | | | | Paired differences | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 95% confidence interval of the difference | | | | |
| | $N$ | Correlation | Sig. | Mean | Std. deviation | Std. error mean | Lower | Upper | $t$ | df | Sig. (2-tailed) |
| Pair 1 (NA–GA) | 20 | −0.077 | 0.748 | 0.55 | 1.669 | 0.373 | −0.23 | 1.33 | 1.473 | 19 | 0.157 |
| Pair 2 (NB–GB) | 20 | 0.771 | 0.000 | 3.05 | 2.743 | 0.613 | 1.77 | 4.33 | 4.973 | 19 | 0.000 |
| Pair 3 (NC–GC) | 20 | 0.261 | 0.267 | 0.45 | 1.146 | 0.256 | −0.09 | 0.99 | 1.756 | 19 | 0.095 |

*Note*. G: graphical 'cake display' prototype; N: numerical display prototype; A: error rates in detecting number of parameter changes; B: error rates in interpreting the meaning of changes; C: error rates in assessing the overall situation.

Error rates during deviation detection include three situations: (A) error rates in detecting the number of deviations; (B) error rates in interpreting the meaning of deviations; (C) error rates in assessing the overall situation. *T*-test was conducted to investigate the error rates in these three situations (Table 2). The hypothesis for situation (A) was that the display format had an effect on error rates in detecting the number of deviations. According to the *t*-test result, since $P > 0.05$, the hypothesis was rejected. This means that there was no significant difference between the numerical display format and the graphical circular display format in terms of error rates in detecting the number of deviations. For situation (B), errors rates in interpreting the meaning of deviations, the hypothesis was that the display format had an effect on these too. According to the *t*-test result, since $P < 0.05$, the hypothesis was accepted. This implies that the graphical *circular display* prototype induced fewer errors in interpreting the meaning of deviations. For situation (C), the hypothesis was that the display format had an effect on error rates in assessing the overall situation. According to the *t*-test, since $P > 0.05$, the hypothesis was rejected, implying that there was no significant difference between the two display formats in terms of error rate in assessing the overall picture of the situation.

### *Subjective results*

The severity of the deviations, by taking into account the overall situation, was related to the number of deviations, values of each deviation and the specific medical care conditions. For both display formats, the number of yellow-coloured parameters was counted as an important clue when judging the severity of the deviations. In addition, the majority of the test subjects took into account the value of each deviation when monitoring with the graphical circular display. Some test subjects thought that medical care conditions such as skin temperature, pupil size, and pulse should also be used in judging the severity of the deviations.

Most of the test subjects thought it was easier to detect changes with the graphical circular display. The reason was that the deviation of the shapes was easy to detect. In addition, some test subjects attributed reasons to advantages of the graphical display, with expressions such as: 'providing less memory load' and 'easier to detect value changes'.

Many of the test subjects found it easier to interpret the overall situation with the graphical circular display than with the numerical display. This was attributed to advantages of the graphical display, with comments such as 'providing hints from shape deviations' and 'easier to compare with the reference field'. However, there were still some test subjects who thought it was easier to interpret the overall situation with the numerical display, because of the alarm thresholds that were presented with numerical values.

When the test subjects were consulted about their preferred display format, the majority of them chose the graphical circular display. They explained by referring to advantages of the graphical display with the following comments: 'easier to detect deviations' and 'better overall picture'. Only a few of the test subjects preferred the numerical display, with reference to advantages such as 'having classical or typical display format for medical devices'.

When the test subjects were encouraged to comment freely and make suggestions to improve the graphical circular display design, these could be divided into four categories: (1) The alarm thresholds were suggested to be presented in numerical values as well in the graphical display. (2) The former values were suggested to be included as the reference values in the graphical display. (3) Yellow and green colours were difficult to distinguish on the graphical display. (4) Enlarging the size of numbers shown in the graphical display was recommended.

## DISCUSSION

### *Test results versus design expectations*

The focus of this study was to evaluate a graphical circular display prototype to determine whether it could make deviations easier to detect. The objective results from the usability tests showed that the graphical circular display was not better than a traditional numerical one in shortening the deviation-detection time. The results also implied that the graphical circular display did not reduce the errors during the detection in comparison with the numerical display. Therefore, the graphical circular display could not satisfy the first point of the design expectation: '*easier to detect when a parameter is deviating*'.

However, the results showed that the graphical *circular display* did offer an advantage in detecting the meaning of changes, when compared with the numerical display. This has to do with *Gestalt* which refers to a configuration, whole pattern, or form. Gestalt psychology deals with whole patterns, and scientists come to notice the importance of pattern recognition and its influence in the designs. Humans have inborn traits of organization that influence the interpretation of objects [8]. Preece [8] points out that graphical representation can make it easier to perceive the trends in data that are constantly changing, the relationships between multidimensional data, and defects in patterns of real-time data. The evaluation results from the present study proved that the graphical design can help people interpret deviations in a pattern easily. This showed that the graphical circular display achieved the second point of the design expectation: '*easier to understand the meaning of the deviations*'.

Vision can be affected by many factors, such as ambient lighting, visual angles and contrast. The graphical circular display format is a preliminary design proposal with no detailed information about the interface design. Therefore, the third point of the design expectation, '*easy to monitor the information at a distance*' was not tested in the evaluation due to the limitation of the prototype design.

The evaluation results showed that the graphical circular display offered no advantage in assessing the overall picture of the patient's condition, which meant that the fourth point of the design expectation, '*easier to get the overall picture of the condition*', was not satisfied by the graphical circular display. The monitoring task in an ICU ward is not a simple screen-monitoring task: it is as mentally complicated as the task of a pilot in a cockpit, in that it requires awareness of a complete situation. Situation awareness (SA) refers to an understanding of the state of an environment, which provides the primary basis for subsequent decision-making and performance in the operation of complex dynamic systems [9–11]. According to Endsley [9], perceiving relevant information and integrating various pieces of information (in conjunction with the present goals) affect the accuracy of predicting future events and the state of the system as well as initiating effective decision-making. In the present study, the overall picture of the patient's condition could be detected and understood only by integrating various pieces of information in conjunction with the corresponding medical treatment modes. The evaluation results of the study indicated that the design of the graphical circular display did not contribute much to interpreting the overall picture of the situation. In other words, it was implied that education and training of ventilation treatment play an important role when interpreting the overall picture of the situation, especially when an acute event occurs.

### *Disagreement between objective and subjective evaluation results*

Agreement between subjective and objective evaluation results is always a concern in many studies. Muckler and Seven [12] stated an assumption that both forms of measure would reflect the same underlying reality if only the right things were measured in the right way. However, Sinclair [13] argued that disagreement between subjective and objective evaluation results may provide more information than agreement would. In addition, it was indicated that the opposite may apply where qualitative judgments or multivariate measurements are concerned, even though experimental instruments can be more sensitive, more reliable and more accurate than human beings [13]. In the present study, the objective measurements on the test subjects' performance and the subjective assessments did not tell a consistent story. Most test subjects thought that the graphical circular display prototype makes it easier to detect deviations, but the objective tests on performance did not prove it. The subjective assessments implied that subjective feeling, for instance '*easier to detect deviations*', is multidimensional. In addition, the disagreement between the subjective and objective results revealed the difference between the design expectations and test subjects' view in interpreting '*easier to detect deviations*'. This implies that improvement is necessary for the graphical circular display design. The comments and suggestions for improvement given by the test subjects should be taken into consideration in further development studies.

### *Hints from the subjective results*

In the subjective assessment, the test subjects were asked to explain how they judged the severity of the situation during the ventilation treatment. Their answers showed that colour coding contributed much to detecting deviations of parameters. In addition, although pattern recognition is helpful in representing the trends in data that are constantly changing, it was shown that numerical coding is still required to help measure the amount of the deviation.

When comparing the two display formats in terms of '*easier to detect the deviations*', only a few test subjects chose the numerical display. This implied the existence of personal differences and personal preferences, which was also shown in the objective results from the usability tests. The analysis of the objective data indicated that personal differences and preferences affected the test results.

### *Limitations of the study*

The graphical circular display is a preliminary design proposal with no concrete detailed information. This adds difficulties to the evaluation work, especially in assessing the achievement of the third point of the design expectation: '*easy to monitor the information at a distance*'. The difference between the environment of the usability laboratory and the real working environment in ICUs is also counted as a limitation. In the present study, the test subjects were highly vigilant because they concentrated on the monitoring task in a silent laboratory. In reality, ventilator nurses are likely to be disturbed and distracted by other things during monitoring work – for example, doctors' and nurse assistants' inquiries or patients' reaction. If the usability tests were conducted in a real-world working situation, the detection time might have been different, and then in favour of the graphical circular display. Additionally, in the present study, only Volume Control mode was chosen for the evaluation since the project time was limited. Treatment differs widely for different mode types, as well as for different working environment settings. More discussion and research should be carried out to determine the applicability of the graphical circular display to other modes of treatment.

### *About graphical display design*

Siegel et al. [4] found that their graphical display designs of haemodynamic and cardiorespiratory parameters facilitated in the recognition of different kinds of pathophysiological conditions such as septic shock, etc. The graphical display design in all their studies is useful for clinicians in analysis and diagnosis. Time was not the key issue here. However, in the present study, the design of the graphical circular display was useful especially for nurses in ICUs. Fast decision-making in a short time is critical in ICUs. The efficiency of a GUI design is affected by many factors, e.g. colour coding, symbols, layout of the display and personal differences. These considerations need to be taken into account in the design process. In the present study, problems with colour coding, text size and personal references were revealed by the results. Nowadays, there is still no universal standard for how to design these factors on the graphical interfaces, because some factors such as specific design purposes, specific use environment, specific users, and specific interface specifications also play dominant roles in the design. Therefore, usability evaluation becomes necessary in the design process. In order to make a good evaluation, a usability test with appropriate users and tasks is always seen as an essential evaluation tool; and in the design process its important role can never be neglected.

## CONCLUSIONS

The following conclusions can be drawn from the present study, based on the usability test results:

1. The graphical circular display is better than the numerical display in helping users to interpret the meanings of parameter deviations accurately.
2. The graphical circular display does not contribute to improving the users' detection time.
3. Concerning the errors in detecting the number of deviations and the error rates in getting the overall picture of the situation, the graphical circular display does not contribute any improvement.
4. The graphical circular display does not facilitate '*ease of detecting deviations*' better than the normal numerical display, although the graphical circular display was subjectively preferred by more of the test subjects.
5. Further improvements and modifications are necessary for the graphical circular display design, and an integrated graphical display which combines both graphical and numerical representation has been suggested.
6. Usability tests are valuable for obtaining information about users' performance and comments on the prototype design in the early design process.

## REFERENCES

1. Gaba D. Human error in dynamic medical domains. In: Human error in medicine. 1994; 197–224.
2. George TB, Stephen DS, Kate W. A graphical object display improves anesthesiologists' performance on a simulated diagnostic task. J Clin Monit 1999; 15: 37–44.
3. Gurushanthaiah K, Weinger MB, Englund CE. Visual display format affects the ability of anesthesiologists to detect acute physiologic changes. Anesthesiology 1995; 83: 1184–1193.

4. Siegel JH, Baker SP, Gennarelli TA. Panel: Current status of trauma severity indices. J Trauma 1983; 23(3): 185–201.
5. Michels P, Gravenstein D, Westenskow DR. An integrated graphic data display improves detection and identification of critical events during anaesthesia. J Clin Monit 1997; 13: 249–259.
6. Nielsen J. Usability engineering. San Diego: Academic Press, 1993: 165–200.
7. McClelland I. Product assessment and user trials. In: Wilson JR, Corlett EN, eds. Evaluation of human work. London: Taylor and Francis, 1995: 118–247.
8. Preece J. Human-computer interaction. Harlow, England: Addison-Wesley, 1994: 78–89.
9. Endsley MR. Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society 32nd annual meeting, 1988: 97–101.
10. Endsley MR. Measurement of situation awareness in dynamic systems. Hum Factors 1995; 37: 65–84.
11. Endsley MR. The role of situation awareness in naturalistic decision making. In: Zsambok CE, Klein G, eds. Naturalistic decision making. Mahwah, NJ: Earlbaum, 1997: 269–283.
12. Muckler FR, Seven SA. Selecting performance measures: 'objective' versus 'subjective' measurements. Hum Factors 1992; 34: 441–455.
13. Sinclair AM. Subjective assessment. In: Wilson JR, Corlett EN, eds. Evaluation of human work. A practical ergonomics methodology. London: Taylor and Francis, 1990: 58–88.