TriControl – A Multimodal Air Traffic Controller Working Position

Oliver Ohneiser, Malte-Levin Jauer, Hejar Gürlük, Maria Uebbing-Rumke
Institute of Flight Guidance, German Aerospace Center (DLR)
Lilienthalplatz 7, 38108 Braunschweig, Germany
{Oliver.Ohneiser;Malte-Levin.Jauer;Hejar.Guerluek;Maria.Uebbing}@DLR.de

***Abstract*—The TriControl multimodal controller working position (CWP) demonstrates a novel concept for natural human-computer interaction in Air Traffic Control (ATC) by integrating speech recognition, eye tracking and multi-touch sensing. All three parts of a controller command – aircraft identifier, command type and value – are inserted by the controllers via different modalities in parallel. The combination of natural gazes at aircraft radar labels, simple multi-touch gestures, and utterances of equivalent values are sufficient to initiate commands to be sent to pilots. This reduces both controller workload and the time needed to initiate controller commands. The concept promises easy, well-adjusted, and intuitive human-computer interaction.**

Keywords—Air Traffic Controller; Human Machine Interaction; Multimodality; Eye Tracking; Automatic Speech Recognition; Multi-touch Gestures; Controller Command; Workload Reduction

I. INTRODUCTION

Current human machine interfaces (HMI) of air traffic controllers mainly focus on the "speech" modality when communicating with pilots. Data link-based communication, wherever available, is generally initiated by mouse or pen input. Controllers usually use mouse and keyboard as interaction devices for keeping system information up-to-date. Multimodal HMIs emphasize the use of richer and more natural ways of interaction by combining different modalities, such as speech, gestures, and gaze. Therefore, they need to interpret information from various sensors and communication channels.

Multimodal systems have the potential to enhance human-computer interaction (HCI) in a number of ways by:

- adapting to a wider range of users, tasks, and situations,
- providing alternative methods for user interaction,
- conveying information via the appropriate communication channel,
- accommodating differences between individual operators by permitting flexible use of input modes,
- improving error avoidance, and
- supporting improved efficiency through faster task completion, especially when working with graphical information.

When people communicate with each other in person they have eye contact, use their hands for gestures and emphasis, and voice for content regarding "facts". Multimodal HMIs represent a new class of user-machine interfaces, applying the same principles from human interaction to human-computer interaction. It is anticipated that they will offer faster, easier, more natural and intuitive methods for data entry.

This capability is a prerequisite for advancing human-machine systems to the point where computers and humans can truly act as a team. Furthermore, air traffic research and development programs like SESAR (Single European Sky ATM (Air Traffic Management) Research Programme) require use and integration of new technologies such as touch- and speech applications for an enhanced controller-system interaction [1]. An efficient way of entering data into the system is also required to enable a beneficial data link application.

The primary goal of TriControl, the DLR demonstrator for a multimodal CWP in the ATC approach area, is to ensure that a human operator can enter data e.g. controller commands more quickly and intuitively. Based upon empirical findings and subjective evaluations we assessed the suitability of different input modes in relation to specific command elements.

To outline the scientific context, chapter II of this paper presents related work on different interaction modalities. Chapter III includes the concept and implementation of the TriControl prototype comprising eye tracking (ET), speech recognition (SR), and multi-touch (MT) gestures. A preliminary evaluation of the implemented system and results of that evaluation are outlined and discussed in chapter IV. Finally, chapter V draws conclusions and identifies future work.

II. Related Work on Multimodal Human Machine Interaction

Implementation of multimodal human-computer interaction concepts is still at an early stage in ATC. Nevertheless, different prototypes using modern interaction technologies as single interaction modalities have been developed.

In fact, any interaction modalities that can be digitally recognized by a computer are conceivable for interaction with the system. Within the SESAR work package 10.10.02 technology screening was carried out in order to assess the suitability of current interaction technologies for controller working positions. The multi-touch, eye tracking and handwriting recognition technologies were investigated [2] on the basis of the screening results. Within this research, the technologies were analyzed and prototypes were evaluated. Consolidated assessments were carried out, particularly for multi-touch and eye tracking. Speech recognition has been substantially developed and evaluated in the AcListant® [3] project.

The most promising interaction technologies currently assumed as being suitable for input are multi-touch (haptic modality), eye tracking (visual modality) and speech (auditive modality). DLR has already successfully evaluated implementations in the field of eye tracking [4], multi-touch [5], and speech recognition [6].

A. Eye tracking (ET)

Eye tracking technology offers at least two different opportunities for use in ATC. Firstly, it has been used to assess mental workload [7] and fatigue of controllers. Secondly, there are a number of other reasons for incorporating eye tracking as an input device for controllers [8]: It allows hand-free interaction and facilitates the manipulation of radar labels or electronic flight strips. Another argument in favor of eye tracking is that eye movements are fast and natural. For instance faster selection times were reported with eye–gaze interaction than with other input devices such as the mouse [9]. According to [8] there is empirical evidence that eye trackers can become an efficient pointing device that can be used instead of the mouse [10] or the keyboard [11].

B. Multi-touch (MT)

By scanning the use of multi-touch technology in the ATC area a prototypic implementation of a workstation – announced as "Indra advanced controller working position" – can be found on Indra's website [12]. Besides stating that "Multi-touch technology is used routinely as a means of interaction" in this CWP, it does not provide a more specific description of what information is gained by MT input or indicate how it is subsequently used.

A master thesis [13] supervised by the German air navigation service provider DFS (DFS Deutsche Flugsicherung GmbH) contains a concept and first application of multi-touch for command input, later to be translated using text-to-speech technology and then sent to the pilots.

This was evaluated using DFS controllers and generated positive feedback on MT usability. The thesis also outlines expectations on deployment in CWPs in the near future.

DLR and DFS collaborated in the SESAR 1 Work package 10.10.02, dealing with ergonomics, hardware and design of the controller working position. Effort was expended in investigating the usability of multi-touch technology at TMA (Terminal Manoeuvring Area) and ACC (Area Control Center) CWPs. The DLR demonstrator with multi-touch interaction was evaluated against a comparable CWP with a mouse interaction concept [5]. In this study fourteen DFS air traffic controllers, aged from 23 to the fifties, were asked to guide approach traffic in a realistic scenario using both the multi-touch and mouse CWP.

Usability (see Figure 1) and workload were assessed. The results revealed higher usability scores for multi-touch technology. Mental effort and task effort were perceived as less of a strain.

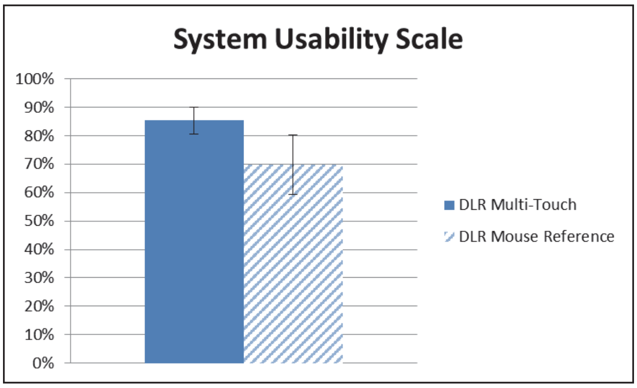

Figure 1. Overall system usability scale (SUS) [14] score multi-touch and mouse reference

The overall investigation indicated that it is likely to be worthwhile to continue developing controller working positions with multi-touch interaction philosophy. The use of multi-touch technology in an experimental context was found:

- not to be a show-stopper due to safety issues,
- to be conceivable at the working position,
- to be error tolerant,
- to be fast and efficient, and
- not to greatly influence controller performance.

The participants therefore encouraged the developers to continue developing the demonstrator.

C. Speech Recognition (SR)

Automatic speech recognition algorithms are capable of converting spoken words or numbers to text. Popular consumer products are for example Siri® [15] or Google's search by voice [16]. The first steps in integrating speech recognition in ATM systems, including ATC training, took place as much as a quarter century ago [17].

SR may also be used to replace pseudo-pilots in ATC simulation environments [18]. The readback of simulated pilots communicating with controllers can be fulfilled by recognizing controllers' utterances (speech-to-text) and repeating the command in correct phraseology again (text-to-speech). Context knowledge of what utterances are most likely in the current air traffic situation makes it possible to improve speech recognition quality [19].

SR can also detect complete controller commands from ATC vocabulary with acceptable reliability. The knowledge of the spoken commands can be used to support different tasks of controllers (e.g. aircraft radar label maintenance of approach controllers via direct automatic input of controllers' uttered clearances) at their working positions [20]. An approach controller is responsible for merging several streams of air traffic into a single final sequence for specific runways. Highly automated decision support tools such as arrival or departure managers have been developed to support human operators in this challenging task. These systems need to adapt to the controller's intentions by providing support for next recommended clearances. Hence, these systems require knowledge of – and input from – their human operators such as given clearances.

Normally, manual input from controllers is necessary. SR can perform the input task automatically by analyzing the radio telephony channel between controller and pilot. The controller only has to check the correctness. This kind of procedure leads to less workload [20].

D. Multimodality

Although the definitions of multimodality differ greatly in literature, there is general consensus that multimodal systems involve multiple senses of the operating human or multiple sensors of the corresponding machine. For example, the European Telecommunications Standards Institute (ETSI) defines the term "multimodal" as an "adjective that indicates that at least one of the directions of a two-way communication uses two sensory modalities (vision, touch, hearing, olfaction, speech, gestures, etc.)" [21].

A multimodal Thales demonstrator called "Shape" already includes eye tracking, a multi-touch device, and voice recognition [22], [23]. However, for example, speech recognition is only used to detect flight numbers. Only controller commands given through the tactile surface as a whole seem to be uplinked to the pilot.

Within a bachelor thesis [24] at DLR the first approach for a multimodal ATC demonstrator was undertaken to integrate the three modalities eye tracking, multi-touch and speech recognition. The main aim of this thesis was to implement and evaluate eye tracking as an input modality for a multi-modal controller working position. For this purpose an existing concept consisting of multi-touch and speech recognition [25] was enhanced by integrating eye tracking in order to enable natural and fast selection of aircraft.

The findings gained from the investigations carried out for that thesis showed that eye tracking is a useful and well accepted input modality when it is accompanied by other intuitive modalities.

However, those modalities should go beyond just serving as a "pointing device" for elements on a screen like many eye tracking applications.

III. Concept of Multimodal Air Traffic Controller Interaction

The motivation for building a prototypic multimodal CWP for approach controllers is based on presumed advantages of multimodal interaction (see chapter I) and promising research results gained from the previously developed unimodal and multimodal prototypes (see chapter II).

TriControl focuses on integrating the three most promising interaction technologies: speech recognition (SR), sensing of multi-touch gestures (MT), and eye tracking (ET) (see [26]). However, these modalities can be combined in a number of ways with respect to the three basic elements of a controller command (aircraft identifier (A), command type (T), and command value (V)). Furthermore, in former investigations some modalities were found to be more suitable for certain standardized command parts than others (Figure 2).

Modality	Aircraft	Type	Value
Speech Recognition	medium	medium	good
Multi-Touch	medium	good	medium
Eye Tracking	good	poor	poor

Figure 2. Matrix with suitability assessment of input modes (SR, MT, ET) with respect to controller command elements aircraft (A), command type (T) and value (V)

Figure 2 shows the favored assignment between input modality and command element. To identify the aircraft (A) that will receive the next command (e.g. DLH123) three possible ways are explained: uttering the callsign (A-SR), touching on its radar target/label on the situation representation or an auxiliary touch display (A-MT), or looking at the radar target/label (A-ET). The command might be transferred to pilots by data link or a text-to-speech interface.

Speech recognition rates of callsigns are good, but it takes some time to utter the whole callsign. Although in previous investigations the direct touch on an aircraft representation was assessed as easy and intuitive, the hand covers the radar screen and hence the traffic situation below. To guarantee a good overview of the whole traffic situation use of a second screen could solve the problem but would create a new issue in that the active gaze has to switch from one screen to the other and back.

However, the controller normally looks at the intended aircraft anyway. Hence, eye tracking seems to be the most convenient option for selecting the first controller command part, just as naturally as one usually makes eye contact in a face-to-face conversation.

Analogue to the aircraft identifier, the three input modalities are also discussed for the command type. SR (T-SR) would recognize International Civil Aviation Organization (ICAO) phraseology conform command types (T) quite well due to the limited search space of different types (e.g. *reduce*, *descend*, etc.). Selecting a type by eye movement (T-ET) – for example from a menu – will be tiring for the human operator as it requires unnatural and active control of gaze. In a human conversation, hands are also used to describe a general direction via gestures. Similarly, a multi-touch device can be used to draw a simple one- or more-finger gesture that is recognized very accurately to code a command type (T-MT).

Three different modalities also exist for entering command values. Selecting exact command values (V) (e.g. 210) with swiping gestures (V-MT), for example, on a visual scale can be difficult. Looking at values in certain menus (V-ET) is as exhausting as selecting command types with one's eyes. However, just uttering the short values works fast and is intuitive (V-SR).

From former investigations we derived a classification of the input mode suitability (poor-medium-good) in terms of a color-coded matrix (see Figure 2). This matrix depicts an initial point for implementation of TriControl (A-ET, T-MT, V-SR).

The chosen combination of modalities for TriControl enables the input of the three most common elements of a controller command (aircraft identifier, command type and value). The number "3" is spoken as "tri" (pronounced as "tree" in English) in radiotelephony to improve the understanding of digits even in bad speech quality, hence the name TriControl was used for the interaction design.

To generate a controller command, the operator has to focus on an aircraft on the radar situation display with his eyes, make a specific gesture on the multi-touch device, and utter a corresponding value for this type of command (see Figure 3 for the setup of modalities and Figure 4 for an example command).

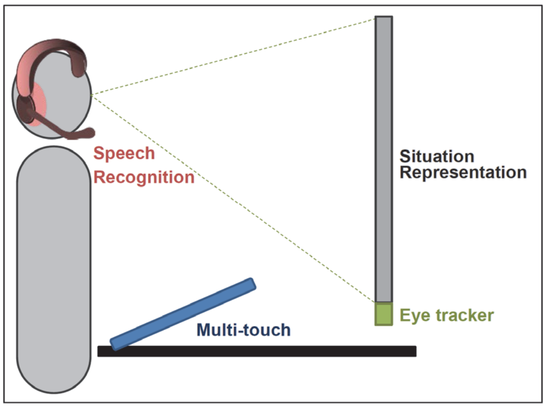

Figure 3. Interaction modalities of TriControl CWP

The information processed by the TriControl CWP is put together as a clearance shown in the far right box of Figure 4.

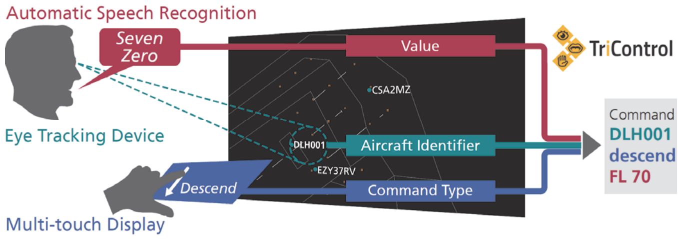

Figure 4. Schematic view of the communication modes and processed information

Our instantiated controller working position may be used by a feeder or pickup approach controller. One of the main tasks of approach controllers is monitoring the radar situation display. Within TriControl it is assumed that the aircraft radar label being looked at by the controller is the focus of attention. Eye gaze measurement is used to continuously calculate the position of the air traffic controller's eye gaze and correlate it with aircraft label positions on the display.

For our demonstrator we used DLR's radar screen software RadarVision [27] showing Düsseldorf airspace. It provides data about the position of aircraft icons, labels, and significant points (runway threshold, initial approach fixes, and waypoints). In TriControl, eye tracking enables aircraft radar labels to be selected as the first part of a command without losing focus of the traffic situation.

In our demonstrator we use a contact-free Tobii [28] infrared eye tracking device (Tobii EyeX Controller) which is mounted at the bottom of a monitor. Calibration is necessary prior to adapting eye tracking quality to people with contact lenses, glasses, or without corrected vision. The position of the user's pupils is followed regarding a standardized screen position set and connected to the display size by the manufacturer software.

Using the resulting display coordinates of the spot being looked at by the user in front of the display, we determine whether an aircraft icon or radar label is displayed. A dwell time of nearly one second was defined as the threshold for highlighting the currently focused aircraft label with a white frame. Otherwise, the controller could be distracted if the highlighting frame jumps around the whole screen, thereby indicating non-intended gazes, particularly while scanning the traffic situation.

Although, radar labels do hardly overlap in the TMA due to lateral aircraft and therefore label separation, manual or automatic deconflicting is possible to select intended aircraft safely via eye tracking. As a safety feature, the controller might fall back to selecting the aircraft callsign from a list on a multi-touch device.

In combination with two-dimensional gestures on a multi-touch display, the controller can add the type of a command to the selected aircraft to start insertion of a clearance.

The controller selects the type using a set of four single- and dual-touch gestures on a tablet – altitude, speed, or heading of the aircraft for example. The direction of the gestures indicates whether the aircraft should, for example, accelerate or decelerate.

Specifically to avoid head-down times the gestures and tablet usage are designed simply and intuitively. Furthermore, the user may perform all gestures at any location on the multi-touch screen while still focusing on the situation representation. We used a standard Wacom [29] multi-touch tablet in our demonstrator.

For the design of specific gestures typical natural gestures and well-known gestures from smartphone use were analyzed and assessed for the use in ATC. So, a one-finger swipe from left to right is recognized as *increase*, the opposite direction is recognized as *reduce*. A one-finger swipe from top to bottom indicates a *descend*, the opposite direction a *climb*. One finger held for more than one second pressed on any point of the multi-touch device is interpreted as a *direct-to* gesture. This gesture is also used for ILS clearance and handover to the next following responsible controller position, but requires different speech input compared to waypoints. Drawing a sector of a circle to the left or right with a two-finger gesture – either using two fingers of either hand – initiates a *heading* command. The multi-touch software evaluates the controller's gesture that results in "*reduce / increase [or-more / or-less], descend / climb [or-above / or-below], turn-right-heading / turn-left-heading, direct-to, handover, cleared-ILS, intercept-localizer*". Thus, the controller inserts the second part of his command – the type – via the haptic modality. If the multi touch device failed, a redundant method was implemented for safety reasons. The commands can also be entered by pressing device hardware respectively software buttons.

For the third and last part of the command – the value – the auditive modality is used. TriControl incorporates specific algorithms to detect spoken values such as numbers. By pressing a foot switch, recording of the subsequent utterance is started. The streamed audio file is the input for an automatic speech recognizer developed by Saarland University (UdS) and DLR [30]. For TriControl this speech recognizer is configured to analyze only command values without value units. There is a broad range of valid value types. The controller is allowed to speak between one and three consecutive digits ("zero, one, two, tree, four, five, six, seven, eight, niner"). For full multiples of ten, hundred or thousand, double numbers ("ten, twenty, thirty,…"), triple numbers (e.g. "two hundred") or a quadruple number (e.g. "four tousand") can be spoken. The system also recognizes special speed phrases ("own discretion, no restriction, minimum clean, final approach"). The speech recognizer accepts keywords for other clearances, e.g. inserting a handover by saying "tower", ILS clearance with the runway name e.g. "two tree right", or a *direct-to* command by a waypoint name in the Düsseldorf airspace ("Bottrop, Metma, Regno, Delta Lima 454 and so on"). Alternatively, values might also be selected from a software menu on the multi-touch device in cases of failure.

The value should of course correspond to a reasonable type to complete all three command parts. When all three modalities have been used, the TriControl system merges SR, ET, and MT data and displays it on the RadarVision screen. The whole command is presented, then to be validated by the controllers. For this visualization, five grey cells have been added to all aircraft radar labels (Figure 5). These cells represent five different command types. Cell one includes flight levels and altitudes, whereas all speeds in knots or Mach are presented in the second cell. The third label line contains the remaining three display areas for other current clearances. Headings, relative turns, waypoints, or transitions are shown there. Cell four includes rates of descent or climb. Cell five contains miscellaneous entries such as handover, ILS and localizer clearance, or holding.

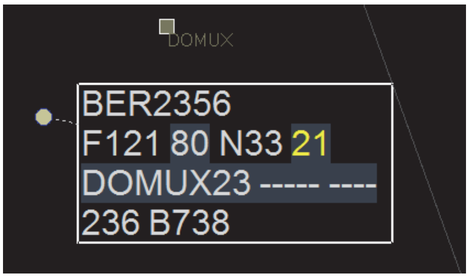

Figure 5. Interactive aircraft radar label cells in RadarVision

The completed command is shown as exactly one yellow value (command part three), in one of the five grey type cells (command part two) at one specific aircraft radar label (command part one). If the data is correct, the controller has to validate the command with a finger tap on the green-only area of the multi-touch device. The yellow value then becomes white. If controllers do not want to confirm the yellow value, they can cancel the entries using a hardware button on the tablet at any time during the process.

If the second or third part of the controller command, i.e. the type or a value, is selected after focusing on an aircraft label the eye tracking feature is locked. The purpose of this feature is to reduce unintentionally assigning command parts to other aircraft. Nevertheless, it is always possible to overwrite command type and value until the whole command has been confirmed or rejected. Furthermore, the obligatory multimodal activation (all three modalities are needed) enables the controller to freely look around without entering commands accidentally into the system.

After the command has been confirmed it will be sent to the aircraft. To yield the desired benefits in communication efficiency this may be done via reliable and fast data link connection. Even though data link technology with CPDLC (controller–pilot data link communications) protocol is now operational at many CWPs, most information exchanges between air traffic controllers and pilots still use voice communication owing to insufficient reliability in data link transfer speed. TriControl is designed to enable use of this digital connection by eliminating the speed bottleneck of human data input.

However, for reasons of compatibility with older aircraft equipment and the migration process from traditional communication, the concatenated command can also be sent via text-to-speech over the conventional radiotelephony channel. The pilot would then only experience a change to a more artificial and standardized voice that always sounds the same.

The following example explains how to insert the three parts of a command similarly to Figure 6.

Figure 6. Multimodal prototypic CWP exhibit TriControl

Figure 6 shows TriControl in a state where the three modalities have been used to insert data into the system: eye tracking (gaze on aircraft radar label of BER8411), multi-touch (two-finger circle-sector gesture indicating command type *heading*), automatic speech recognition (utterance of "two hundred"), situation data display (Düsseldorf approach area), and the resulting yellow input value (200 in grey "direction" cell) in aircraft label before validation of the controller command.

To reach this state the controller firstly looks at the BER8411 label on the radar situation display for nearly one second (see Figure 7).

Figure 7. TriControl setup with radar display attached eye tracker, headset, and multi-touch device

This may be achieved very naturally by the controller merely checking the label of the aircraft that is to be addressed. In this way the aircraft with the given callsign is selected as the aircraft which is to receive a new command.

Secondly, the controller touches the multi-touch device with two fingers, rotates them on the screen, and lifts his fingers again. This is understood as a *heading* gesture.

Thirdly, the controller presses the foot-switch and says "two hundred" using his headset. SR evaluates the speech and will deliver "200" as a result. All three parts of the command are concatenated to "BER8411 *heading* 200", which means that flight Air Berlin 8411 must turn its heading to 200 degrees.

As the three interaction modes can be used simultaneously, the air traffic controller's intention is entered into the ATC system fast. In our opinion, the controller will roughly need only one third of the time needed to utter the whole command with its three parts "air berlin eight four one one turn heading two hundred". The time needed to utter "two hundred" is simultaneous to the heading gesture and looking at the aircraft radar label BER8411.

In addition, the unilateral workload for verbal communication will be reduced and balanced with other modalities. It greatly relieves the strain on the voice from talking. The reduction in the total time needed to issue one command frees up the controller's cognitive resources.

This may even result in higher mental capacity and more efficient work if controllers can manage more aircraft at a time through reduced communication contact times. This could then also increase air traffic operational capacity.

IV. Evaluation of Multimodal CWP Demonstrator *TriControl*

For preliminary evaluation of the multimodal system usability we gathered structured feedback data of fair guests at DLR's World ATM Congress 2016 booth in Madrid who 'worked' intensively on our exhibit and agreed to participate in the inquiry. The survey comprised ten items from the System Usability Scale questionnaire (SUS) [14], three additional questions – one on each modality, and one summarizing item on the complete system.

Participants had to rate 14 statements on a Likert scale [31] from 0 to 4 meaning from "strongly disagree" to "strongly agree". The questionnaire items consisted of seven positively/negatively formulated statements. Twelve people (many of them air traffic controllers) took part in the survey.

A SUS score between 0 and 60 indicates poor usability; good usability starts at just over 75, becoming excellent the closer the score gets to 100.

The average SUS score in our survey was 79 and no single participant score was below 60. Two participants even rated usability with a SUS score of 90. Hence, usability of the whole multimodal CWP can be assumed as good.

The worst single item rating (2.9) was obtained for the SUS question on "Frequent Use", with the best (3.3) being obtained for questions on "Simplicity of Use" and "Using without Training" (see Figure 8). Black bars indicate the standard error as the quotient between variance and square root of sample size.

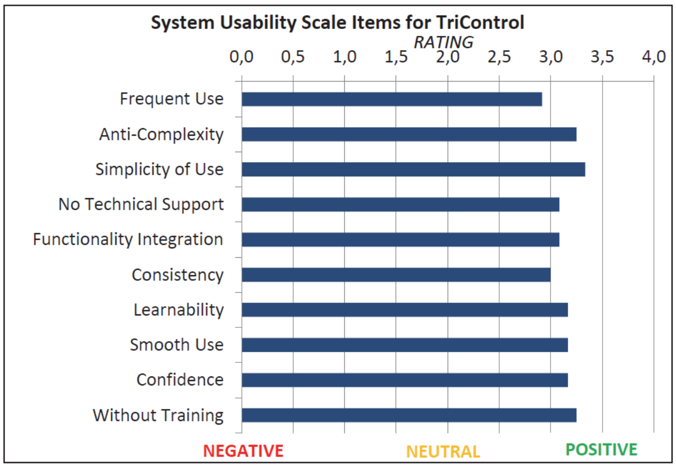

Figure 8. Participants ratings on system usability questionnaire items

A few minutes of exhibit use are not sufficient to contemplate a steady and more restricted use compared to current CWPs. However, simplicity was rated best (3.3). This demonstrates the clarity and intuitiveness of the multimodal concept.

All other item ratings lay between 3.0 and 3.3 (inversion of negatively formulated statement ratings for better comparability) showing good usability for different aspects of the prototypic multimodal CWP concept.

Multi-touch gesture recognition was rated best of the additional questions (3.3) (see Figure 9). Hence, after a quick training phase, the four different gesture types proved to be easy to remember and apply. Speech recognition of command values (3.1) worked very well for most participants. However, in a handful of speakers accents led to slightly lower recognition rates and demand for adjustments.

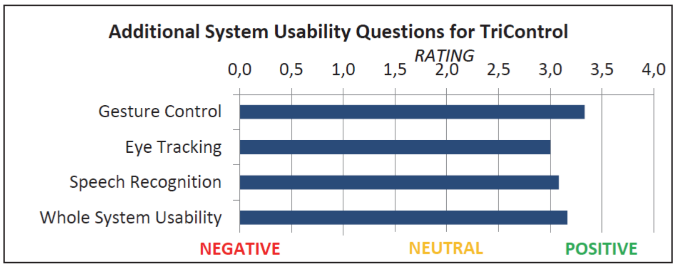

Figure 9. Participant ratings on additional system usability questions

Eye tracking was the most interesting and surprising modality being integrated. It worked fairly well for people without vision correction, with contact lenses, or glasses. Recalibration for each individual participant improved the eye tracking feature for aircraft labels.

Nevertheless, after changing seat settings or head position, the feature needed clearer gazes to react and was rated at 3.0, which is still within the good usability range. The great majority of participants were positively impressed by the overall performance of all three integrated modalities as reflected by the rating 3.2.

V. SUMMARY AND OUTLOOK

TriControl was the first ATC interaction demonstrator to combine eye tracking, multi-touch gestures, and speech recognition to generate full-featured controller commands. Dozens of air traffic controllers from roughly twenty different countries and all continents tested the TriControl exhibit in addition to those who took part in the survey. The feedback was broadly unanimous: training is needed, especially for simultaneously use of all modalities, but thereafter interaction is intuitive, fast, and straightforward.

The training need includes "handling" different devices simultaneously without looking at them. The training effect is similar to the difference between new and experienced drivers. Drivers have to manage different foot pedals, gearshift, steering wheel, indicator lights and other on-board equipment, whilst constantly watching the traffic, other drivers, pedestrians, road signs, etc. Thus, there is a general consensus that it should be fairly easy to acquire multimodal ATC interaction skills.

ATC experts also encourage further investigations into the advantages and drawbacks of multimodal interaction for controllers. Hence, other combinations of interaction modalities should be tested and compared. As ATC must satisfy stringent safety standards, the reliability and accuracy of these input modes must be very high.

In order to accommodate individual differences and preferences, the next phase anticipates allowing users to choose the modalities that they wish to use to interact with the system. Different extracts from the complete three times three matrix, comprising interaction modalities and controller command parts (see Figure 2), will be implemented in an enhanced version of TriControl. The speed gain for command input and interaction with TriControl should be measured against conventional systems. With rapid development of other innovative input technologies by the consumer industry, the mentioned matrix may grow further. When expanding this matrix to a tensor including parameters like personal preferences or variations over user workload even more combinations could be investigated.

Furthermore, each of the devices for interaction may be changed. The low-cost eye tracker could be replaced by a camera system, tracking head and eye position to improve accuracy and allow for greater freedom of body positions while working (see Figure 10).

A number of following studies shall prove and improve various aspects of TriControl. First, operational feasibility and suitability to controllers' requirements will be investigated.

Afterwards, experiments concerning user acceptance, usability, and related operational improvements will be performed and evaluated. Finally, capacity and safety will be analyzed. To this end, we will identify conditions for better and safer use of certain modalities.

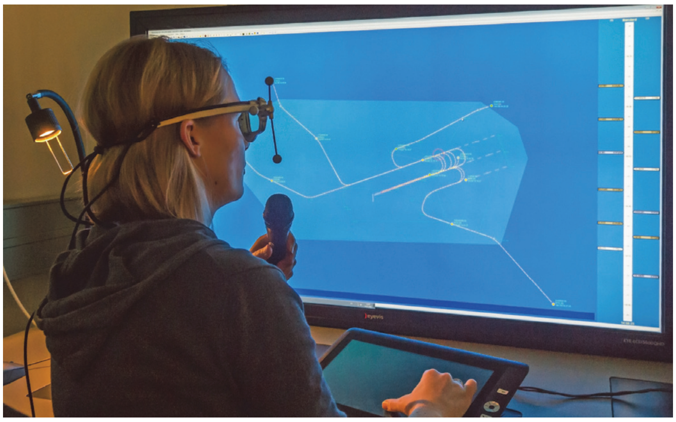

Figure 10. Advanced eye tracking device at CWP

With DLR's knowledge in CWP design and its validation infrastructure for executing realistic high-quality simulations, initial results and empirical evidence on the usefulness of multimodality for air traffic control will be gained in a continuative development phase to find out the best ways of achieving multimodal interaction.

REFERENCES

[1] SESAR, "The roadmap for delivering high performing aviation for Europe – European ATM Master Plan," Brussels, 2015.

[2] A. Labreuil, D. Bellopede, M. Poiger, K. Hagemann, M. Uebbing-Rumke, H. Gürlük, M.-L. Jauer, V. Sánchez, and F. Cuenca, "SESAR 10.10.02 D93, Innovation Analysis Report 2013," April 2014.

[3] AcListant® homepage: www.AcListant.de/wp.

[4] C. Möhlenbrink and A. Papenfuß, "Eye-data metrics to characterize tower controllers' visual attention in a multiple remote tower exercise," ICRAT, Istanbul, 2014.

[5] M. Uebbing-Rumke, H. Gürlük, M.-L. Jauer, K. Hagemann, and A. Udovic, "Usability evaluation of multi-touch displays for TMA controller working positions," 4th SESAR Innovation Days, Madrid, 2014.

[6] H. Gürlük, H. Helmke, M. Wies, H. Ehr, M. Kleinert, T. Mühlhausen, K. Muth, and O. Ohneiser, "Assistant based speech recognition – another pair of eyes for the Arrival Manager," 34th DASC, Prague, 2015.

[7] U. Ahlstrom and F. J. Friedman-Berg, "Using Eye Movement Activity as a Correlate of Cognitive Workload," in International Journal of Industrial Ergonomics 36: 2006, pp. 623–636.

[8] R. Alonso, M. Causse, F. Vachon, P. Robert, D. Frédéric, and P. Terrier, "Evaluation of head-free eye tracking as an input device for air traffic control," Taylor & Francis Group, Toulouse, France; Québec, Canada, 2012.

[9] L. E. Sibert and R. J. K. Jacob, "Evaluation of Eye Gaze Interaction," Proceedings of the CHI '00 Conference on Human Factors, in Computing Systems, New York, NY: ACM, 2000, pp. 281-288.

[10] P. Majaranta, I. S. MacKenzie, A. Aula, and K. J. Räihä, "Effects of Feedback and Dwell Time on Eye Typing Speed and Accuracy," in Universal Access in the Information Society 5: 2006, pp. 199–208.

[11] K. Kotani, Y. Yamaguchi, T. Asao, and K. Horii, "Design of Eye-typing Interface Using Saccadic Latency of Eye Movement," in International Journal of Human–Computer Interaction 26: 2010, pp. 361–376.

[12] Indra, "Advanced Controller Working Position," http://www.indracompany.com/sites/default/files/indra-indra_advanced_controller_working_position.pdf.

[13] D. Wald, "Implementation and evaluation of touch based air traffic control", original German title: "Programmierung und Evaluierung einer Touch basierten Flugverkehrskontrolle," Master Thesis, Langen, 2011.

[14] J. Brooke, "SUS - A quick and dirty usability scale," in Usability Evaluation in Industry, P. W. Jordan, B. Thomas, I. L. McClelland, and B. A. Weerdmeester, Eds. London, Taylor and Francis, 1996, pp. 189–194.

[15] SRI International, "Siri-based virtual personal assisstant technology," http://www.sri.com/engage/ventures/siri.

[16] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope, "Google search by voice: A case study," in Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics, Springer, 2010, pp. 61–90.

[17] C. Hamel, D. Kotick, and M. Layton, "Microcomputer System Integration for Air Control Training," Special Report SR89-01, Naval Training Systems Center, Orlando, FL, USA, 1989.

[18] D. Schäfer, "Context-sensitive speech recognition in the air traffic control simulation," Eurocontrol EEC Note No. 02/2001 and PhD Thesis of the University of Armed Forces, Munich, 2001.

[19] K. Dunkelberger and R. Eckert, "Magnavox Intent Monitoring System for ATC Applications," Magnavox, 1995.

[20] H. Helmke, O. Ohneiser, T. Mühlhausen, and M. Wies, "Reducing Controller Workload with Automatic Speech Recognition," 35th DASC, Sacramento, 2016.

[21] ETSI, EG, "Human Factors (HF); Multimodal interaction, communication and navigation guidelines," ETSI EG 202 191 V1.1.1, 2003, p. 7.

[22] Thales, "SHAPE - Innovative and Immersive ATC working position," https://onlineexhibitormanual.com/60atcaAnnual/PDF/Brochure_exhiReg477326_Shape.pdf, 2014.

[23] A. Schofield, "Thales' radical concept for controlling air traffic," in Things With Wings, http://aviationweek.com/blog/thales-radical-concept-controlling-air-traffic, 2014.

[24] P.-E. Seelmann, "Evaluation of an eye tracking and multi-touch based operational concept for a future multimodal approach controller working position", original German title: "Evaluierung eines Eyetracking und Multi-Touch basierten Bedienkonzeptes für einen zukünftigen multimodalen Anfluglotsenarbeitsplatz", Bachelor Thesis, Braunschweig, 2015.

[25] M.-L. Jauer, "Multimodal Controller Working Position, Integration of Automatic Speech Recognition and Multi-Touch Technology", Bachelor Thesis, Braunschweig, 2014.

[26] DLR Institute of Flight Guidance, "TriControl – multimodal ATC interaction", 2016, http://www.dlr.de/fl/Portaldata/14/Resources/dokumente/veroeffentlichungen/TriControl_web.pdf.

[27] O. Ohneiser, "RadarVision - Manual for Controllers", original German title: "RadarVision - Benutzerhandbuch für Lotsen", German Aerospace Center, Institute of Flight Guidance, Internal Report 112-2010/54, Braunschweig, 2010.

[28] Tobii homepage: www.tobii.com.

[29] Wacom homepage: www.wacom.com.

[30] H. Helmke, J. Rataj, T. Mühlhausen, O. Ohneiser, H. Ehr, M. Kleinert, Y. Oualil, and M. Schulder, "Assistant-Based Speech Recognition for ATM Applications", in 11th USA/Europe Air Traffic Management Research and Development Seminar (ATM2015), Lisbon, Portugal, 2015.

[31] R. A. Likert, "Technique for the Measurement of Attitudes," in Archives of Psychology 22, No. 140, 1932, pp. 5–55.