

Evaluation of a Touch-based User Interface for a Naval Command & Control System

J. Schwarz, O. Witt

Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE)

Abstract – Operators on naval ships are assisted in their assessment of the tactical situation around the ship and their subsequent actions by Command & Control Systems (C2-Systems). As the development of C2-Systems was mainly focused on a wider automation of working processes, little effort was spent on the ergonomic design of the Human-Machine-Interface (HMI). In a study assigned by the German Navy the aim was to develop a concept for an improved user interface for a naval C2-System. The concept was assessed in three experimental evaluations – in comparison with two alternative design concepts, in comparison with the current C2-System of a Corvette and on a moving platform which simulates ship movements. Results indicate that the new design concept with a touch-based user interface is a promising alternative to conventional interfaces for naval operations.

Keywords

Evaluation, touch-based user interfaces, naval Command & Control Systems.

INTRODUCTION

The domain of naval warfare contains highly time and safety critical tasks. For example, if an airborne object suddenly stops following the air route and starts heading towards the own ship, it is crucial, that the operator in the ship's Command Center gets aware of this event at an early stage, checks the identity of the object, and, if necessary, initiates an engagement in time.

A Command & Control System (C2-System) supports the operator for these tasks. The operator can see the objects (so-called "tracks") which are detected by sensors in the surrounding of the ship on a radar screen (Tactical Display Area - TDA). The system also provides information on the tracks' properties like altitude, velocity, bearing or its IFF (Identification Friend or Foe) response. IFF is an

electronic radio based identification system, which enables identification on the basis of a code request.

Since many steps in the identification and engagement processes run semi-automatically, the problem arises that the operator gets out of the loop – the so-called "irony of automation" (Bainbridge, 1987) which is enforced by a complex user interface. Therefore Fraunhofer FKIE was assigned by the German Navy to develop a user interface for a C2-System with respect to the tasks „tactical picture compilation“ and „engagement“ with the purpose to enhance situation awareness and guarantee quick and proper operating conditions by improved visualisations, workflows and appropriate interaction techniques.

The fulfilment of goals and requirements was tested in three experimental scenario-based evaluations: The first evaluation was a comparative assessment of three alternative design concepts. Secondly a comparison of the final design concept with the user interface of a current C2-System was conducted. Finally the design concept was tested on a moving platform, which simulates ship movements.

EVALUATION OF THREE DESIGN CONCEPTS

First, three design concepts were evaluated which differed in hardware, input devices and visualisations (see figure 1). The first design concept is called "conventional" as the input techniques consist of a computer mouse and a keyboard. Additionally a keypad with programmable soft keys can be used for inputs. For the display device a 30" monitor is used.

In contrast to the first concept the second design concept is characterised by the fact that all inputs are done via a 12" Tablet-PC with a pen. Therefore this concept is called "central input panel". For the display several monitors are used that provide the possibility to show all the information simultaneously without having to open additional windows or tabs.

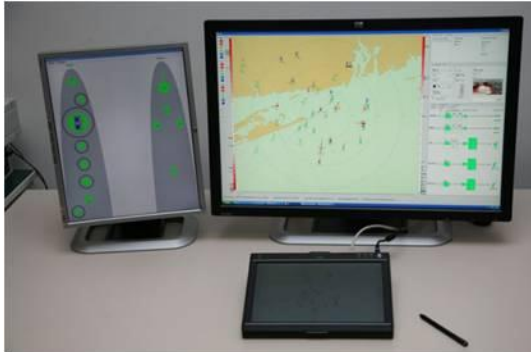
The third design concept is called "multitouch application" as it only consists of a 24" multitouch monitor. This monitor is used for display as well as input. All inputs occur by touch; so the design requires touch-specific layouts and functionalities.

Corresponding author

Name: Dipl.-Psych. Jessica Schwarz
Affiliation: Fraunhofer Institute for Communication,
Information Processing and Ergonomics
Address: Neuenahrer Str. 20
53343 Wachtberg
Germany
Email: jessica.schwarz@fkie.fraunhofer.de
Phone: +49-(0)228-9435491



„conventional“ design concept



„central input panel“



„multitouch application“

Figure 1. Design concepts 1-3 (from top to bottom)

Method

Participants

Eighteen research associates of Fraunhofer FKIE participated in the evaluation. The participant group consisted of ten men and eight women aged 19 to 39 years.

Task Description

A scenario-based test was conducted for the assessment of these design concepts. Subjects had to perform tactical picture compiling tasks on each concept. Since most of the subjects just had little knowledge of this work domain, tasks were designed in such a way that they are operable by inexperienced subjects as well. The tasks included the following steps:

- Selection of a track on the TDA,

- Initiation of an IFF-interrogation (this is done by selecting the buttons for the code request),
- Identification of the track on the basis of the IFF-information (assignment of the ID friendly, neutral or hostile),
- Classification of the track (determining the type of a track e.g. deciding if it is a fighter or air plane),
- Retrieval of information on the track (e.g. velocity, altitude and course).

These tasks had to be performed consecutively for each track that appeared as a new object on the TDA.

Variables and Experimental Design

The evaluation was conducted in a within-subject design with $N = 18$ subjects. The order of presenting the design concepts was changed systematically in order to avoid sequence effects. Three subjects always assessed the design concepts in the same order due to six possible variations.

For the assessment of the three design concepts the following dependent variables were measured:

- Task performance (accuracy and required time for task execution),
- Subjective workload according to NASA-TLX (Hart & Staveland, 1988),
- Subjective assessment of the design concept with questionnaires.

The design concept and the type of task were used as independent variables.

Procedure

For each design concept the evaluation started with an introduction and a training phase in which the task execution could be practised. This was followed by the real test scenario which consisted in performing the tasks for eight tracks. Afterwards, subjects filled out the NASA-TLX together with a questionnaire for subjective rating. The evaluation took about one hour for each design concept.

Results

Task performance

Regarding the accuracy of task execution, the error rate was very low. So, it did not provide any insights on differences between the concepts. To analyse the time spent for task execution, a repeated-measures ANOVA was conducted with the two factors “design concept” and “type of task”. The ANOVA indicated significant differences between the type of tasks, $F(4,68) = 60.19$; $p < .001$ as well as a significant interaction between the two factors design concept and type of task, $F(8, 136) = 20.61$; $p < .001$. The results are illustrated in figure 2.

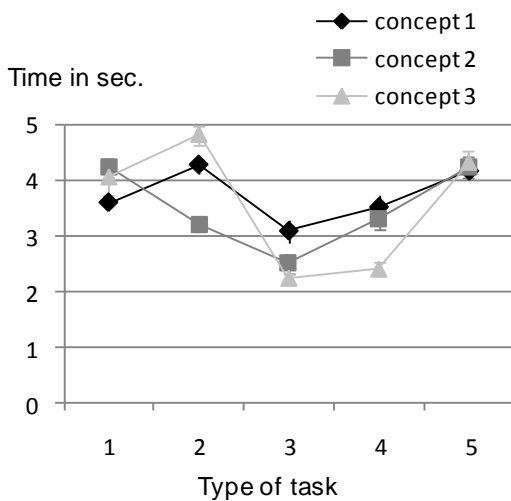


Figure 2. Mean time spent on task execution (1 = Selection, 2 = IFF-Interrogation, 3 = Identification, 4 = Classification, 5 = Retrieval of information).

As can be seen, the time spent for the various design concepts is strongly influenced by the type of task. Post hoc testing revealed significant differences between the concepts for the tasks IFF-interrogation, $F(2,34) = 71.57$; $p < .001$, Identification, $F(2,34) = 21.35$; $p < .001$ and Classification, $F(2,34) = 13.53$; $p < .001$. While concept 3 required the most time for the task IFF-interrogation, it turned out to be the fastest for the tasks Identification and Classification.

Subjective Workload

The subjective workload was analysed for the subscales of NASA-TLX by using the mean value of every scale in a repeated measures ANOVA. The ANOVA indicated a significant main effect for the concept type, $F(2,34) = 6.16$; $p < .05$.

A descriptive analysis revealed that the workload was rated a bit higher for concept 1 ($M = 3.76$; $SE = 0.37$) than for the other two concepts ($M = 2.76$; $SE = 0.23$ for concept 2; $M = 2.94$; $SE = 0.31$ for concept 3). But since the used scale ranges from 1 (very low workload) to 15 (very high workload) the overall workload level is rather low for all three concepts.

Subjective Rating

For each design concept subjects were requested to rate how well the concept supports task execution. In an ANOVA the overall differences between the concepts did not turn out to be significant. The analysis revealed that all concepts were rated quite positive.

But in a final rating, which was conducted after all three concepts had been tested, clearer results could be found. Subjects were asked to name their most and their least favourable concept and the concept which was easiest to use. Table 1 shows how frequently each concept was named in this survey.

	Concept 1	Concept 2	Concept 3
Most favoured	5	2	11
Least favoured	4	13	1
Easiest to use	6	4	8

Table 1. Number of times each concept was named as the most favourable, least favourable and easiest to use concept.

The results clearly indicate that concept 3 was the most favoured concept and concept 2 the least favoured. Concept 3 was also most often named as the concept that was easiest to use.

Discussion

The evaluation indicated no large differences between the design concepts with respect to task performance. This might be due to the fact that each task could be accomplished in quite a short time of only a few seconds which narrowed the variance. The low error and workload rates also indicate rather easy to perform tasks. Nevertheless, workload ratings lead to significant differences between the concepts, with the highest workload rating for the conventional design concept (concept 1). So, the rather new interaction techniques which are used in the concepts 2 and 3 (pen and touch input) don't seem to affect the workload level in a negative way.

The subjective rating, especially the direct comparison of the concepts at the end of the evaluation, clearly revealed, that subjects liked the touch application the most. The interaction per touch was regarded as very intuitive and it was positively stated that no avert of one's gaze was required during task execution. This was also named as a reason why concept 2 was regarded as the least favoured concept by most of the subjects. As it consists of several screens while inputs are done on the Tablet-PC, the head has to be turned up and down and subjects have to reorientate themselves on the screens.

On the basis of these results it was decided to consider the multitouch-application as the final design concept. In a next step, it was expanded with functionality required for the engagement process (see Witt, Schwarz & Özyurt, 2010 for a detailed description) and compared to the user interface of a current C2-System in the second evaluation.

EVALUATION IN COMPARISON WITH A CURRENT C2-SYSTEM

A comparative scenario-based evaluation was conducted in order to examine if the touch-based design concept is an improvement over the user interfaces of current C2-Systems. For the

comparison, a C2-System of a Corvette was used, which was most recently placed into service and is therefore equipped with the most up to date C2-System.

Method

Participants

The participant group consisted of $N = 12$ operators of the German Navy (11 men and 1 woman). Half of them had good knowledge of the C2-System used in the evaluation. The other half was unfamiliar with the system. The age of the operators ranged from 22 to 48, $M = 26$.

Task Description

The evaluation consisted of a scenario-based test with respect to the tasks “tactical picture compilation” and “engagement” and a subjective assessment of the user interfaces. For the task tactical picture compilation the test was divided into two phases which differed in the degrees of freedom the operators were given.

Phase 1 is characterised by a rather high standardisation of task execution which should ensure comparable outcomes and a high internal validity. The task was to identify a certain track on the TDA every 30 seconds according to previously announced ID-Criteria. The situation awareness was measured in this phase by certain questions the operator had to answer during the scenario. According to the technique SPAM (Durso & Dattel, 2004) the time spent to answer the questions were taken as a measure of situation awareness. As an example a question was: “How many air tracks are heading towards the own ship?”

In phase 2 task execution corresponded more to real working conditions and was characterised by a low standardisation which should lead to a higher external validity. The operator was commissioned to observe the tracks on the TDA and to indicate, when tracks show changes in their behaviour that call for a change of their ID. Additionally new unknown tracks had to be recognised and identified self-dependently.

For the task engagement only a test with high standardisation of task execution was conducted due to technical restrictions. The test was based on an artificial scenario where tracks were heading towards the own ship at an interval of two minutes. For each approaching track operators had to initiate an engagement in a predefined way and answer questions about the current engagement process e.g. “With which weapon can the target be engaged at the earliest possible time?”.

Variables and Experimental Design

The evaluation was conducted in a within-subject design with $N = 12$ subjects. The experimental design consisted of three independent variables:

- User interface (touch-based interface vs. interface of current C2-System),
- Task (tactical picture compilation vs. engagement),
- Experience (operators with high and low experience with respect to the C2-System).

The order in which the touch-based user interface and the user interface of the current C2-system was tested and assessed was varied in order to avoid sequence effects.

For each task and each user interface the following dependent variables were measured:

- Task performance (accuracy and required time),
- Situation Awareness,
- Subjective Workload,
- Subjective assessment of the interfaces with questionnaires.

Procedure

The evaluation started for each user interface with the task tactical picture compilation. After an introduction and a training phase the test for phase 1 was conducted. Subsequently subjects had to rate their workload. Phase 2 started without a training phase and was followed again by a workload rating. At the end the user interface was assessed using questionnaires with respect to the task tactical picture compilation. The procedure was the same for the task engagement with the difference that the second phase was left out. Overall the assessment of both interfaces took one day for each participant.

Results

Task performance regarding Tactical Picture Compilation

As a measure of accuracy the error rate was used. That is the percentage of times a track was assigned with either a wrong or no ID within a 30 seconds timeslot. For the touch-based user interface the error rate was a bit higher (17%) than for the user interface of the current C2-System (14%) but the difference did not turn out to be significant.

Regarding the time spent for the identification the average identification time for the touch-based user interface was $M = 14.4$ with $SD = 3.4$ and for the interface of the current system $M = 16.2$ with $SD = 3.0$. It took a bit more time to set the ID in the current system, even though the difference was not significant.

Task Performance regarding Engagement

For the task engagement, all tasks could be accomplished without errors with both interfaces except one task where the error rate was 17% for the user interface of the C2-system. The task was to select the track with the least time to impact. With the touch-based user interface no errors occurred.

With respect to the required time, the user interfaces only differed significantly on two of the five tasks. The tasks were to select the track with the least time to impact and to initiate an engagement with a specific weapon. For these tasks significantly more time was required with the current C2-System than with the touch-based user interface, $t(11) = 6.09$, $p < .001$ and $t(11) = 3.71$, $p < .01$.

Situation Awareness

In phase 1 of the task tactical picture compilation the situation awareness was measured by questions on the current situation. The time to answer these questions did not lead to significant differences but a significant difference could be found regarding the error rate, $t(11) = 3.53$; $p < .01$. The error rate was 13% for the touch-based user interface and 22% for the current C2-System.

An indication of the situation awareness can also be inferred from the events that were recognised in the second phase. The user interfaces do not differ significantly in the total percentage of recognised events (92% for the touch-based interface, 84% for the user interface of the C2-System) but a separate examination for each event shows that almost every event was recognised by more subjects on the touch-based interface. As it is shown in Figure 3 especially event 5 was recognised by all subjects on the touch-based interface but by only five subjects on the interface of the C2-System.

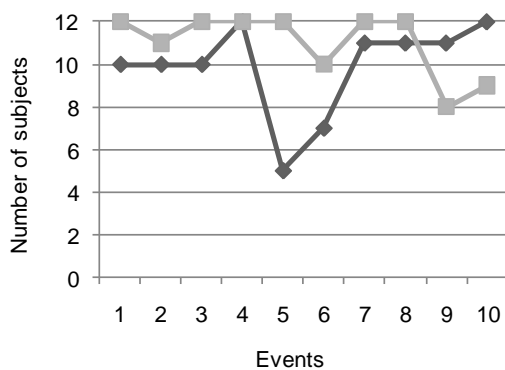


Figure 3. Number of subjects who recognised the respective event (dark grey line: user interface of C2-System, light grey line: touch based user interface)

Workload

For the analysis of workload a repeated measures ANOVA with the factors user interface and task (tactical picture compilation phase 1, phase 2 and

engagement) was conducted. The analysis revealed significant main effects for both factors ($F(1,11) = 23.37$, $p < .01$; $F(2,22) = 27.11$, $p < .001$) but no significant interaction. As figure 4 shows, the workload was rated higher for every task on the user interface of the C2-System. For both interfaces Phase 2 of the tactical picture compilation task led to the highest workload level.

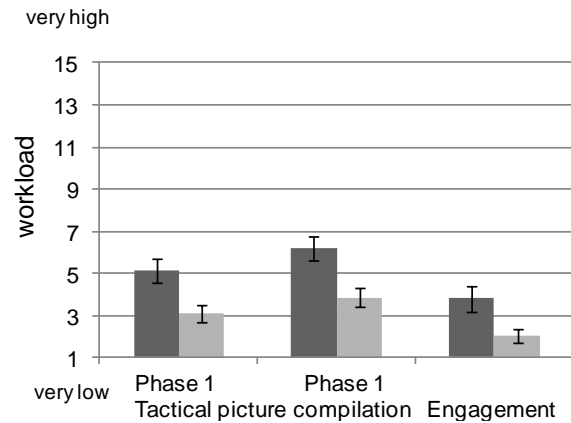


Figure 4. Mean workload levels (dark = user interface of C2-System, light = touch-based user interface).

Subjective Assessment

In the subjective assessment of the user interfaces subjects had to rate the interfaces according to some criteria which were essential for the tasks tactical picture compilation and engagement. At the end subjects were also asked to rate general aspects of the interfaces like the input technique and the visualisation of information. Figure 5 shows the mean values of the ratings summarised for the questions related to tactical picture compilation, engagement and the general rating. In all three categories the touch-based user interface was rated significantly more positive than the interface of the C2-System.

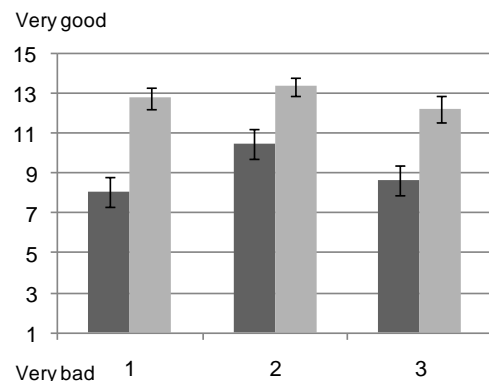


Figure 5. Mean values of the subjective rating (1=tactical picture compilation, 2=engagement, 3=general rating)

Discussion

The results of this evaluation reveal that the newly developed touch-based interface has some

advantages over the user interface of the current C2-System. Even though they do not differ much with respect to the task performance, it could be shown that the touch-based interface leads to a lower workload level and is rated more positively by the operators. It also seems to enhance situation awareness as can be especially seen in case of event 5 in phase 2 of the task tactical picture compilation. The event was that an air track left the airroute and was heading towards the own ship which was better detectable on the TDA of the touch-based interface.

For the interpretation of the results it has to be taken into account that some of the operators already knew the current C2-System but had to get to know the touch-based interface. However, when investigating if the experience of the operators has an effect on the outcome no significant results could be found. This could be due to the fact that other C2-Systems are similar in layout and functionality and operators could profit from their general knowledge of C2-systems.

TEST ON A MOVING PLATFORM

Finally the touch-based concept was tested on a moving platform. The purpose was to examine if a proper interaction per touch is still possible under moving conditions which can occur on naval vessels. In terms of a comparison, the test was also done with a magnetic mouse on the same interface.

Method

Due to limitation of time the test was only conducted by four subjects in a repeated-measures-design. In the test three independent variables were varied:

- Input technique (touch vs. mouse),
- Task (identification vs. engagement),
- Moving conditions (without movement, movements according to wind force 4, movements according to wind force 5).

Subjects had to absolve the test for each of the 12 possible factor combinations. First, the tests for the condition without movement were done, followed by wind force 4 and wind force 5. With respect to the input technique the order was changed between the subjects to control for sequence effects. After every condition the workload was measured by the questionnaire NASA-TLX. Subjects were also asked after every moving condition and task which input technique they preferred.

Results and Discussion

Since the participant group was quite small it was not intended to get effects of stastical relevance. However, a descriptive analysis gives at least some

insight to the practicability of the input techniques under moving conditions.

As figure 6 shows exemplary for the subtask selection, interaction times were quite similar for both input techniques under the condition of no movement and under the condition of wind force 4. Regarding wind force 5 outcomes are characterised by quite high inter-individual differences.

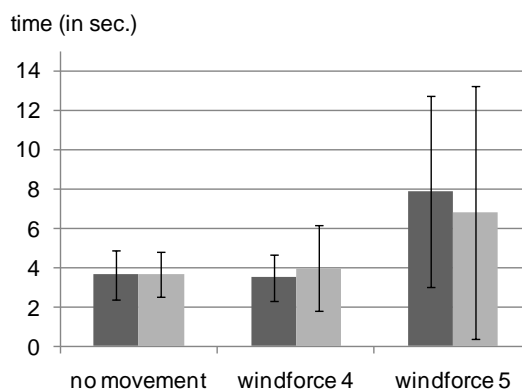


Figure 6. Selection times for touch input (dark) and mouse (light) on the three moving conditions.

In an analysis on an individual level for wind force 5 it could be seen that only one subject had considerably higher selection times. With respect to the task engagement all subjects interacted faster with the touch input. Results were mixed for the task identification.

The analysis of workload did not reveal big differences between the input techniques. But the analysis of the subjective preference provided an interesting result. Subjects were asked a total of six times which input technique they preferred. When aggregating the number of votes (in total 24), it turns out that touch input was named 12 times as the preferred input technique, mouse was named 3 times and no preference was stated 9 times.

All in all the results suggest that touch input is presumably a slightly better input type under moving conditions than the input with a magnetic computer mouse. However, it has to be considered that data of only four subjects are not representative and just show up a tendency which has to be confirmed with a bigger sample size.

CONCLUSION

In order to develop a new improved interface for Command & Control Systems three evaluations were conducted. The results indicate that a touch application can be a promising alternative to existing interfaces that use more conventional input techniques. However, further tests are required to assure that the touch-based interface also enables a proper and fast interaction under stress and in prolonged interaction periods.

REFERENCES

- Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, and J. Leplat (eds.), *New Technology And Human Error*, pp. 271–283: Wiley: New York, 1987.
- Durso, F.T. & Dattel, A.R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* Aldershot, UK: Ashgate pp. 137-154 2010
- Hart, S.G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (pp. 139-183).
- Witt, O., Schwarz, J. & Özyurt, E. (2010). Touchbasierte Benutzungsoberfläche für ein Marine-Einsatzsystem. In M. Grandt & A. Bauch (Hrsg.), *Innovative Interaktionstechnologien für Mensch-Maschine-Schnittstellen (DGLR-Bericht 2010-01, S. 133-145)*. Bonn: Deutsche Gesellschaft für Luft-und Raumfahrt e.V.