Usability Evaluation of Multi-Touch-Displays for TMA Controller Working Positions

Maria Uebbing-Rumke, Hejar Gürlük, Malte Jauer
Institute of Flight Guidance
DLR, German Aerospace Center
Braunschweig, Germany
Maria.Uebbing@dlr.de; Hejar.Guerluek@dlr.de; Malte-Levin.Jauer@dlr.de

Konrad Hagemann, Andreas Udovic
DFS, German Air Navigation Services
Langen, Germany
Konrad.Hagemann@dfs.de; Andreas.Udovic@dfs.de

***Abstract*—Since smartphones and tablets were introduced into consumer electronics, multi-touch technology has been increasingly used for professional purposes. The aim of our research within the SESAR project 10.10.2 task "Innovation Analysis 2013" was to determine whether controllers working in the approach controller position could also use gesture-based natural user interfaces for their work. We built a mock-up with multi-touch interaction that was evaluated against a comparable mouse interaction concept. In our study, fourteen air traffic controllers nearly all with active licenses were asked to guide air traffic in a realistic scenario with both the multi-touch and a mouse mock-up. Usability and workload were assessed. The results revealed higher usability scores for the multi-touch mock-up, and the workload is assessed to be decreasing in many aspects, too.**

Keywords- multi-touch; interaction design; human-computer-interaction; controller working position; usability; air traffic control; TMA; mouse

I. INTRODUCTION

In dense airspaces such as Frankfurt, approach controllers, especially feeders, have a heavy workload to achieve a compact sequence of aircraft on the centerline. Appropriate commands to the pilot need to be given timely optimized while observing separation rules. This entire process requires continuous communication over radiotelephony (R/T). In the future air navigation service providers will increase the effort for being able to use data link connections for transferring commands from controllers to pilots. To this end other interaction technologies will be used in controller working positions. We investigated the usability of multi-touch interaction compared to traditional mouse interaction. The mouse at controller working positions (CWP) is commonly used for navigation on the actual display and at several implementations for clearance tracing, too.

II. STATE OF THE ART

An innovation analysis was conducted by the SESAR WP 10.10.2 (CWP Human Factors Design) members at the beginning of the investigation. For this innovation analysis the literature was reviewed for statements assessing the usability aspects of multi-touch applications by a predefined set of criteria like used context, user groups, type of multi-touch technology, modes of interaction, system's level of maturity, ergonomic studies, and number of users in parallel [1]. Since the initial search revealed that studies aiming at a direct comparison of multi-touch with conventional technology in the context of air traffic control (ATC) were rather rare, a wide variety of sources also from other domains has been considered for the literature review. Cross-domain factors were extracted and reported.

The advantage seen by many authors is the potential for collaborative usage if the screen is big enough [2]. The ability to communicate over the HMI without using speech is given. Even more than two people can work together on one device. A tabletop device may increase the mutual awareness during collaborative decision making tasks. This kind of collaboration may be suitable for pre-tactical planning in the ATC environment. Multi-touch devices offer the chance to a more natural human computer interaction. A benefit may result from reduced selection time for direct-touch with one finger compared to an indirect selection with a mouse; bi-manual interactions can result even in faster triggered actions than using one finger [3]. Some authors also see the aspect of hedonistic user experience i.e. fun when executing his task [4]. No intermediate mechanical device as pen or mouse is needed. The simple construction of multi-touch devices (one device only) makes their use easy and fail-safe in many applications. The combination of multi-touch and tangible artifacts may offer a potential to support a natural way of interaction with digital information.

The core design recommendation extracted from the literature can be expressed as: The gestures have to be intuitive, i.e. easy to remember. Then you will get the positive effect of easy-learning and less need for instruction material.

There are few disadvantages figured out in several studies, e.g. the lack of preciseness in detecting the interaction touch or gesture. In addition, if the display is very large it may reduce supervision and reachability. Mouse interaction is assessed to be more precise and proposed to be always preferred if only single-point interaction is needed [3]. Target objects may be hidden by the hand. If there are several users sharing one screen it is difficult to share information between seated users as they might be restricted when trying to reach distant objects or interfere the action of their neighbor. It is stated that visual and haptic feedback is missing (e.g. virtual keyboard), so handling while looking somewhere else is complicate. A list of the considered research studies and more detailed extracted results can be found in the Innovation Analysis Report 2013 document [1].

In many implementations of CWP single touch input devices (TID) are already in operation, e.g. the paperless strip system (PSS) at the DFS. In PSS the controllers document their clearances by stylus input. The focus lies on providing quickly and easily reachable shortcuts for important functions.

III. RESEARCH PROTOTYPE

Our research meets the issue that studies could be hardly found that contained a direct comparison of multi-touch and conventional technology in the context of ATC.

The Institute of Flight Guidance in the German Aerospace Center (DLR) designed and implemented a CWP with multi-touch interaction philosophy [5]. A CWP with mouse input was designed and implemented for different purposes before. The application of the two mock-ups with different interaction philosophies can be one-to-one compared because of using the same features, so they only differ by the interaction technology and the number of screens belonging to the CWP, two for multi-touch and one for mouse. The main design principles for the multi-touch mock-up were intuitiveness and simplicity. Gestures for controller commands and support functions are intended to be easy to remember and their application follows intuitive movement directions. The CWP is designed for the approach control feeder position where the feeder is supported by an arrival management system (AMAN) [6] providing an optimized sequence proposal. Both mock-ups comprised a rather reduced set of accepted commands for the feeder position, e.g. heading and direct to waypoint commands were not available, all aircraft were flying on standard routes.

A. *Multi-touch Mock-up*

The multi-touch mock-up as depicted below in Figure 1 consists of a two-display installation, with one 30” sized situation data display (upper screen) serving merely as a traffic situation display and a multi-touch display (lower screen) serving as interaction display. The controller commands are given by single-touch respectively multi-touch gestures. In contrast to the mouse mock-up where the controllers are sitting on a chair in front of their display, they have to stand upright in front of the multi-touch display that is placed on a desk while guiding the traffic by gestures. Of course, the body posture in this mock-up cannot be considered to be optimal, but the ergonomically optimal integration of the monitor into a CWP was beyond the scope of this innovation analysis. The effect of the given commands can be seen on both the multi-touch and the upper situation data display. The components of the multi-touch display in the marked icons are briefly described here.

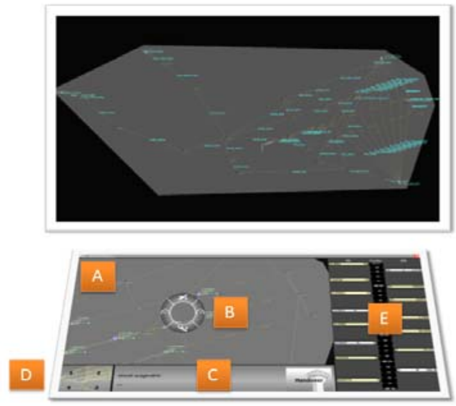

Figure 1: Multi-touch Mock-up

Multi-Touch Interaction Area (Icon A in Figure 1**):** This is the main area where the multi-touch interaction takes place. The interaction area only depicts a part of the whole traffic situation depending on which quadrant (see D. Quad Selection) the controller has pressed. The interaction area enables the controller to perform single- and multi-touch gestures for the assignment of several commands such as speed, altitude, turn-to-base etc.

Option Wheel (Icon B in Figure 1)**:** The option wheel (see Figure 2) depicts a graphical element, which indicates to the controller the possible actions that can be carried out, i.e. speed command by moving the finger on the horizontal plain as well as altitude command by moving the finger on the vertical plain. Whenever the controller touches an aircraft icon the option wheel pops up. As indicated by the Figure 2, the descending aircraft (arrow downwards) and climbing aircraft (arrow upwards) represent the possible altitude commands. Whenever the controller swipes his finger vertically upwards a higher flight level is commanded. The speed-indicator symbol to the left represents deceleration whereas a swipe gesture to the right represents acceleration.

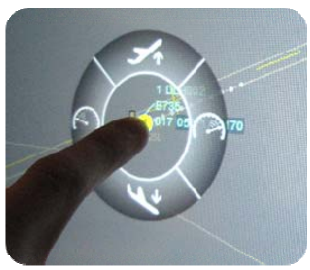

Figure 2: Option-wheel for possible controller commands

Aircraft Bar (Icon C in **Figure 1**)**:** The aircraft bar depicts the callsign of the currently selected aircraft. Besides that, this is the area where a handover to the tower position is carried out, by touching the handover button.

Quad Selection (Icon D in **Figure 1**)**:** The quad selection (see Figure 3) enables the controllers to choose between four defined areas of the given airspace thus they select the focus they are interested in by single-touching one of the four quadrants.

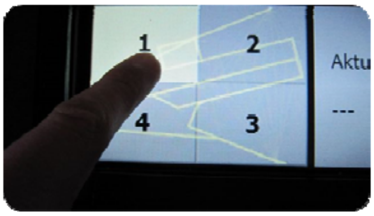

Figure 3: Quadrant selection for detailed airspace view

AMAN Timeline (Icon E in **Figure 1**)**:** The AMAN timeline displays the planned sequence of aircraft. The interaction with the AMAN timeline is limited to the highlighting of a label that will be colored red when touched within the timeline widget. The corresponding aircraft label displayed within the interaction area is colored in light blue.

During the multi-touch test sessions the controllers had to guide the aircraft by the following commands designed for multi-touch interaction:

Speed and altitude commands: The speed and altitude commands are designed as sliders that appear when the user initiates a left/right respective up/down swipe gesture starting on the option wheel. A deceleration is done with a swipe gesture to the left whereas acceleration is executed with a swipe gesture to the right. A flight level change is performed in the vertical plane, meaning a swipe gesture downwards corresponds to commanding a lower, and a swipe gesture upwards a higher flight level (see **Figure 4**).

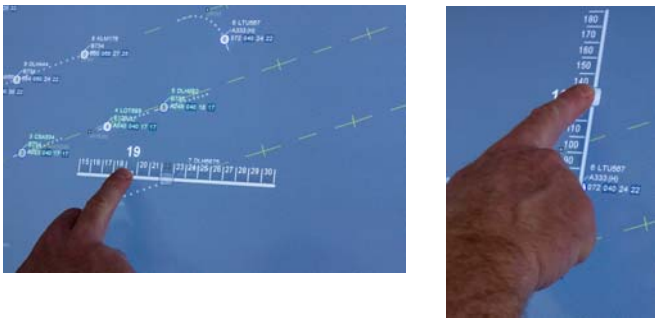

Figure 4: Speed and Altitude Commands

Turn to base command: The turn-to-base command may initiate a turn left or turn right. The complete execution of the gesture starts with evoking the option wheel by single touch with one finger. A second finger is used outside the wheel to make a turning wipe gesture to the left respective to the right. Turn indicator arrows appear showing the chosen turn direction (see **Figure 5**).

Figure 5: Turn-to-base Command

Handover to tower command: The handover to tower command is executed by simply selecting the corresponding aircraft and then touching the big handover button in the aircraft bar (see Figure 6).

Figure 6: Handover to Tower Command

Distance Measuring: As an additional feature the user can initiate distance measuring actions in order to monitor the separation between aircraft in nautical miles (see Figure 7). The distance measuring gesture may be executed optionally with one or two hands with simultaneous or consecutive touches. It is possible, to display several measuring events at once. The calculated distance is updated continuously.

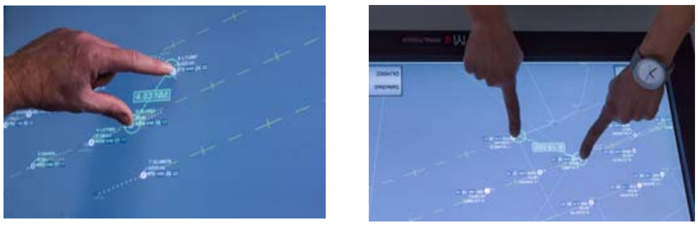

Figure 7: Distance Measuring with one and two Hands

B. *Mouse Mock-up*

The mouse mock-up functionalities are designed in analogy to the multi-touch mock-up. The controller uses the same commands for guiding and has the distance measuring feature for control purposes. The standard design elements known for mouse usage are part of this application: Popup and context menus defined by clicking different icons by right or left mouse click, scrolling of values by turning the mouse wheel etc. (see Figure 8).

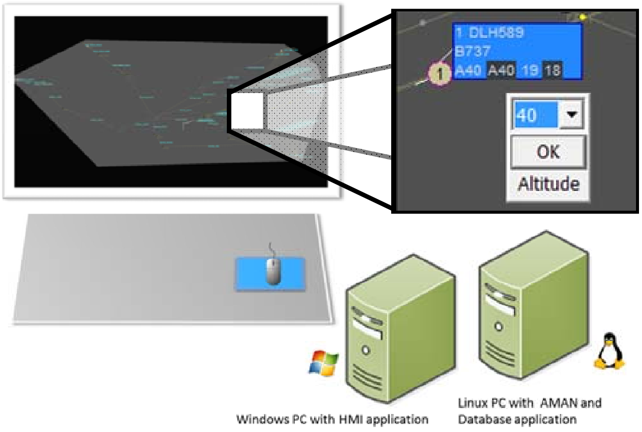

Figure 8: Mouse Mock-up with popup menu for descend command

IV. Usability Study

A. *Evaluation Concept*

For SESAR projects that would like to implement a multi-touch system in order to support the human system dialog this evaluation provides a systematic assessment using quantitative and qualitative data on the potential usability benefits for this technology within the TMA/ACC environment. The main objective of the study was to evaluate the usability, user experience, and workload of a new operator concept, realized by means of a multi-touch HMI. This was achieved by a hypotheses testing approach, whereby a multi-touch and a mouse-based operator concept for controller working positions were compared with each other. For the evaluation of the different operator concepts, realistic approach scenarios were used. Derived by the innovation analysis task description in SESAR the following research questions were answered by the study:

- Do controllers think multi-touch is a feasible technology for their working positions? How well is it accepted?
- How does the use of technology influence the workload and quality (i.e. safety and efficiency) of air traffic control operations (directly compared the identical scenarios with mouse and multi-touch technology)?
- What are the benefits and flaws of a multi-touch interface at the TMA/ACC CWP?
- Which guidelines can be derived from the data concerning the usage of multi-touch interfaces at the TMA/ACC CWP?

Usability is defined by DIN EN ISO 9241-10 [11]. The following general criteria are frequently associated with usability and used within the ATC context:

- Controller Performance
- Mental and physical workload
- Usability, which is divided into general usability and user experience

Reports from the literature (see [1]) indicate a variety of usability advantages for the usage of multi-touch HMIs. Subsequently, the following hypotheses were formulated for the evaluation:

Controller Performance

$H1_A$: Controller performance will be higher with the multi-touch HMI compared to the mouse-based HMI.

$H0_A$: There will be no difference in controller performance with the multi-touch HMI compared to the mouse-based HMI.

Measures: Overall Separation Accuracy, Total Number of Landings, Separation Violations, Overall rating

Mental and physical workload

$H1_B$: Mental workload will be lower when using the multi-touch HMI compared to the mouse-based HMI.

$H0_B$: There will be no difference in mental workload when using the multi-touch HMI compared to the mouse-based HMI.

Measures: NASA TLX questionnaire

$H1_C$: Physical workload will be higher when using the multi-touch HMI compared to the mouse-based HMI.

$H0_C$: There will be no difference in physical workload when using the multi-touch HMI compared to the mouse-based HMI.

Measures: NASA TLX questionnaire + 1 item of the DLR usability questionnaire

Usability

General usability

$H1_D$: General usability will be higher when using the multi-touch HMI compared to the mouse-based HMI.

$H0_D$: There will be no difference in general usability when using with the multi-touch HMI compared to the mouse-based HMI.

Measures: General Usability questionnaire, DLR usability of Mock-up Functions questionnaires, System Usability Scale (SUS) questionnaire

User Experience

$H1_E$: User experience will be higher when using the multi-touch HMI compared to the mouse-based HMI.

$H0_E$: There will be no difference in user experience when using with the multi-touch HMI compared to the mouse-based HMI.

Measures: User Experience Questionnaire (UEQ)

Each testing day contained debriefing sessions (semi-structured group interviews) during which the participants have reported their thoughts and experiences with the system. During the testing the participants were asked to think aloud, i.e. share their thoughts with the investigators while performing the task. The collected statements were overall analyzed, summarized and set into relation with the other measures

B. Technical Setup

During the evaluation, two different system configurations with their operator consoles were assessed in the same evaluation environment. The two system configurations comprised the following system components: Air traffic simulation, traffic situation data display (SDD), interaction device, AMAN decision support tool to advice sequence and runway assignment.

Scenario: The two system configurations were assessed in the same evaluation environment using a realistic Frankfurt approach scenario. The scenario showed a medium dense traffic in the Frankfurt airspace structure from 2010. There was one single reference scenario for both interaction devices in order to ensure that the results can be compared.

Controller Tasks: With both system configurations, the feeder controllers had to guide the aircraft to landing within a 25-minute trial. Their main task was to achieve a sequence on the centerline as compact as possible while complying with separation rules. For this task the following commands were used: Descend and reduce, turn to base and intercept final, runway allocation, and handover to tower.

Furthermore the controllers were asked during a trial to apply the distance measuring for several times. They were not forced to stick to the AMAN advices concerning sequence or runway. The AMAN recognized deviating guidance procedures and adapted to controller intentions. The transfer of speed or altitude commands to the traffic simulator could be traced by the controller when the entry in the label had changed. Runway allocation and turn-to-base processing had to be monitored through observing the radar track.

R/T was not used for the test runs. Based on future SESAR ATC concepts, it was assumed that there is a quick and reliable data link connection between the controller input system (mouse or multi-touch) and the aircraft. In the evaluation environment the simulated pilots complied with the commands without delays or further questions. The investigation focused on the evaluation of multi-touch as a possible interaction technique for future CWPs.

C. Metrics

Variables: Essentially, the independent variable of the study was the operator concept. The factor "operator concept" has two levels: mouse vs. multi-touch. The operator concept, thus, is regarded as a predictor variable which is expected to cause differences in the parameters. The dependent variables, in term, were the following below. To assess the quality of the simulation runs we tracked the following parameters:

Number of Landings: The number of aircraft guided to landing is evaluated. Similarly to the parameter separation accuracy, the parameter number of landings is included in the quality of the control task.

Separation Violations: A separation violation is detected when the distance between two aircraft falls below the ICAO separation with preceding aircraft at the threshold and its follower on the final.

Separation Accuracy: It determines the degree of deviation from the prescribed ICAO separation for aircraft of different weight turbulence categories when landing. In this context, the mean deviation from the standard separation is evaluated. This parameter can be rated as an indicator for the quality of the control task. As exceeded and undercut separations may even up each other, separation violations have to be considered adequately in the overall rating

Overall Rating: An overall rating indicator is calculated. It is an overall score for the quality of the simulation runs calculated from the previous parameters.

To assess general usability parameters additional to the quality of the controller task performance we evaluated usability and workload questionnaires:

Workload: The workload is assessed by the NASA-TLX workload index using the raw scales [9].

User Experience: It is assessed by the User Experience questionnaire [10], which besides user experience partly assesses also classical aspects of usability (efficiency, conformity of user expectations, etc.).

Usability: It is assessed both by the generic usability questionnaire System Usability Scale (SUS) [8] and a tailor-made usability questionnaire considering general aspects and the handling of specific functions of the DLR multi-touch and mouse mock-ups. These questionnaires were derived from the ISONORM questionnaire described in [11].

To assess statistical significance, we conducted hypothesis tests for paired samples. In case of approximately normal distributed samples, we performed the t-test, otherwise the non-parametrical Wilcoxon-test. P-values lower than 0.05 indicate a statistical significant difference between the two samples. The calculated significant values are provided in the results analysis. Observation protocols were used to register usability issues during the interaction and general statements about the HMI and interaction concept. Technical affinity was assessed as a control variable via a specific questionnaire [7].

D. Execution

Fourteen DFS air traffic controllers, two of them female, aged from 22 to 59 years, participated in the study. They were experienced approach or upper airspace controllers, thirteen of them bearing an active sector family license (5 for Frankfurt approach, 2 for Munich approach, and 6 for upper airspace from UAC Karlsruhe). At the beginning of the evaluation, participants were asked to fill out a socio-demographic questionnaire, including e.g. questions about their experience with digital media. Moreover, subjects were asked to sign an informed consent for the study. Both participants per day were jointly briefed by using a power point presentation that explained the objectives of the study, the setup, the tasks, and data measurements.

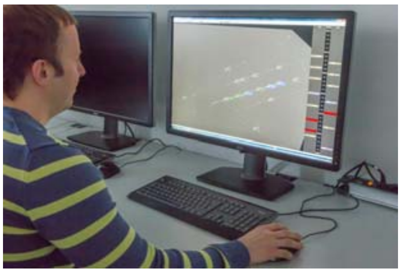

Figure 9: Test Person at Mouse Mock-up

The experiments took place from 29th of January to 6th of February 2014 at the Braunschweig DLR site. On seven days two controllers participated in the simulations with both mock-ups. The trials were conducted counterbalanced in order to avoid sequence effects. Thus, one participant started with mouse the other with multi-touch interaction trials.

Figure 10: Test Person at Multi-touch Mock-up

Before the simulation runs the participants were trained on the according interaction device for about 10 minutes. Immediately after the simulation runs of 25 minutes they had to fill in the questionnaires. Afterwards debriefing sessions took place. At the end of the day, a joint debriefing with both test persons was conducted.

E. *Results*

The following sections show in detail the results for the assessed parameters, explanations and interpretations followed by the according diagrams. More detailed explanations of the used measures can be found in the "Innovation Analysis Report 2013" document [1]. Section 8) contains controller statements from observation protocols and debriefing sessions.

1) *Control Variable Technical Affinity*

After the visual inspection of the individual scores it can be stated that the selected population showed very coherent scores on technical affinity as assessed by the technical affinity questionnaires at the beginning of the evaluation day [7]. The standard deviation (SD) depicted as the grey-dotted corridor was rather small (see **Figure 11**). Therefore, technical affinity was not considered as an influencing factor in the subsequent analysis.

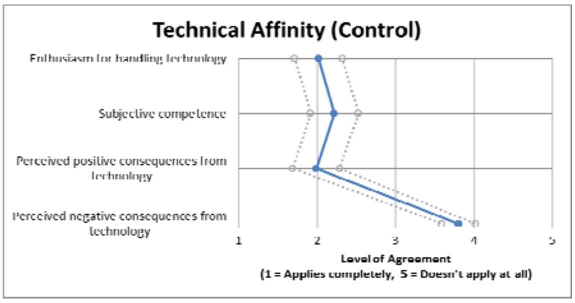

Figure 11: Control Variable Technical Affinity (dotted lines mean SD)

Overall, the data implies a rather open-minded participant population concerning technical innovations as reflected by the mean scores to technical enthusiasm and perceived positive consequences from new technology. Perceived negative consequences from new technology received the lowest agreement scores within the selected population.

2) *System Usability Scale*

For the assessment of the general usability from the SUS questionnaire an overall SUS score was calculated which can be interpreted as a percentage value: 100% would represent an ideal, perfect system without any usability problems, Scores ranging between 70% and 90% indicate an excellent usability. Scores between 50% and 70% can be interpreted as a mediocre to good usability, whereby scores lower than 50% indicate poor usability.

For the selected population of controllers the mouse mock-up obtained a SUS score [8] of 69.8% whereas the multi-touch mock-up received 85.4%. The scores differ with statistical significance after the Wilcoxon-test ($z = -2.45$; $p = 0.01^*$). According to the SUS interpretation scale (see Figure 12) the mean usability assessment of the mouse mock-up can be classified as good, although few participants evaluated the usability of the mouse as poor. However, overall assessment of the multi-touch mock-up usability obtained even an excellent usability result, demonstrating a more positive bias towards multi-touch input when compared to mouse input. Nevertheless, it has to be considered that both SUS scores reflect the usability evaluation of the examined HMIs with only limited functionalities, i.e. further investigation is necessary whether a larger and more complex set of possible commands will still yield the same results.

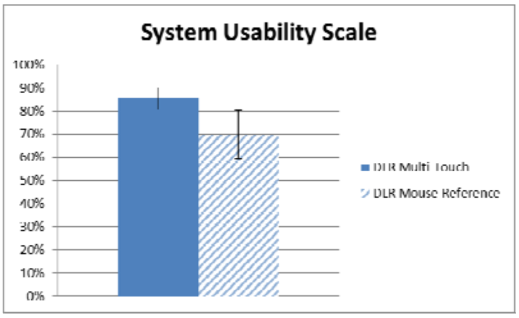

Figure 12: System Usability Scale

3) User Experience Questionnaire

The analysis of the user experience questionnaire (see Figure 13) revealed average values at a slightly above medium rating level. However, no significant effect between mock-up conditions multi-touch vs. mouse reference could be observed. The variable novelty just failed to reach significance in the statistical evaluation (t = 2.16, p = 0.05). This may somehow reflect the impression that multi-touch has reached an “everyday usage level” among most of the participants’ estimation.

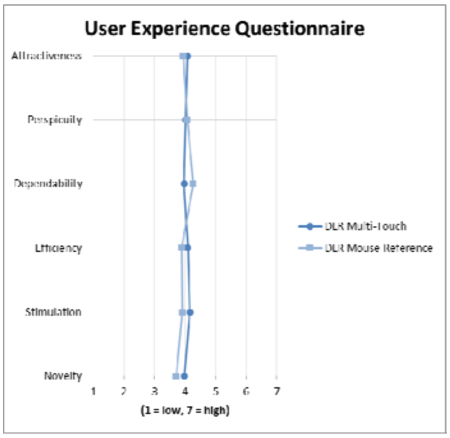

Figure 13: User Experience Questionnaire

4) General Usability Questionnaire

The mean rating results of the general usability questionnaire are displayed in Figure 14. As described in section III.A the participants had to stand upright in front of the multi-touch mock-up. This gives an explanation for the poor rating for a relaxed working posture (significantly poorer; t-test: z=-2.45; p=.01*). As zooming and panning were not possible at that mock-up the rating of individualization is quite poor, too. For all other usability criteria the multi-touch interaction HMI was rated better than the mouse mock-up.

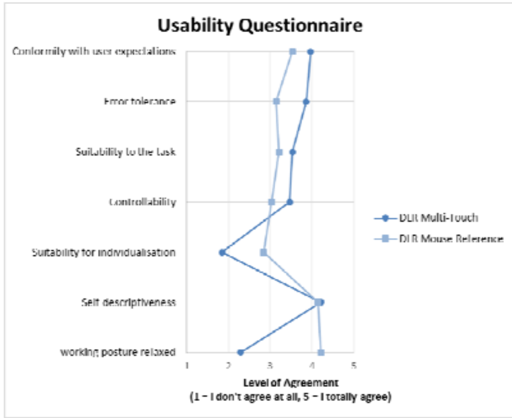

Figure 14: General Usability Questionnaire

5) DLR Usabilty of Mock-up Functions Questionnaire

The assessment of the usability of the DLR specific mock-up functions was carried out with a tailor-made questionnaire. The overall results revealed a clear positive bias towards the usage of multi-touch (see **Figure 15**). This was especially the case for the ratings regarding the execution of turn-commands (significant after Wilcoxon z = -2.44; P = 0.01*) or distance measurements (z = -2.84; P = 0.00**) between two aircraft. Although the usability of the speed and altitude commands was also viewed better for multi-touch, the difference among both input modalities seemed not to be that big. This could be due to the stated problems of the touch accuracy, especially when sliding the finger in the vertical plane (altitude slider). Another interesting point is that the two-finger gestures (turn-to base, distance measuring of two aircraft) obtained high usability ratings and showed the biggest differences in relation to the ratings for the mouse. A possible explanation for that could be that necessary multiple inputs can be performed faster and in a more comfortable manner by multi-touch than by mouse.

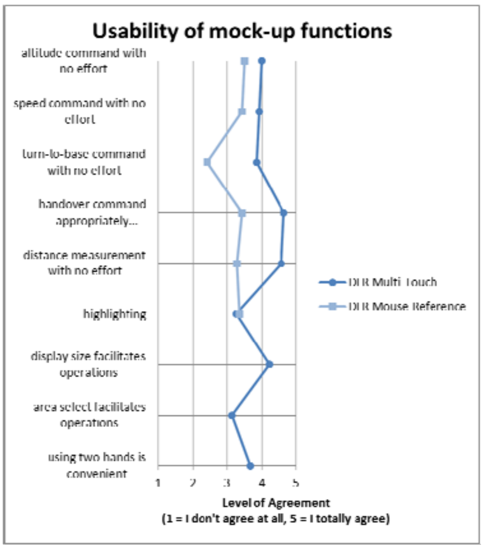

Figure 15: Usability of Mock-up Functions

Furthermore, the handover to the tower showed a significant (z = -2.85; p = 0.00**) advantage of the multi-touch, probably caused by the simplicity of the multi-touch input (one finger-tap on the handover-button) in comparison to the mouse context menu.

6) NASA TLX Workload Assessment

The analysis of the NASA TLX questionnaire [9] revealed lower scores for the various task load scales for the multi-touch mock-up compared to the mouse reference mock-up, except for the physical demand (see Figure 16). Participants evaluated physical demand to be higher with the multi-touch mock-up, although the difference failed to reach significance after the t-test (t = 0.786; p = 0.446). We consider this to be primary due to the standing position in comparison to a sitting position while working with the mouse mock-up. As a possible result of the reduced menu depth (inexistent context menus) and intuitive design of the application, the test persons perceived the multi-touch device as significantly less straining concerning the mental demand and effort of the tasks (Wilcoxon-test z = -2.3; p = 0.02* respectively z = -2.3; p = 0.02*). These factors might have also influenced the higher rating for the frustration level, which directly comes along with the better perceived error tolerance and conformity of user expectations as stated in the usability questionnaire (see above). Another significant (t-test: t = 0.786; p = 0.446) difference shows up on the temporal demand. A reason could be the faster interaction with direct input, for example when commanding a turn or measuring the distance as reported by several participants in the debriefing interviews.

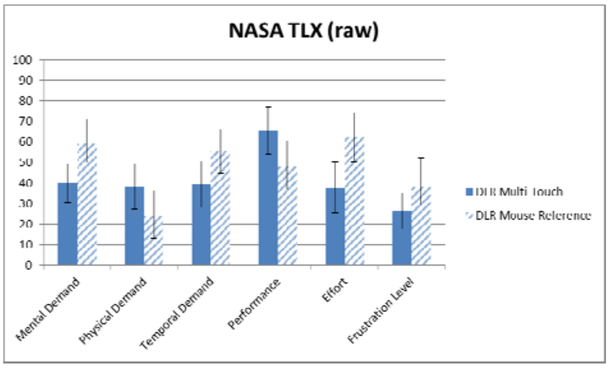

Figure 16: NASA TLX Workload Indicators

Although the test persons rated the performance of the multi-touch mock-up slightly better than the mouse reference; this does not correspond to the results of the controller task performance analysis (see Figure 17).

7) Quality of Controller Performance

Since the controller performance data was obtained from various heterogeneous parameters, which are therefore not easy to compare, we decided to scale the values to a range between zero and one. We assumed that the best possible guidance will pilot a maximum number of aircraft with minimal separation from the ICAO standard differentiation to the landing in the given time; while maintaining the wake vortex separations. Hence we extracted three main parameters: average separation deviation from ICAO, separation violations and number of touchdowns. Although separation deviation and separation violations imply a lower score to be better, the values are scaled in a way to correspond zero as the worst and one as the best recorded value of all test persons (on both interaction devices). Finally we calculated an overall performance rating from the three parameters, which is a weighted average with a weighting of 20% for average separation deviation, 50% for separation violations, and 30% for the number of touchdowns, which is a major indicator for the efficiency of the guidance.

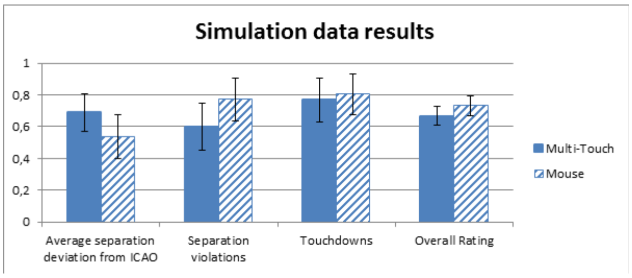

Figure 17: Simulation Data Assessment for Controller Performance

As to see in Figure 17, the overall rating for both devices is very similar. Of course, the limited set of controller commands make the results barely comparable to reality, but comparable between both mock-ups. The overall rating of the mouse mock-up is slightly higher, but with no statistical significance (Wilcoxon-test: z = -1.60, p = 0.11). When comparing the different parameters, a more dense guidance of the aircraft with the multi-touch device can be observed due to the better score in the average separation. However, this might directly depend on the separation violations; more frequent separation violations with the multi-touch system obviously yielded a lower separation average score. This is the reason, why the separation violation score is weighted at 50 percent; to prevent a balancing of those parameters, hence separation violations are unacceptable. A possible reason for the more (sometimes too) dense guidance can be the ease of use of the multi-touch gestures, as well as the smaller amount of time a person needs to give a command via gestures. Furthermore the tasks of tower controllers were not simulated, so that late corrections as they are possible in reality were not processed.

8) Controller Ratings

The most prominent statements from the debriefing sessions are listed below, sorted by statements concerning the trial setup and multi-touch mock-up assessment:

Trial setup:

- The test scenario was well set up with medium dense traffic

- Mock-ups hardly imaginable with real traffic, because too few functionalities are implemented
- Data link deployment with reasonable response times is far in the future

The following section summarizes the positive and negative key statements of multi-touch mock-up assessment.
Pros:

- Gestures are innovative and intuitive
- The infrared technology allows comfortable and smooth input with the fingers
- The abortion of an action and restart of a new action was very fast
- Not many unintended actions occurred
- Multi-touch is intuitive and a fast manner to interact with
- Direct touch philosophy is easy to apply
- Low menu depth with multi-touch leads to faster actions
- Analogies used for the implementation of commands as for instance the "turn-to-base command" (e.g. right-turn, rotating the finger to the right) are more easily understood and memorized
- Multi-touch is safer because of being easier to use

Cons:

- Touch accuracy is sometimes a problem when selecting moving objects
- Information on screen was sometimes covered by menus or too big icons (head symbol) or by hands
- Some controllers did not like a two display concept
- Infra-red based multi-touch sometimes produced unwanted ghost touches
- Ergonomic problems could arise when working for a longer period of time
- Some controllers wonder whether it is possible to handle a big airspace like en-route with direct operations

V. CONCLUSIONS

A. *Assessment of Experiment Hypothesis*

In the end summarized conclusions are provided on the results by referring to the initial hypotheses for the usability evaluation written down in section IV.A.

The data implies that compared to traditional mouse operation using a multi-touch input device in an ATC approach scenario may offer a suitable support to the operator by freeing scarce mental resources as shown by lower scores on the NASA TLX scales "mental workload" and "effort" (hypotheses $H1_B$ confirmed). Moreover, a tendency towards less temporal demand with multi-touch could be observed in the NASA TLX data speaking for the intuitiveness of the input device.

As the participants had to stand in front of the mock-up looking downward at the multi-touch display in contrast to the mouse mock-up where the participants sat in front of a desktop display, the physical demand (e.g. for the neck) was reported to be higher with the multi-touch device. This was according to our expectations ($H1_C$ confirmed).

With regard to data on the general usability, the current developmental stage of the multi-touch mock-up seems to be well accepted in general as reflected by the score "excellent" on the system usability scale and by the ratings of the DLR questionnaire ($H1_D$ confirmed). However, there were also a few exceptions (e.g. the possibility to adjust the HMI to one's own needs such as area selection on the radar screen etc.). Certain touch gestures such as the "turn-to-base", "distance measuring" and "highlighting" functions were clearly favored in comparison to similar functions in the mouse reference system. The data inspection of the user experience questionnaire showed average scores on a medium level, but no clear difference between system types ($H1_E$ rejected). This may reflect that nowadays multi-touch technology has become a widely used and accepted technology in consumer products as well as in professional domains.

The simulation data pointed to a slightly different direction. As stated in section IV.E.7), the choice of interaction technology does not seem to influence the controller performance in neither direction, i.e. positively nor negatively (H1A rejected). This indicates that there were no severe usability issues in operating the multi-touch mock-up in comparison to a reference mouse-based system.

B. *Benefits*

The overall investigation comparing multi-touch vs. mouse revealed a trend that further developing of controller working positions with multi-touch interaction philosophy is worth to follow up. The participants encouraged the developers to move further on with the prototype. The usage of multi-touch technology in the given context of the experimentation (mock-up) is found to be:

- no show-stopper due to safety issues
- imaginable at the working position
- error tolerant
- quick
- efficient
- not much influencing the performance of the controller (even at this early stage)

The acceptance will be influenced by the appropriate design and integration of further functionalities and the installation into well-designed furniture that allows sitting as well as standing body posture while working.

VI. Outlook

With this study a first step is done to assess the interaction technique multi-touch for the ATC environment. The results are promising, and encourage us to keep on working in the future with this new technology. Further investigations should be extended by the following aspects to get even closer to controllers' every day work.

- Implementation of complex commands that have to be issued consecutively within one single command (e.g. speed, then heading, then ILS clearance)
- Extension of the implemented design by adding intuitive gestures to invoke complex and additional commands
- Adequate solutions regarding ergonomic issues as for instance body posture of the controller as well as optimal integration of a multi-touch display into working table
- Use navigation gestures known from consumer devices for zooming and panning
- Optimal HMI element design regarding its size, in order to avoid coverage of the traffic situation on the one hand and touch accuracy problems on the other hand
- Consider the integration of other input modalities like speech, as R/T communication will stay an indispensable part of CWPs in future ATC concepts

VII. Disclaimer

References

[1] A. Labreuil, D. Bellopede, M. Poiger, K. Hagemann, M. Uebbing-Rumke, H. Gürlük, M. Jauer, V. Sánchez and F. Cuenca, "Innovation Analysis Report 2013," Deliverable D93 SESAR WP 10.10.2, 2014.

[2] S. Conversy, H. Gaspard-Boulinc, S. Chatty, S. Valés, C. Dupré and C. Ollagnon, "Supporting Air Traffic Control Collaboration with a TableTop System", in CSCW 2011 (Hangzhou, March 2011), ACM Press, pp. 425-434.

[3] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan, "Direct-Touch vs. Mouse Input for Tabletop Displays", in Proceedings of CHI 2007 (San Jose CA, May 2007), ACM Press, pp. 647-658.

[4] P.K. Sudarshan, R. Bruder and A. Ray "Joy of use in automative touch screen UI for the driver", in Proceedings of Automotive UI 2010 (Pittsburgh PA, November 2010).

[5] H. Gürlük and M. Uebbing-Rumke,"Operational Concept Multi-Touch HMI for Feeder Controllers", German Title: "Bedienkonzept Multi-Touch HMI für Feeder Lotsen," DLR Internal Report 112-2013/24, 2013.

[6] H. Helmke, L. Christoffels, R. Hann, H. Sievert, M. Temme, M. Uebbing-Rumke, "FAGI Modules of 4D-CARMA." DLR-Internal Report. 112-2010/22, 2010.

[7] K. Karrer, C. Glaser, C. Clemens and C. Bruder, "Technikaffinität erfassen – der Fragebogen TA-EG", in A. Lichtenstein, C. Stößel, C. Clemens (Hrsg.), Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme (ZMMS Spektrum 2009, Reihe 22, Nr. 29, pp. 196-201). Düsseldorf: VDI Verlag GmbH.

[8] System Usability Scale, SUS, © Digital Equipment Corporation, 1986.

[9] S.G. Hart and L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). Amsterdam: North-Holland, 1988.

[10] B. Laugwitz, T. Held and M. Schrepp, "User Experience Questionnaire", in A. Holzinger (Ed.): USAB 2008, LNCS 5298. Heidelberg: Springer-Verlag.

[11] J. Prümper, "Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität" in: R. Liskowsky u.a. (Ed.): Software-Ergonomie '1997, Stuttgart.