Article

Operational Feasibility Analysis of the Multimodal Controller Working Position "TriControl"

Oliver Ohneiser [1],*, Marcus Biella [1], Axel Schmugler [2] and Matt Wallace [3]

[1] German Aerospace Center (DLR), Institute of Flight Guidance, Lilienthalplatz 7,
    38108 Braunschweig, Germany; Marcus.Biella@DLR.de

[2] Technische Universität Dresden, School of Science, Faculty of Psychology, 01062 Dresden, Germany;
    Axel.Schmugler@Mailbox.TU-Dresden.de

[3] Deutsche Flugsicherung GmbH, Academy, Am DFS-Campus, 63225 Langen, Germany;
    Matthew.Wallace@DFS.de

* Correspondence: Oliver.Ohneiser@DLR.de

Abstract: Current Air Traffic Controller working positions (CWPs) are reaching their capacity owing to increasing levels of air traffic. The multimodal CWP prototype TriControl combines automatic speech recognition, multitouch gestures, and eye-tracking, aiming for more natural and improved human interaction with air traffic control systems. However, the prototype has not yet undergone systematic evaluation with respect to feasibility. This paper evaluates the operational feasibility, focusing on the system usability of the approach CWP TriControl and its fulfillment of operational requirements. Fourteen controllers took part in a simulation study to evaluate the TriControl concept. The active approach controllers among the group of participants served as the main core target subgroup. The ratings of all controllers in the TriControl assessment were, on average, generally in slight agreement, with just a few showing statistical significance. However, the active approach controllers performed better and rated the system much more positively. The active approach controllers were strongly positive regarding the system usability and acceptance of this early-stage prototype. Particularly, ease of use, user-friendliness, and learnability were perceived very positively. Overall, they were also satisfied with the command input procedure, and would use it for their daily work. Thus, the participating controllers encourage further enhancements to be made to TriControl.

Keywords: air traffic controller; human–machine interaction; usability; multimodality; eye-tracking; automatic speech recognition; multitouch gestures; controller command; feasibility analysis

1. Introduction

Facing the growing levels of air traffic and increased automation, conventional working methods and workstations are no longer adequate for air traffic controllers (ATCOs). Hence, there is a need for a faster, more efficient, and, ideally, more natural way of working, considering that a huge amount of ATCO work includes communication with aircraft pilots to issue commands. The communication between ATCOs and pilots is still mostly based on radio telephony. Hence, current human–machine interfaces (HMIs) in air traffic control (ATC) support the single voice interaction mode. However, this principle contradicts natural, intuitive, and efficient human communication. For instance, when explaining the route and distance to the airport to someone, verbal instruction is most often simultaneously accompanied by gestures for the direction and eye contact. In order to comply with these everyday interactions, the modalities should also be available at the controller

working position (CWP), which mainly comprises an interactive air traffic situation data display and the communication infrastructure.

First, controllers observe and analyze the air traffic at their situation data display. Thus, it is only a small step to utilizing eye-tracking technology to select the currently observed aircraft as the one to receive the next command. Secondly, verbal commands are an everyday concept in ATC. However, utterances can be reduced to only articulating command values. Thirdly, performing simple and fast multitouch gestures for command types—as are widely used nowadays on electronic consumer products—complements an easy and natural way of creating commands, which the controller finally confirms.

Multimodal HMIs combine different interaction modalities, aiming to support a natural [1] and efficient way of human communication [2,3]. Recent research has revealed that reasonable interaction technologies [4] for a CWP should recognize touch, speech, and gaze [5–8]. In accordance with these findings, the German Aerospace Center (DLR) has developed the multimodal CWP TriControl concept, which combines automatic speech recognition, multitouch gestures with one or multiple fingers on a touch input device, and eye-tracking via infrared sensors located at the bottom of the monitor. These modalities can be used to input the three basic ATC command parts, i.e., aircraft callsign, command type, and value, into the ATC system [9]. Hence, conventional subsequent command parts that are uttered verbally are replaced by parallel ATC system input with different modalities [10].

First analyses with the multimodal CWP prototype TriControl showed an acceleration of command input by up to 15% [10]. Furthermore, the artificial voice broadcast or data-link transmission of commands resulting from combined command parts of the parallel ATC system input in TriControl can also reduce misunderstandings in verbal communication caused by various "foreign language" English accents [11], that might even lead to serious accidents [12].

However, the operational feasibility including the system usability and acceptability of TriControl have not yet been systematically evaluated with ATCOs in a realistic environment [9]. As ATCOs work in a highly safety-critical domain, they and the air navigation service providers are very cautious with respect to new technologies [13].

The goal of this paper is to evaluate the multimodal CWP prototype TriControl in practice, i.e., mainly based on questionnaires after simulation runs, to receive input from target users for future development [14]. The evaluation concentrates on operational feasibility in terms of system usability and analyzing the fulfillment of operational requirements.

In the next section, we present the relevant background on multimodal HMIs, the validation methodology, and the TriControl system. Section 3 is the method part, introducing the participants, setup, and contents of the TriControl feasibility analysis study. All analysis results on system usability, acceptability, and performance are presented in Section 4. The main results and further comments are briefly presented in Section 5. Finally, Section 6 summarizes and concludes this paper, and sketches out future work.

2. Background of Multimodal Interfaces, Feasibility Analysis, and the CWP prototype TriControl

Many studies have investigated the advantages and disadvantages of specific interaction technologies as well as of multimodal HMIs. The most important results regarding multimodal HMIs and their relationship with the ATC domain are outlined in the following section. Furthermore, a theoretical background regarding the main aspects of the feasibility analysis study is outlined, i.e., concerning the validation methodology, the concepts of usability and acceptability, as well as the user-centered design. Finally, the functionality of the multimodal CWP prototype TriControl that was evaluated in a feasibility analysis is explained.

2.1. Multimodal HMIs and their benefits for ATC

Human–machine systems comprise reciprocal interaction between system components such as hardware and software as well as humans to achieve specific goals [15]. The communication channel for information between human and machine is called the "modality" [16]. HMIs usually utilize

visual-, audio-, and sensor-based modalities [17]. Hence, there can be unimodal or multimodal HMIs, depending on the number of utilized modalities. The HMI serves as a means for information exchange, including input and output. Well-designed HMIs can lead to better operational performance, safety [18], and efficiency [19], in particular, for high risk systems such as air traffic control workstations [20].

ATCOs shall ensure safe and efficient air traffic flow, also avoiding fuel and noise pollution, [21] being supported by their CWP. The basic tasks of ATCOs follow a cognitive cycle that consists of checking external information, searching for conflicts, issuing commands, and updating their mental picture [22]. The radar screen of their CWP is a central means to receiving external information, to search for conflicts, and to updating the ATCOs' mental picture. Thus, the information flow from the machine to the ATCO is mostly visually based. For issuing of commands—and the check of pilots' readbacks—ATCOs use audio-based radio telephony in two-way communication with pilots [23] and other ATCOs. Furthermore, there are touch-based input devices at modern CWPs, which are used to update the information in the system. However, due to a lack of reasonable parallel usage of those modalities, current CWPs are more like a set of different unimodal HMIs instead of a multimodal HMI as sketched.

The limits of human cognitive resources for information processing compared to the machine especially are a disadvantage of unimodal HMIs [24]. If the speed of information input or output, natural interaction, error-proneness, or individualization is essential, unimodal HMIs are often not the best concepts [17]. Conversely, multimodal HMIs take into account that humans process information modality-dependent and potentially simultaneously via different modalities considering the cognitive load theory [25] and the working memory model [26]. Respective interfaces and their analysis started in the 1980s [27].

Nowadays, many versions of multimodal HMIs exist, but all of them contain the principle of "more than one modality" for a human to interact with a machine [16,28,29]. Focusing more on the technical aspects and with regard to TriControl, a definition of multimodal interfaces could be "two or more combined user input modes—such as speech, pen, touch, manual gestures, gaze, and head and body movements—in a coordinated manner with multimedia system output." [30]. Multimodal HMIs have several advantages compared to unimodal HMIs. They promise to be more intuitive [31], to be better fitting to human needs for system control [32], to be faster, safer, and more reliable [33], to be less error-prone [34,35], to be more robust [35], to offer the best-suiting modality for users' choice [36], to reduce cognitive workload of users [37], and to comprise briefer and less disfluent input [34,38].

In a tested use case, roughly 95% of target users would prefer the multimodal over unimodal interaction [34]. However, users often mix unimodal and multimodal interaction [39,40]. Particularly in case of low cognitive load, users may even prefer unimodal interaction [41]. If users established their individual way of using multiple modalities, this interaction style would hardly change [42]. Overloading one or multiple modalities with information can result in less trust in the HMI [43].

Further findings on the three interaction modalities (automatic speech recognition, eye-tracking, and multitouch inputs as used by TriControl) are outlined in the following. The use of speech recognition in ATC started decades ago [44]. As ATCOs and pilots are obliged to use standard phraseology, a set of rules and terms for verbal radio telephony communication defined by the International Civil Aviation Organization (ICAO), the number of words and word sequences is limited compared to natural language [45]. Transcribing controller utterances word-by-word is only a small step for further applications [46]. The interpretation of those words also needs to be annotated to understand the respective command parts [47] especially if they do not follow the ICAO phraseology.

Given the commands, it is possible to highlight aircraft labels, to assess ATCO workload [48], or to enable safety functions [49,50] in operational environments or for the support of ATC training and simulations [51]. However, low command error rates are crucial for those applications. Thus, assistant-based speech recognition has been introduced [52]. A command hypotheses generator provides the most probable and reasonable commands in a given situation to a speech recognition

engine. This reduces the command error rate down to 1.7%. The respective radar label maintenance task supported by speech recognition reduces ATCO workload for ATC system input by more than 30% [53].

Eye-tracking interaction has already been analyzed in the air traffic domain [54,55]. The freeing of hands to be used for other manual interaction is one central advantage of this interaction means [56]. However, the visual selection of elements after a gaze dwell time does not seem beneficial [57]. The use of gestures in the ATC context has also been investigated. Earlier prototypes mostly include multitouch surfaces for more complex gestures [58]. Further examples from ATC research prototypes combine gestures with eye-tracking [59] or speech recognition [60]. Another application uses visual gesture recognition of air marshallers [61,62].

Touch gesture recognition has been investigated in the context of multimodal CWP [13,63]. Multitouch based interaction was evaluated as natural and fast enough for ATC applications [13]. Furthermore, users were able to work with the tested modalities quickly and perceived easy interaction [13]. Speed gains of up to 14% could be achieved compared to mouse inputs [64]. As ATCOs work in a highly safety-critical environment, the acceptability and trust by ATCOs of their HMIs is essential.

The user-centered design process takes target users into account in each design step of the HMI. Hence, there are early opportunities to influence the HMI development to the needs of ATCOs also in low technology readiness levels respective to validation phases.

2.2. Validation Methodology for Feasibility Analysis with Usability, Acceptability, and User-centred Design

The European Operational Concept Validation Methodology 3 (E-OCVM 3) developed by the European Organisation for the Safety of Air Navigation (EUROCONTROL) [65] provides a processual approach for the validation of air traffic management (ATM) operational concepts. The methodology shall include all relevant stakeholders and support the development process. The E-OCVM concept lifecycle model encompasses eight steps for maturing concepts based on iterative loops for design and evaluation. The steps for "validation phases" (V) are "ATM needs (V0)", "Scope (V1)", "Feasibility (V2)", "Pre-Industrial Development and Integration (V3)", "Industrialization (V4)", "Deployment (V5)", "Operations (V6)", and "Decommissioning (V7)" focusing on V1 to V3 for the concept validation methodology. Many of those phases are similar to the more popular "technology readiness levels" (TRL) 1 to 9 [66]. Hence, V1 corresponds to TRL2 "Technology concept and/or application formulated", V2 corresponds to TRL4 "Component validation in laboratory environment", and V3 corresponds to TRL6 "System/subsystem model or prototype demonstration in an operational environment" [67].

The TriControl CWP prototype is assumed to fulfill step V2 "feasibility" of E-OCVM 3 or TRL4 "Component validation in laboratory environment" respectively. In this step, the technological concept of a prototype in the ATC domain should be elaborated to be operationally feasible in normal and non-normal conditions, the latter e.g., comprising emergency flights or severe weather. An initial functional prototype should undergo a simulation for further analysis and revelation of further development needs. The aspect of feasibility itself is again subdivided into operability respectively usability, (system) acceptability, and performance.

Usability as one aspect of feasibility is defined as a construct with many facets including the aspects being "easy to learn, efficient to use, easy to remember, having a low error rate, and meeting user satisfaction" [68]. It can also be seen as the extent to which a system can be used regarding specified users and context to reach effectiveness, efficiency, and satisfaction [69]. Much background of the concept "usability" is given in [70]. The focus on users and environments next to just the tasks in the tool development is a central factor resulting from usability concerns [71,72]. If usability is taken into account, this can increase productivity, reduce training and support needs, improve users' acceptance [73], or even lead to higher efficiency [74]. Usability can be measured directly or indirectly [75,76] via questionnaires and interviews on perceived usability respectively via behavioral and interaction data from system experiments [77,78]. Therefore, a combination of evaluation methods improves the usability assessment [79]. System usability problems can be

detected with small user sample sizes. In specific studies, five study subjects were able to find 80% of usability problems and 15 study subjects detected all usability problems [80,81], however there may be hierarchies in those problems so that fixed numbers of subjects might not make sense [82].

Acceptability as another aspect of feasibility can be defined as the perceived usefulness and ease of use of a system to fulfill a task [83–85]. This affects the attitude towards the system as well as the behavioral intention and actual use of this system following the Technology Acceptance Model (TAM) [86]. The TAM has broadly been applied and developed high reliability to become a valuable acceptability assessment model [87]. If acceptability is taken into account during concept and system development especially in complex environments, this can avoid user resistance [88] and avoid negative use of the system such as obstruction or under-utilization [89]. Acceptability can be measured via Likert attitude scales [90] in questionnaires. A widely used questionnaire with 12 items [91] has high reliability [92].

Those aspects of feasibility can be assessed early in the applied "user-centered design" process. User-centered design encompasses the involvement of all relevant stakeholders in iterative design steps having an appropriate view on the requirements of tasks, users, and task distribution [73]. If user-centered design is applied, this can result in better acceptance of users [93], higher satisfaction [94], improved usability [95], and less training needs [96]. The iterative design loop of DIN EN ISO 9241–210 includes an analysis of system usage context, followed by a deduction of requirements and the development of a design solution that is evaluated afterwards. Non-satisfied user requirements of the evaluation tests lead back to the beginning of the design loop. This methodology has also been applied to certain extents to other ATC interfaces [97] next to TriControl.

2.3. Multimodal CWP TriControl

Nowadays, an ATCO usually issues verbal clearances to pilots via radio telephony and enters the clearances' contents manually into the ATC system. The clearances include structured information about the necessary pilot actions. The central contents are an aircraft callsign, a command type, and a command value. Pilot actions after readback of the clearance can lead to trajectory changes of the aircraft, e.g., due to speed, altitude, and heading changes; or can be of organizational manner, e.g., to handover the controlling responsibility for an aircraft to an ATCO of the adjacent airspace sector.

When using the multimodal CWP TriControl—combining three input modalities—an ATCO is able to generate a clearance with an aircraft callsign via gaze, a command type via multitouch gesture, and a command value via verbal utterance (see Figure 1).
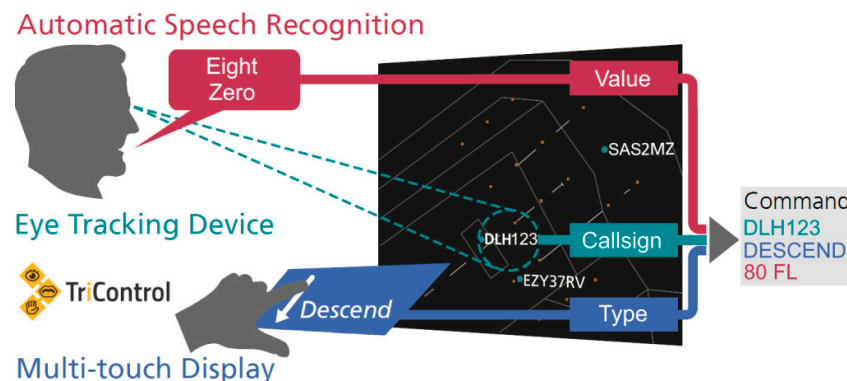

Figure 1. Multimodal interaction with TriControl combining the inputs from eye-tracking via gaze, automatic speech recognition via utterance, and multitouch display via gesture to a controller command (adapted from [98]).

The eye-tracking device detects the spot at the radar screen the ATCO is fixing for a certain dwell time. If there is an aircraft label or icon, the respective callsign is selected as the aircraft that will then receive the next command. Afterwards, the ATCO can input the command type and value

in sequence or in parallel. The targeted aircraft is locked as soon as the system detects that the ATCO performs a gesture or utters something, to avoid the clearance being sent to another aircraft that may be looked at thereafter.

The command type is entered into the system via two-dimensional gestures on the multitouch display. The one-finger gestures swipe down results in "descend", swipe up for "climb", swipe left for "reduce", swipe right for "increase", and long press for "cleared ILS/handover/direct to" depending on the value. When rotating semi-circle wise with two fingers, this will be recognized as a "heading" type. In an additional multitouch interaction mode that can be activated and deactivated with a button press on the multitouch device, some aspects of the graphical user interface of TriControl can be adjusted. This is done with two further multifinger gestures: Five fingers are used to zoom in and out on the radar map via spreading/contracting or they are just moved to pan the map; two fingers are needed to attach the distance measurement tool circles on two selectable aircraft.

The command value is spoken and recorded with a microphone. This value only consists of digits or is a waypoint, runway, or controller position name, respectively, i.e., "one five zero", "two hundred", "delta lima four five five", "two three right", "tower", etc. The command type gesture and command value utterance can be entered in parallel as Figure 2 demonstrates.

The callsign, type, and value are then combined to a controller command displayed to the ATCO. The uttered value is displayed in yellow in the type field of the corresponding aircraft label. For the validation trial, there was an additional top bar on the radar screen [99] with the three input elements next to a yellow value in the corresponding command type label field of the respective aircraft as shown in Figure 2. This visualization of the generated controller command before issuing it helps to detect mistakenly entered or falsely recognized command parts. Thus, the controller can either completely cancel the clearance or overwrite a wrong callsign as detected by gaze recognition, a wrong command type as analyzed by the multitouch device, or a wrong value as recognized by the speech recognition. Hence, there is even one more additional manual check for correctness with TriControl than just to listen to the pilot readback to determine if a conventional—completely verbal—clearance might contain an error.

If the ATCO acknowledges the completed command via a confirmation tap on the multitouch device, it is entered into the ATC system and could be further processed to influence the aircraft trajectory. Hence, the command could be sent to the aircraft via datalink or could be read by an artificial voice via the usual radio telephony channel. More details on the background of TriControl as well as functionalities especially with respect to aspects of command element input orders and timing can be found in [10].
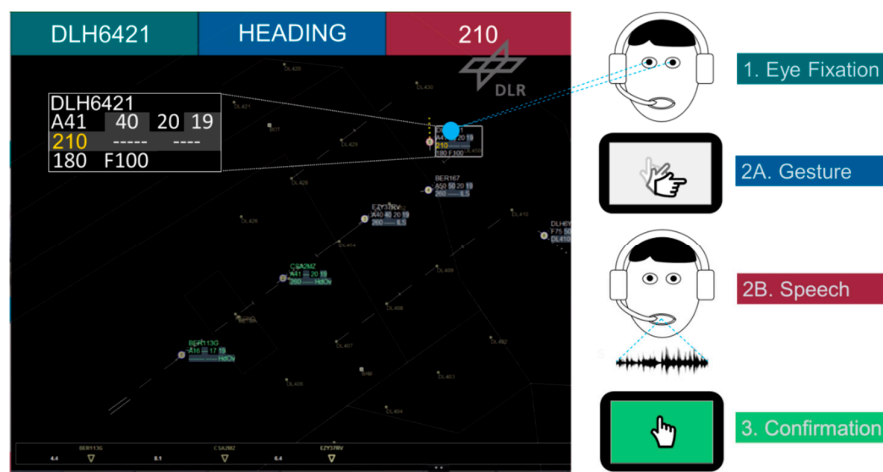

Figure 2. Input order of TriControl starting with eye fixation (on label of callsign DLH6421), followed by potentially parallel touch gesture (rotating two-finger semi-circle swipe) and speech utterance

(two one zero) being recognized, terminated with a confirmation tap (short press) to finalize a command.

3. Multimodal CWP Feasibility Analysis

The participants, setup, and tasks of the feasibility analysis study as well as questionnaires and study hypotheses are explained in the following sections.

3.1. Evaluation Site and Study Participants' Characteristics

The TriControl human-in-the-loop feasibility analysis study took place at the German air navigation service provider DFS Deutsche Flugsicherung GmbH in Langen, Germany in April 2017. Fourteen DFS ATCOs with an average age of 47 years (standard deviation (SD): 10 years) participated as study subjects. They had an average professional ATC experience of 21 years (SD: 12 years). The ATCOs worked at different positions, such as approach (7xAPP), area control center (5xACC), upper area center (2xUAC), tower (4xTWR), and as a generic instructor, so multiple answers were possible, and multiple perspectives were obtained. Four ATCOs were identified as the core target group being active APP ATCOs. This is due to the fact that TriControl is an approach control CWP, and current controlling skills, i.e., not being retired, influences the performance during the simulation study.

Some characteristics and experiences of ATCOs were asked as this could be relevant for the efficient usage of the different input modalities. Five participants had previous experience with eye-tracking, 3 with gesture-based inputs, and 10 with automatic speech recognition. Four participants did not use vision correction devices, two participants wore contact lenses, and eight participants wore glasses. Twelve of 14 participants were right-handed. All participants had appropriate English language skills as needed for air traffic control. The participants' native languages were German (9), English (3), Hungarian (1), and Hindi (1).

3.2. Tasks during the Human-in-the-Loop Study for Feasibility Analysis

The complete study included four different phases: Introduction, training, simulation run, and evaluation with debriefing. The TriControl concept and functionalities were described during the 15 min introduction. This included a standardized presentation about project goals as well as the system handling with the three interaction modalities and the graphical user interface taught by the technical supervisor.

The training phase consisted of a practice human-in-the-loop simulation run and lasted roughly 15 min. The traffic scenario used Düsseldorf approach airspace and comprised less aircraft than the later evaluation trial. This gave study participants time for repetition and familiarization with multiple and very different command inputs. Furthermore, they could focus on gathering information using the new radar screen environment. In addition, the eye-tracking device was calibrated to the participants' physical requirements. This phase was accompanied by the technical and psychological supervisors who answered open questions and corrected possible mistakes. As soon as the study participants stated his/her comfort and confidence with the system, the practice run was finished.

The simulation run in which participants worked with the TriControl CWP lasted a bit more than 30 min. The hardware comprised commercial-off-the-shelf devices (laptops, monitor, touchpad, eye-tracker, headset, foot switch). The Düsseldorf approach area with only active runway 23R was used as simulation setup (see Figure 3).

Figure 3. Setup of TriControl feasibility simulation study with a participant before uttering a command value, after performing a command type gesture and selecting an aircraft with his eyes.

The air traffic scenario comprised 38 arriving aircraft including seven of wake turbulence category "Heavy" and 31 "Medium". The scenario did not encompass departure traffic and was sufficient for one-hour maximum simulation time. Each participant's task was to work as a "Complete Approach" controller (meaning combined pickup/feeder ATCO in Europe and combined feeder/final ATCO in the US, respectively). The traffic scenario used standard arrival routes and there were no unusual traffic or weather conditions. The aircraft followed the issued command instructions directly after confirmation. Hence, the participants got an impression of TriControl's functional mechanisms in a standard ATC approach environment.

During the final evaluation and debriefing phase, all study participants needed to fill out questionnaires regarding feasibility with usability and acceptability, demographics, and profession-related data in the presence of the psychological supervisor. More precisely, 10 questions about personal data as well as 146 statements comprising the system usability scale (SUS) as well as the topics TriControl concept (T), eye-tracking (E), clearances (C), gestures (G), speech recognition (S), input procedure (I), and radar screen (R) plus 30 lines for optional comments on certain elements needed to be handled. Examples for those statements contain the ability to guide air traffic with TriControl (topic T), usefulness of eye-tracking (topic E), ability to issue different command types (topic C), learnability of gestures (topic G), user-friendliness of speech recognition (topic S), satisfaction with command input procedure (topic I), and identification of radar information (topic R) (see Appendix A for all statements to be rated). Together with further 17 categories for notes taken by the psychological supervisor during the experiment, this sums up to 203 lines of raw data for each of the 14 participants.

Three classes of requirements have been defined for the feasibility analysis of TriControl: (1) Multimodal interface fitness for intended use in the "TriControl concept", (2) "information retrieval", and (3) "command issuing". The developed questionnaires apply the norm DIN EN ISO 9241–11 (2017) with the subcategories effectiveness, efficiency, and satisfaction, as well as acceptability. For class (1–TriControl concept), the category effectiveness consisted of controlling air traffic as the core task of an ATCO. The category efficiency orients on DIN EN ISO 9241–11 Dialogue Principles (2006) for HMIs. The general requirements of this main norm were used for the category satisfaction. The category acceptability is assessed with widely used items of system use aspects [86]. For class (2–information retrieval), category effectiveness took the retrieving of information into account, category efficiency bases on DIN EN ISO 9241–12 Presentation of Information (2000), categories satisfaction and acceptability used the same sources as for class (1). For class (3–command

issuing), category effectiveness again took the core task of issuing commands into account, categories satisfaction and acceptability consider the E-OCVM 3 demands.

3.3. System Usability and Feasibility Analysis Questionnaire

To answer the research questions on basic feasibility and usability of the TriControl prototype in a quantitative and qualitative way, different assessment approaches were necessary. The quality of the operational concept and system usability of TriControl in general needed to be analyzed with a globally comparable measure. Therefore, the System Usability Scale (SUS) [100,101]—a subjective system usability assessment tool—was chosen. The SUS questionnaire consists of 10 statements to be rated on a scale comprising five possible answers coded as 0 to 4 points. The statements alter their positive-negative formulation respectively to prevent bias [102]. All ten items are multiplied with 2.5 to span a range from 0 to 100 whereas a higher score indicates better perceived system usability. The SUS proved to be highly reliable with an α of 0.911 [103] and to represent an overall trend [104]. Furthermore, the SUS scale has been used to evaluate TriControl in an earlier phase [9] and thus allows for better comparability and continuing the system usability assessment.

For the current analysis, the SUS score should indicate a sufficient value to represent a system usability as "ok", i.e., be at least at 50.9 as investigated in the literature [104]. Thus, the formal hypotheses on system usability of TriControl are:

$$H_{01}: \bar{x} < 50.9 \tag{1}$$

$$H_{11}: \bar{x} \geq 50.9 \tag{2}$$

The SUS score represents an overall score of ten recorded items [100] that is used for a point estimation. The SUS score is analyzed for a confidence interval (condition $\alpha=0.05$) for an interval estimation. It also investigates possible significant deviations from the critical cutoff value [79].

The feasibility was tested with a newly developed Likert-scale questionnaire based on user requirements. The self-based assertions aimed to evaluate the single elements of the TriControl system in a systematic manner [105]. The respective scale ranged from 1 (strongly disagree) to 6 (strongly agree) and included two further items (not important) and (not affected) [106]. The newly developed feasibility questionnaire should indicate at least a positive evaluation above the average score of 3.5 on the Likert-scale [90] ranging from 1 to 6 for all items. Hence, the formal hypotheses on feasibility of TriControl's operational concept are:

$$H_{02}: \bar{x} < 3.5 \tag{3}$$

$$H_{12}: \bar{x} \geq 3.5 \tag{4}$$

The non-parametric binomial test was used for the statistical significance analysis due to the small sample size of *N*=14. However, taking into account the robust binomial distribution supporting the null hypothesis, results will less likely be significant with respect to the desired direction [107]. The binomial test for each item included the n answers actually given by ATCOs, a test ratio of 0.5, an α of 0.05, and an expected mean value of 3.5 as the answers lay within 1 to 6. The further qualitative analysis was structured content-wise to deduct recommendations for certain feasibility elements according to [108].

Additionally, verbal remarks by the study subjects on the human–machine interface during non-task-interfering times were noted in a similar version compared to the Thinking Aloud technique [109]. Furthermore, non-verbal mistakes when using the prototype were noted [110].

4. Results of the Feasibility Study

The questionnaire results on system usability and feasibility as well as the most important comments of the 14 ATCOs are reported in the following sections.

4.1. Score of System Usability Scale (SUS)

The average SUS of TriControl for all 14 ATCOs was 60.9 (SD=21.9; lower and upper confidence interval limits: 48.3/73.5). Hence, the mean value indicates a system usability between "ok" (50.9) and "good" (71.7) [104]. However, the confidence interval overlapped the cutoff value. So, the mean value does not significantly deviate from the null hypothesis value of 50.9. It has to be rejected that the TriControl prototype offers a valid operational concept for ATC at the current stage. Though, when reducing the sample set to the core target group of active approach (APP) ATCOs (N=4), the results dramatically improved, i.e., the mean increased to 79.4 (SD=9.7; lower and upper confidence interval limits: 73.8/85.0). This would indicate a system usability evaluation of TriControl between "good" and "excellent" [104] as shown in Table 1. The SUS score of 79 also equaled an older non-systematic pre-evaluation of TriControl [9]. Table 1 also lists the 10 single SUS items S01-S10 representing a similar result regarding the ratings of active APP ATCOs. They did not perceive TriControl as cumbersome (S08) or inconsistent (S06) but would even like to use the system frequently (S01) with 3.5 points or more on the scale up to 4 points. Furthermore, the four usability statements S11-S14 on the three different input modalities and the combination of it was rated above the scale mean and, again, better from active APP ATCOs.

Table 1. Scores for system usability and four extra statements for single items and total system usability scale (SUS) score.

No.	System Usability Score Items[1]	N=14 (all ATCOs)		N=4 (active APP ATCOs)	
		M	SD	M	SD
S01	think that I would like to use the system frequently.	2.1	1.5	3.5	0.6
S02	found the system unnecessarily complex.[2]	2.6	1.2	3.3	0.5
S03	thought the system was easy to use.	2.5	1.2	3.3	0.5
S04	think that I would need the support of a technical person to be able to use the system.[2]	2.7	1.3	3.3	1.0
S05	found the various functions in the system were well integrated.	2.3	1.1	3.0	0.8
S06	thought there was too much inconsistency in the system.[2]	2.4	1.2	3.5	0.6
S07	would imagine that most people would learn to use the system very quickly.	2.2	1.1	2.5	1.0
S08	found the system very cumbersome to use.[2]	2.5	1.6	4.0	0.0
S09	felt very confident using the system.	2.1	1.1	3.0	0.0
S10	needed to learn a lot of things before I could get going with the system.[2]	2.9	1.1	2.5	1.7
	Total SUS score	**60.9**	**21.9**	**79.4**	**9.7**
S11	found that TriControl multitouch gestures for command selection are intuitive and easy to learn.	2.8	1.2	3.5	0.6
S12	think that the use of eye-tracking feature for selecting aircraft is disturbing.[2]	2.3	1.4	2.5	1.0
S13	think that automatic speech recognition is a good way to enter values.	2.2	1.4	2.8	1.5
S14	found the use of multiple modalities (eye gaze, gestures, speech) is too demanding.[2]	2.6	1.2	3.0	1.2

[1] Rating per single item from 0 "worst rating" to 4 "best rating", multiplied by 2.5 for Total SUS score. M represents the mean, SD the standard deviation.

[2] Statement rating has been "inverted" due to negative formulation, i.e., 0.5 points in the raw data are presented as 3.5 points here to enable better comparability of all items.

4.2. Feasibility Questionnaire Ratings

All 25 statements on the TriControl concept (T) are presented in Table A1. The 44 statements on command input in different categories (E/C/G/S/I) are shown in Table A2. Table A3 lists all 63 statements on the used prototypic radar screen (R). The tables include values for ratings' mean,

standard deviation, number of answers, and number of positive answers. They also list the *p*-value of the binomial test for significance analysis, i.e., to assess if the mean value significantly deviates from the null hypothesis value of 3.5. In roughly 85% of all 132 items of those three tables—especially except in the majority of items in category "R1.2 Coordination"—the rating of the active APP ATCOs was equal or better than the rating of all ATCOs. More than 55% of active APP ATCO ratings, on average, were equal or even above 5 points on the six-point scale. Some meaningful results per category are highlighted in the following.

4.2.1. Ratings on TriControl concept (T)

The active APP ATCOs rated the statements on the TriControl concept with an average of 4.8 points (on a scale from 1 to 6, see Table A1). Except for the statement on "T6.1 need of suitability for individualization", the active APP ATCO ratings were better than of all ATCOs, i.e., almost one point higher.

ATCO (in particular active APP ATCOs) were able to guide aircraft to their destination in an efficient way following the common safety requirements with TriControl (Controlling T1.1–T1.3). The TriControl interface was rated as appropriate for the intended use. ATCOs—especially with the parallel command input—felt supported to quickly and effectively achieve their best performance (Task Adequacy T2.1–T2.3). All ATCOs were aware of TriControl command input states and knew which and how actions could be executed to perform their controlling tasks due to the average ratings (Self-Descriptiveness T3.1–T3.4). They were also able to intuitively interact with TriControl as it matched common CWP conventions (Expectation Conformity T4.1–T4.2).

Furthermore, the statement ratings on timing and issuing of commands were rated above the scale mean (Controllability T5.1–T5.3). Particularly, active APP ATCOs felt safe to issue commands with little time and mental extra effort in case of a mistake (Error Tolerance T6.1). Active APP ATCOs less likely wanted to be able to adapt TriControl's interface to personal preferences than all ATCOs on average, even though they preferred to have the settings options (Suitability for Individualization T7.1). The satisfaction, notably of active APP ATCOs with TriControl, was good (T8.6). There were high ratings for the ease of use, user-friendliness, and learnability (Satisfaction and Acceptability of TriControl T8.1–T8.8). Some even wished to use TriControl in their daily work if they had the option.

To sum it up, almost all ratings were in the positive half of the scale, indicating a feasible TriControl concept even if not being statistically significant in all cases. Some circumspection existed to state that TriControl is preferred over common ATC interfaces, even in its current prototypic stage.

4.2.2. Ratings on command input

Every single active APP ATCO rating on the command input statements was better than that of the group of all ATCOs, i.e., more than 0.7 points better in average on the six-point scale (see Table A2). Almost one-third of the statements even had a significantly positive rating.

Ratings on eye-tracking (E)

The eye-tracking modality worked fine for aircraft selection. ATCOs perceived the eye-tracking as useful, user-friendly, as well as easy to use and learn (Aircraft Selection E1.1-E1.2; Satisfaction and Acceptability of the Eye-Tracking Feature E2.1–E2.8). Ratings of active APP ATCOs were mostly in the positive scale range.

Ratings on clearances (C)

According to the ratings, ATCOs were able to issue each type of clearance that TriControl offers. They also knew the command state they were in and could even simultaneously enter command type and value. Almost all statements were rated statistical significantly positive (Issuing Commands C2.1–C2.9).

Ratings on gestures (G)

The multitouch gestures to input the command type were perceived as useful, user-friendly, as well as easy to use and learn (Satisfaction and Acceptability of the Gesture based Command type input G2.1–G2.8) and could be an option for ATCOs' daily life CWPs with respect to ratings.

Ratings on speech recognition (S)

ATCOs were also satisfied with the automatic speech recognition modality for command value input in average, especially true for the active APP ATCOs (Satisfaction and Acceptability of Speech-Recognition based command value input S2.1–S2.9). Speech recognition was rated as useful, user-friendly, as well as easy to use and learn. Furthermore, the majority did not have problems verbalizing only the value instead of a whole command.

Ratings on input procedure (I)

The ratings on the complete command input are very similar to those of single input modalities (Satisfaction and Acceptability of the complete command input procedure I2.1–I2.8). TriControl received positive ratings for usefulness, user-friendliness, as well as ease to use and learn. In particular, active APP ATCOs were satisfied with eye-tracking, multitouch gesture recognition, speech recognition, and command confirmation elements of TriControl.

If all ATCOs are considered, the ratings on daily work use of TriControl and preference over conventional ATC interfaces are only around the scale mean. It is noticeable that all statements on effectiveness of the single interaction modes and TriControl as a compound, respectively (E/G/S/I.2.2 and T8.2) were rated rather negatively below the scale mean of 3.5. As TriControl would replace or support an APP CWP, respectively, it is also not expected that it would be more effective. Controlling air traffic via commands is still possible and is still the valid method to actively guide the traffic. The term effectiveness does not say anything about the efficiency of TriControl. The potential for efficiency gains—comparing TriControl with pure speech commands and following manual ATC system input—has been reported in [10].

4.2.3. Ratings on radar screen (R)

The majority of ATCOs' ratings on the radar screen used for TriControl were in the positive scale range and even statistical significantly positive (Aircraft within and aircraft heading to ATCOs' sector, orientation aids, centerline separation range (CSR), information design, as well as satisfaction and acceptability R1.1.1–R6.8). Active APP ATCOs had some difficulties to obtain weight classes and alphanumeric distances at the linker line between two aircraft as the appearance was different from their usual radar screen. The basic radar screen appearance should represent a common state-of-the-art like radar display. This was confirmed by ATCOs as it was usable for monitoring, designed to be user-friendly, and was easy to learn (R6.3–R6.5). The active APP ATCO ratings on the radar screen (see Table A3) were slightly better than the ratings of all ATCOs, i.e., more than two-thirds of the statements had better scores. However, even more than 60% of the statements for all ATCOs also had a significantly positive rating.

For the TriControl concept itself, it was important that ATCOs rated the statement on discriminability between different command states within the aircraft label (inactive, active, received, confirmed) positively (T3.2).

4.3. Feasibility Questionnaire Comments of all ATCOs

On the one hand, there are hints for improvements. Some ATCOs recommended that speech recognition needs to recognize multiple accents better as it did not work reliably for some ATCOs during the simulation trials. One ATCO perceived the foot pedal for speech recording as not helpful. Thus, a button at the headset microphone would be preferred. The technical issue with the non-reliable confirmation gesture recognition should of course be solved. One ATCO claimed that

he disliked the two-finger selection for separation assessment and that the left hand is completely unused. Moreover, the additional graphical user interface mode of the multitouch device offers too few functionalities to deserve its own category. A number of ATCOs reported that the eye-tracking was too slow reacting for them. Furthermore, aircraft have been deselected when looking away during command issue phases. The precision of the mouse has been rated better than eye-tracking especially in cases of vertically separated aircraft with partially overlapping labels. Also, the label itself seems to present too much information. Furthermore, many requests for additional functionalities were formulated such as the option to input combined and conditional clearances, to enter vertical speeds, to differentiate between left and right turns, to see the pilot statements in the aircraft radar labels as well, and a possibility for rotating or moving labels. The labels themselves can be analyzed separately as there are other dedicated studies and developments of labels and label interactivity. The main focus of this paper is on the multimodality.

Some further comments were made with respect to the simulation capabilities and the radar screen layout, which have not been central aspects of the feasibility study. Therefore, ATCOs wanted to have a better trail history and heading needles that are not overlapped by radar labels. Some aircraft did not follow instructions and the descent rates were too slow. Some ATCOs wanted better highlighting of heavy aircraft and information about whether an aircraft is on his/her frequency. Furthermore, a traffic forecast for the next ten minutes would be helpful. For some ATCOs, it was uncommon to work with a dark background radar map and without minimum vector altitude and airspace maps.

One ATCO remarked that the focus would change from "watching the traffic" to "watching if the system complies". Thus, the system feels like an extra step. In addition, there would be a great potential for confusion and errors due to convoluted features. Command issuing via voice was perceived as easier by some of the ATCOs, because it allows better situational awareness. A fallback redundancy would be necessary next to a safety assessment during further system development.

On the other hand, there are many aspects liked by ATCOs. Some ATCOs would prefer TriControl over mouse and screen input. Another ATCO still liked the Paperless Strip System (PSS) better; however, the mouse menu in the labels was perceived to be worse than in TriControl. Other ATCOs remarked that the use of TriControl is fun, it is easy to learn, and that they liked the system. The idea of combining three input methods was appreciated. One ATCO experienced no problem at all. For another, the speech recognition worked fantastically. The eye-tracking input was interesting and worked well after short practice for a number of ATCOs. Hence, it has potential with more development. If the system input was successful, the response is much quicker than with common systems to overall save time due to other ATCOs. It could also lead to fewer misunderstandings. Further thoughts were on the helpfulness in ATC training. An on-the-job-training instructor, who teaches ATCOs to be instructors, noticed that TriControl would be a good system to easily see what the controller is thinking and doing. The centerline separation range support functionality was especially appreciated as it was helpful without needing deduction. It was also reported that the plausibility check of command elements is a good idea and better than the solutions of competitors. Many ATCOs reported that their performance improved with practice and would further improve, so TriControl would be a good aid to ATCOs. They also encouraged following up the project.

5. Discussion of TriControl Feasibility

The system usability of TriControl as rated by all ATCOs was in a range up to a "good" result. This system usability score increased to a range up to "excellent" if considering only active APP ATCOs. However, these values should not be taken as equivalent due to the small sample size of only four active APP ATCOs. It can be a bias indicator with respect to the working positions the ATCOs usually work with. The system usability results were also reflected in the same range by the single system usability statements and the additional items for the specific interaction modalities.

The 132 feasibility analysis statements on TriControl concept, command input, and radar screen were slightly positive, whereas active APP ATCOs again agreed far more positively in the majority

of items. Particularly, ATCOs appreciated user-friendliness, usefulness, as well as ease to use and learn.

It is worth mentioning that there were great differences in the TriControl concept ratings depending on some personal and technical abilities of ATCOs during the simulation run. TriControl concept was rated much better by ATCOs—than by other ATCOs—if they:

- Were able to perform parallel input with different modalities,
- Hardly experienced any malfunction with the multitouch pad correspondence,
- Did not forget to perform the confirmation gesture after command completion,
- Did not perform wrong gestures,
- Did not experience some troubles with eye-tracking,
- Experienced more reliable speech recognition,
- Did not make other interaction mistakes, such as:
  - Too long press for confirmation and thus turning into a direct_to command,
  - Forgetting to toggle back from the multitouch device's graphical user interface mode,
  - Pressing the foot pedal for voice recording during complete command creation.

All of the above criteria do fit for the active APP ATCOs. However, it is not completely clear why the four active APP ATCOs performed much better and almost error-free compared to the other 10 ATCOs even if TriControl was designed as an APP CWP. The average age might be an indicator for that, i.e., the four active APP ATCOs were 37 years, the other ATCOs 52 years in average. Assuming that younger people are more familiar with modern interaction technologies from their daily life, this could explain the better ratings of active APP ATCOs. Furthermore, the simulation run time of 30 min might have been too short for ATCOs usually working on other positions to familiarize with the APP environment in addition to the new input modalities.

6. Summary, Conclusions, and Outlook

The feasibility of the multimodal CWP prototype TriControl—integrating eye-tracking, multitouch gesture, and speech recognition for command input—has been analyzed with 14 ATCOs in a human-in-the-loop study. Feasibility, system usability, and acceptability were judged slightly positive. The subgroup of active approach controllers agreed even more positively due to statistical analysis of questionnaire results. They were also motivated to further improve the TriControl system to bring it closer to operational needs as the achieved feasibility scores do not indicate significant showstoppers.

The SESAR2020 (Single European Sky ATM Research Programme) project PJ.16-04 CWP HMI "Workstation, Controller productivity" also dealt with automatic speech recognition, multitouch inputs, and eye-tracking in three different activities. However, those interaction technologies are not combined, but investigated stand-alone. Further research activities on the three interaction technologies will be continued in SESAR2020's wave 2 projects PJ.10-96 "HMI Interaction modes for ATC centre" and PJ.05-97 "HMI Interaction modes for Airport Tower". Hence, there is and will be research on modern interaction technologies in air traffic control. However, TriControl is one of the few concepts integrating multiple promising technologies to extract the benefits of each of them.

Recent iterations of ATC system development in general—and in particular, interface developments—have resulted in significant efficiency gains in the ability to process increased traffic levels, which soon rise in the real world to reach the new system limitations. A significant limitation in all further developments, however, seems to be the "bottleneck" of frequency congestion. The process of getting clearances clearly and safely submitted from the ATCO to aircraft and checking pilot readbacks for correctness is time consuming. The TriControl system seeks to address this with what could potentially be the effective removal of the existing bottleneck, allowing a greatly improved capacity increase.

According to the above study results, and to already revealed increased efficiency potential, further development of the early prototype TriControl will be performed to overcome the revealed malfunctions and integrate many suggestions for improvement. Afterwards, TriControl should be applied in different contexts also comprising non-nominal conditions such as weather influence,

high-density air traffic, and emergency aircraft. Then, TriControl's operational concept can be compared with current systems including cognitive workload assessment. Overall, the feasibility analysis motivated to foster multimodal interaction for air traffic control.

7. Patents

TriControl can serve as an input means and usable environment for the command generator European patent application 17158692.8.

Appendix A

Table A1. Evaluation of statements on TriControl core concept in different categories.

No.	Statement[1]	N=14 (all ATCOs)					N=4 (active APP ATCOs)	
		M	SD	n	k	p	M	SD
	Controlling							
T1.1	I was able to guide the aircraft to their destination.	4.8	1.1	14	11	.03	5.8	0.5
T1.2	I was able to guide the aircraft following the common safety requirements.	4.5	1.2	14	10	.09	5.3	1.0
T1.3	I was able to guide the aircraft in an efficient way.	3.4	1.7	14	8	.40	5.3	0.5
	Task Adequacy							
T2.1	The interface sufficiently supported me achieving my best performance.	3.6	1.5	14	7	.60	4.3	1.0
T2.2	The parallel command input enabled me to issue commands fast and effectively (type and value).	3.4	2.0	14	7	.60	5.3	0.5
T2.3	All in all the command procedure was appropriate for its intended use (input and feedback).	4.4	1.4	14	11	.03	5.0	0.8
	Self-Descriptiveness							
T3.1	I was always aware of the state of use I was currently operating in (monitoring, issuing commands).	3.9	1.1	14	9	.21	5.0	0.0
T3.2	I was always aware of the state the command input was in (inactive, active, receiving, received, accepted).	3.6	0.9	14	8	.40	4.0	0.8
T3.3	I always knew which actions I was able to execute at any given moment.	4.4	1.1	14	11	.03	5.3	0.5
T3.4	I always knew how those actions had to be executed.	5.1	0.6	14	14	.00	5.3	0.5
	Expectation Conformity							
T4.1	I was always able to intuitively interact with the interface the way I needed to.	3.6	1.6	14	9	.21	5.0	0.0

No.	Statement	M	SD	n	k	p	M	SD
T4.2	TriControl matched common conventions of use (content, depictions, specificity of numeric information etc.)	3.6	1.6	14	8	.40	4.8	0.5
Controllability								
T5.1	I was able to start the command issuing exactly when I wanted to.	3.8	1.8	14	9	.21	5.0	0.8
T5.2	I was able to control the command issuing the way I wanted to (proceed, cancel, or confirm).	4.1	1.5	14	11	.03	5.3	0.5
T5.3	I was able to control the pace at which the commands were entered.	3.6	1.7	14	9	.21	4.0	0.8
Error Tolerance								
T6.1	In case of a mistake, a command could still be issued with little extra effort (time and mental effort).	3.8	1.7	13	9	.13	5.0	0.0
Suitability for Individualization								
T7.1	I would like to be able to adapt the interface to my personal preferences.	4.2	1.5	9	7	.09	4.0	2.0
Satisfaction and Acceptability of TriControl								
T8.1	TriControl is useful for managing routine approach air traffic.	3.6	1.6	14	8	.40	4.8	1.0
T8.2	Working with TriControl is more effective than working with common interfaces.	2.8	1.4	14	3	.99	3.3	0.5
T8.3	TriControl is easy to use.	4.3	1.6	14	9	.21	5.5	1.0
T8.4	TriControl is user friendly.	4.4	1.6	14	10	.09	5.8	0.5
T8.5	It is easy to learn to use TriControl.	4.9	1.0	14	13	.00	5.8	0.5
T8.6	Overall, I am satisfied with TriControl.	4.2	1.4	14	10	.09	5.3	0.5
T8.7	I would want to use TriControl for my daily work if I had the option.	3.3	1.8	12	5	.81	4.0	2.2
T8.8	I would prefer TriControl over common ATC interfaces.	3.1	1.5	14	6	.79	3.5	1.3

[1] Rating per single item from 1 "worst rating" to 6 "best rating", other/missing ratings are ignored.

M represents the mean, SD is the standard deviation, n is the number of given valid ratings, k is the number of "successful" ratings above the scale mean ($\geq$4), and p is the p-value of the binomial test (1-tailed) that is underlined if equal or below 0.05 to state significance.

Table A2. Evaluation of statements on TriControl command input in different categories.

No.	Statement[1]	N=14 (all ATCOs)					N=4 (active APP ATCOs)	
		M	SD	n	k	p	M	SD
Aircraft Selection								
E1.1	I was able to select every aircraft I wanted to.	3.8	1.3	14	10	.09	4.5	0.6
E1.2	Only little effort was needed to select aircraft.	4.0	1.6	13	10	.05	5.0	0.8
Satisfaction and Acceptability of the Eye-Tracking Feature								
E2.1	The eye-tracking method is useful for aircraft selection.	4.1	1.4	14	11	.03	5.0	0.0
E2.2	The eye-tracking method works more effectively than conventional aircraft selection methods.	2.9	1.4	14	6	.79	3.0	1.4
E2.3	The eye-tracking method is easy to use.	4.1	1.5	14	10	.09	5.0	1.4
E2.4	The eye-tracking method is user-friendly.	3.9	1.4	14	10	.09	4.8	1.3
E2.5	It is easy to learn to use the eye-tracking method.	4.9	1.0	14	12	.01	5.5	0.6
E2.6	Overall, I am satisfied with the eye-tracking as a method of aircraft selection.	3.7	1.4	14	9	.21	4.8	0.5
E2.7	I would want to use it for my daily work if I had the option.	3.3	1.9	14	7	.60	3.8	2.2
E2.8	I would prefer it over conventional input methods.	3.1	1.8	14	7	.60	3.3	1.7
Issuing Commands								
C2.1	I was able to issue altitude clearance.	4.9	0.9	14	13	.00	5.8	0.5
C2.2	I was able to issue speed clearance.	4.9	0.9	14	13	.00	5.8	0.5
C2.3	I was able to issue heading clearance.	4.8	1.0	14	13	.00	5.5	1.0
C2.4	I was able to command heading to a certain waypoint.	5.0	0.7	12	12	.00	5.7	0.8

ID	Statement	M	SD	n	k	p		
C2.5	I was able to command hand over to tower.	4.0	1.8	14	9	.21	5.0	2.0
C2.6	I was able to identify when I was able to issue commands.	4.8	1.2	13	12	<u>.00</u>	5.8	0.6
C2.7	I was able to identify when my commands were being received.	4.7	1.3	14	12	<u>.01</u>	5.8	0.5
C2.8	I was able to identify when my commands were being accepted by the system.	4.8	1.1	14	13	<u>.00</u>	5.3	1.0
C2.9	I was able to enter command type and command value simultaneously.	3.9	1.5	14	9	.21	4.5	1.3
Satisfaction and Acceptability of the Gesture based Command type input								
G2.1	The gesture-based command type input is useful for the input of command types.	4.6	1.2	14	11	<u>.03</u>	5.5	0.6
G2.2	The gesture-based command type input is more effective than common approaches.	3.2	1.4	14	7	.60	3.5	1.3
G2.3	The gesture-based command type input is easy to use.	4.4	1.4	14	10	.09	5.3	0.5
G2.4	The gesture-based command type input method is user friendly.	4.1	1.5	14	9	.21	5.0	0.8
G2.5	It is easy to learn the gestures.	5.1	0.5	14	14	<u>.00</u>	5.5	0.6
G2.6	Overall, I am satisfied with the gesture-based command type input.	4.0	1.5	14	9	.21	5.3	0.5
G2.7	I would want to use it for my daily work if I had the option.	3.4	1.4	14	7	.60	4.0	1.2
G2.8	I would prefer it over common methods of command type input.	3.1	1.4	14	6	.79	3.5	1.3
Satisfaction and Acceptability of Speech-Recognition based command value input								
S2.1	Speech recognition is useful for the input of command values.	4.2	1.4	14	9	.21	5.3	0.5
S2.2	The speech recognition command value input is more effective than common approaches.	3.3	1.3	13	6	.71	4.3	1.2
S2.3	The speech recognition is easy to use.	4.0	1.6	13	8	.29	5.3	0.5
S2.4	The speech recognition-based command value input is user friendly.	4.1	1.5	14	9	.21	5.3	0.5
S2.5	It is easy to learn to use the speech recognition.	4.4	1.4	14	10	.09	5.3	0.5
S2.6	It was easy to get used to only verbalize the command value and not the whole command.	4.3	1.4	14	11	<u>.03</u>	4.8	0.5
S2.7	Overall, I am satisfied with the speech recognition-based command value input.	3.8	1.3	14	7	.60	4.8	0.5
S2.8	I would want to use it for my daily work if I had the option.	3.2	1.6	14	6	.79	3.8	1.5
S2.9	I would prefer it over common methods of command value input.	3.0	1.2	14	6	.79	3.3	1.0
Satisfaction and Acceptability of the complete command input procedure								
I2.1	TriControl command input procedure is useful for issuing commands.	4.3	1.3	14	10	.09	4.8	0.5
I2.2	The command input procedure is more effective than common approaches for command issuing.	2.9	1.3	14	5	.91	3.0	1.4
I2.3	TriControl's command input procedure is easy to use.	4.2	1.6	14	10	.09	5.0	0.8
I2.4	The combination of eye-tracking, gestures, speech recognition and confirmation is user friendly.	3.8	1.7	14	8	.40	4.8	1.3
I2.5	It is easy to learn to use the command input procedure.	4.7	1.1	14	13	<u>.00</u>	5.0	0.8
I2.6	Overall, I am satisfied with the command input procedure.	3.9	1.4	14	9	.21	5.0	0.0
I2.7	I would want to use the command input procedure for my daily work if I had the option.	3.1	1.4	14	7	.60	3.8	1.0
I2.8	I would prefer the command input procedure over common methods of command value input.	2.6	1.3	14	4	.97	2.8	1.0

[1] Rating per single item from 1 "worst rating" to 6 "best rating", other/missing ratings are ignored.

M represents the mean, SD is the standard deviation, n is the number of given valid ratings, k is the number of "successful" ratings above the scale mean (>=4), and p is the p-value of the binomial test (1-tailed) that is underlined if equal or below 0.05 to state significance.

Table A3. Evaluation of statements on radar screen prototypic design used for TriControl in different categories.

No.	Statement[1]	M	SD	n	k	p	M	SD
				N=14 (all ATCOs)			N=4 (active APP ATCOs)	
Aircraft within my sector: Identification								
R1.1.1	I was able to identify every aircraft's presence.	4.9	1.2	14	12	.01	5.3	0.5
R1.1.2	I was able to identify every aircraft's location.	5.1	0.9	14	13	.00	5.5	0.6
R1.1.3	I was able to identify every aircraft's call sign.	5.4	0.6	14	14	.00	5.8	0.5
R1.1.4	I was able to identify every aircraft's weight class.	4.0	1.8	14	7	.60	3.3	1.9
Aircraft within my sector: Coordination								
R1.2.1	I was able to obtain information regarding every aircraft's altitude.	4.7	1.1	14	12	.01	4.5	1.7
R1.2.2	I was able to obtain information regarding every aircraft's cleared altitude.	4.6	1.1	14	12	.01	4.5	1.7
R1.2.3	I was able to obtain information regarding every aircraft's speed.	4.7	1.1	14	12	.01	4.5	1.7
R1.2.4	I was able to obtain information regarding every aircraft's cleared speed.	4.6	1.1	14	12	.01	4.5	1.7
R1.2.5	I was able to obtain information regarding every aircraft's heading.	4.6	1.1	14	12	.01	4.5	1.7
R1.2.6	I was able to obtain information regarding every aircraft's cleared heading.	4.6	1.1	14	12	.01	4.5	1.7
R1.2.7	I was able to obtain information regarding every aircraft's next selected waypoint.	4.6	1.1	13	11	.01	4.3	2.1
R1.2.8	I was able to obtain information regarding every aircraft's distance to another aircraft.	3.8	1.5	14	9	.21	2.8	2.2
R1.2.9	I was able to obtain information regarding every aircraft's sequence number suggested by the AMAN.	4.3	1.3	11	8	.11	5.0	0.0
R1.2.10	I was able to obtain information regarding every aircraft's miscellaneous information (Cleared ILS, Handover to Tower).	4.2	1.4	13	8	.29	4.0	1.8
Aircraft heading into my sector: Identification								
R2.1.1	I was able to obtain the information that aircraft were heading into my sector.	4.0	1.1	11	7	.27	5.0	0.0
R2.1.2	I was able to obtain the information how many aircraft were heading into my sector.	3.8	1.2	12	6	.61	4.3	1.2
R2.1.3	I was able to obtain the call sign of every aircraft heading into my sector.	4.8	0.8	13	12	.00	5.3	0.5
Aircraft heading into my sector: Coordination								
R2.2.1	I was able to obtain every aircraft's estimated time of arrival (ETA).	2.5	0.5	6	0	.99	-	-
R2.2.2	I was able to obtain every aircraft's point of entry.	2.4	1.0	7	1	.99	-	-
Orientation Aids								
R3.1	I was able to obtain the runway location.	5.3	0.6	13	13	.00	5.3	0.6
R3.2	I was able to obtain the runway orientation.	5.4	0.5	13	13	.00	5.3	0.6
R3.3	I was able to obtain the extended runway centerline.	5.4	0.5	14	14	.00	5.3	0.5
R3.4	I was able to obtain the standard arrival routes (STAR).	5.2	0.6	11	11	.00	5.3	0.6
R3.5	I was able to obtain the borders of my airspace sector.	5.1	0.8	11	10	.01	5.3	0.6
R3.6	I was able to obtain GPS waypoints.	5.3	0.5	14	14	.00	5.3	0.5
The Centerline Separation Range								
R4.1	I was able to obtain the location of aircraft in final descend.	4.8	1.0	13	11	.01	5.3	0.5
R4.2	I was able to obtain the separation between aircraft and neighboring elements (runway, different aircraft).	4.8	1.0	13	11	.01	5.3	0.5

R4.3	I was able to obtain the weight class of aircraft in final descend.	4.3	1.6	13	9	.13	3.5	2.4
Information Design: Clarity								
R5.1.1	I was able to obtain all information quickly.	4.1	1.3	14	11	.03	4.3	1.7
R5.1.2	All information is as specific as I need it to be.	4.0	1.1	14	10	.09	4.3	1.0
Information Design: Discriminability								
R5.2.1	I was able to discriminate between different radar screen elements in general.	4.7	0.9	14	12	.01	5.0	0.8
Information Design: Discriminability - Aircraft								
R5.2.1.1	I was able to easily discriminate between different aircraft within my sector.	4.8	1.1	14	13	.00	5.5	0.6
R5.2.1.2	I was able to easily discriminate between different aircraft heading into my sector.	4.6	1.1	14	12	.01	5.3	1.0
R5.2.1.3	I was able to easily discriminate between different information within the label.	4.5	1.2	14	11	.03	5.3	1.0
R5.2.1.4	I was able to easily discriminate between different command states within the aircraft label (Inactive, active, received, confirmed).	4.1	1.5	14	9	.21	5.0	0.8
R5.2.1.5	I was able to easily discriminate between different indicated weight classes.	4.1	1.2	13	9	.13	3.8	1.7
R5.2.1.6	I was able to easily discriminate between different Arrival Manager order suggestions.	3.8	1.5	10	7	.17	5.0	1.4
R5.2.1.7	I was able to easily discriminate between different heading directions.	3.9	1.4	14	10	.09	4.3	1.7
Information Design: Discriminability – Orientation Aids and Centerline Separation Range (CSR)								
R5.2.2.1	I was able to easily discriminate between different categories of orientation aids in general.	4.7	0.8	11	10	.01	4.0	1.4
R5.2.2.2	I was able to easily discriminate between different GPS waypoints.	5.0	0.8	11	11	.00	5.3	0.8
R5.2.2.3	I was able to easily discriminate between different runways.	5.2	0.6	10	10	.00	5.3	0.8
R5.2.2.4	I was able to easily discriminate between different aircraft on the CSR.	5.3	0.7	8	8	.00	5.5	0.7
R5.2.2.5	I was able to easily discriminate between different distances between aircraft on the CSR.	5.0	0.9	8	7	.04	5.3	0.6
Information Design: Consistency								
R5.3.1	The format of the information given was consistent with what I expected it to be.	4.4	1.0	14	13	.00	4.8	1.0
Information Design: Compactness								
R5.4.1	I obtained all the information I needed to monitor the area effectively.	4.0	1.5	14	10	.09	5.0	0.8
R5.4.2	The radar screen didn't present any unnecessary information.	4.5	0.9	14	12	.01	4.3	1.0
Information Design: Detectability								
R5.5.1	I was able to direct my attention towards the currently necessary information.	3.9	1.4	14	10	.09	4.8	0.5
R5.5.2	The radar screen didn't divert my attention towards currently unnecessary information.	4.4	1.2	14	11	.03	5.3	1.0
Information Design: Readability								
R5.6.1	I was able to easily read alphanumeric information concerning the aircraft.	4.9	1.0	14	13	.00	5.5	0.6
R5.6.2	I was able to easily read alphanumeric information concerning the orientation aids.	5.1	0.7	14	13	.00	5.5	0.6
R5.6.3	I was able to easily read alphanumeric information within the CSR.	5.2	0.6	11	11	.00	5.5	0.7
Information Design: Comprehensibility of coded meaning								
R5.7.1	I was able to easily understand the coded information in	4.5	1.1	13	11	.01	5.3	0.6

		M	SD	n	k	p		
	general.							
R5.7.2	I perceived the used coding of information as unambiguous.	3.9	1.4	12	8	.19	4.0	1.7
R5.7.3	I was able to easily interpret all used codes.	4.1	1.3	13	9	.13	5.3	0.6
R5.7.4	I found it easy to deduce the coded meaning of the given information.	4.2	1.3	12	9	.07	5.3	0.6
Satisfaction and acceptability of the radar screen								
R6.1	The information design used in the radar screen is useful for sector monitoring.	4.1	0.8	14	12	<u>.01</u>	4.0	0.8
R6.2	The radar screen depicts information more effectively than conventional models.	2.8	1.2	13	3	.99	3.0	1.4
R6.3	The radar screen is easy to use for monitoring.	4.1	0.9	14	11	<u>.03</u>	4.3	1.0
R6.4	The radar screen design is user friendly.	4.2	1.0	13	11	<u>.01</u>	4.5	1.0
R6.5	It was easy to learn to use the radar screen.	4.4	1.0	14	13	<u>.00</u>	4.5	1.7
R6.6	Overall, I am satisfied with the radar screen information design.	3.9	1.2	14	10	.09	3.8	2.1
R6.7	I would want to use it for my daily work if I had the option.	3.0	1.3	13	6	.71	3.0	1.4
R6.8	I would prefer it over conventional radar screen designs.	2.7	1.2	13	4	.95	2.8	1.0

[1] Rating per single item from 1 "worst rating" to 6 "best rating", other/missing ratings are ignored.

M represents the mean, SD is the standard deviation, n is the number of given valid ratings, k is the number of "successful" ratings above the scale mean ($\geq$4), and p is the p-value of the binomial test (1-tailed) that is underlined if equal or below .05 to state significance.

References

1. Quek, F.; McNeill, D.; Bryll, R.; Kirbas, C.; Arslan, H.; McCullough, K.E.; Furuyama, N.; Ansari, R. Gesture, speech, and gaze cues for discourse segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2000 (Cat. No.PR00662), Hilton Head Island, SC, USA, 15 June 2000; Volume 2, pp. 247–254.
2. Oviatt, S.L. Multimodal interfaces. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*; 2003, pp. 286–304.
3. Oviatt, S.L. Advances in Robust Multimodal Interface Design. *IEEE Comput. Graph. Appl.* **2003**, 23, 62–68.
4. Koons, D.B.; Sparrell, C.J.; Thorisson, K.R. Integrating simultaneous input from speech, gaze, and hand gestures. In *Intelligent Multimedia Interfaces*; Maybury, M.T., Ed.; American Association for Artificial Intelligence: Menlo Park, CA, USA, 1993; pp. 257–276.
5. Uebbing-Rumke, M.; Gürlük, H.; Jauer, M.-L.; Hagemann, K.; Udovic, A. Usability evaluation of multi-touch displays for TMA controller working positions. In Proceedings of the 4th SESAR Innovation Days, Madrid, Spain, 25–27 November 2014.
6. Gürlük, H.; Helmke, H.; Wies, M.; Ehr, H.; Kleinert, M.; Mühlhausen, T.; Muth, K.; Ohneiser, O. Assistant Based Speech Recognition—Another Pair of Eyes for the Arrival Manager. In Proceedings of the 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, 13–17 September 2015.
7. Helmke, H.; Ohneiser, O.; Mühlhausen, T.; Wies, M. Reducing Controller Workload with Automatic Speech Recognition. In Proceedings of the 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016.
8. Möhlenbrink, C.; Papenfuß, A. Eye-data metrics to characterize tower controllers' visual attention in a multiple remote tower exercise. In Proceedings of the ICRAT, Istanbul, Turkey, 26–30 May 2014.
9. Ohneiser, O.; Jauer, M.-L.; Gürlük, H.; Uebbing-Rumke, M. TriControl—A Multimodal Air Traffic Controller Working Position. In Proceedings of the 6th SESAR Innovation Days, Delft, The Netherlands, 8–10 November 2016.
10. Ohneiser, O.; Jauer, M.-L.; Rein, J.R.; Wallace, M. Faster Command Input Using the Multimodal Controller Working Position "TriControl". *Aerospace* **2018**, *5*, 54, doi:10.3390/aerospace5020054.
11. Tiewtrakul, T.; Fletcher, S.R. The challenge of regional accents for aviation English language proficiency standards: A study of difficulties in understanding in air traffic control-pilot communications. *Ergonomics* **2010**, *2*, 229–239.

12. ICAO. The Second Meeting of the Regional Airspace Safety Monitoring Advisory Group (RASMAG/2). 2004. Available online: https://www.icao.int/Meetings/AMC/MA/2004/RASMAG2/ip03.pdf (accessed on 14 January 2020).

13. Chatty, S.; Lecoanet, P. Pen Computing for Air Traffic Control. In Proceedings of the CHI'96: SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 13–18 April 1996; pp. 87–94.

14. Schmugler, A. Feasibility Analysis of the Multimodal Air Traffic Controller Working Position Prototype "TriControl". Master Thesis, Technische Universität Dresden, Germany, 2018.

15. Czaja, S.J.; Nair, S.N. Human Factors Engineering and Systems Design. In *Handbook of Human Factors and Ergonomics*; Salvendy, G., Ed.; 2006, doi:10.1002/0470048204.ch2.

16. Bernsen, N. Multimodality Theory. In *Multimodal User Interfaces. Signals and Communication Technologies*; Tzovaras, D., Ed.; Springer: Berlin/Heidelberg, Germany, 2008, doi:10.1007/978-3-540-78345-9_2.

17. Adkar, P. Unimodal and Multimodal Human Computer Interaction: A Modern Overview. *Int. J. Comput. Sci. Inf. Eng. Technol.* **2013**, *2*, 8–15.

18. Norman, D.A. Design Rules Based on Analysis of Human Error. *Commun. ACM* **1983**, *4*, 254–258.

19. Nachreiner, F.; Nickel, P.; Meyer, I. Human factors in process control systems: The design of human-machine interfaces. *Saf. Sci.* **2006**, 44, 5–26.

20. Sheridan, T.B. Humans and Automation. System Design and Research Issues. In *Wiley Series in System Engineering and Management: HFES Issues in Human Factors and Ergonomics Series*; Human Factors and Ergonomics Society, Eds., Santa Monica, CA, USA; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2002; Volume 3.

21. EUROCONTROL. *Integrated Task and Job Analysis of Air Traffic Controllers—Phase 2: Task Analysis of En-route Controllers*; 1999.

22. EUROCONTROL. *Integrated Task and Job Analysis of Air Traffic Controllers—Phase 3: Baseline Reference of Air Traffic Controller Task and Cognitive Processes in the ECAC Area*; 2000.

23. Cardosi, K.M.; Brett, B.; Han, S. *An Analysis of TRACON (Terminal Radar Approach Control) Controller-Pilot Voice Communications*; (DOT/FAA/AR-96/66); Washington, DC, USA, DOT FAA, 1996.

24. Proctor, R.W.; Vu, K.-P.L. Human Information Processing: An Overview for Human-Computer Interaction. In *Human Computer Interaction Fundamentals*; Sears, A., Jacko, J.A., Eds.; CRC Press: Boca Raton, FL, USA, 2009; pp. 19–38.

25. Oviatt, S.L. Human-centered design meets cognitive load theory: Designing interfaces that help people think. In Proceedings of the 14th Annual ACM international Conference on Multimedia, New York, NY, USA, 23–27 October 2006; pp. 871–880.

26. Baddeley, A.D. Working Memory. *Science* **1992**, 255, 556–559.

27. Bolt, R.A. Put-that-there: Voice and gesture at the graphics interface. *Comput. Graph.* **1980**, *3*, 262–270.

28. Nigay, L.; Coutaz, J. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In Proceedings of the INTERCHI'93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, 24–29 April 1993; pp. 172–178.

29. Bourguet, M.L. Designing and Prototyping Multimodal Commands. In Proceedings of the Human-Computer Interaction INTERACT'03, Zurich, Switzerland, 1–5 September 2003; pp. 717–720.

30. Oviatt, S.L. Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems. *Adv. Comput.* **2002**, 56, 305–341.

31. Manawadu, E.U.; Kamezaki, M.; Ishikawa, M.; Kawano, T.; Sugano, S. A Multimodal Human-Machine Interface Enabling Situation-Adaptive Control Inputs for Highly Automated Vehicles. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1195–1200.

32. Pentland, A. Perceptual Intelligence. *Commun. ACM* **2000**, *4*, 35–44.

33. Seifert, K. Evaluation of multimodal computer systems in early development phases, original German title: Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen. Ph.D. Thesis, Technische Universität Berlin, Berlin, Germany, 2002, doi:10.14279/depositonce-530.

34. Oviatt, S.L. Multimodal interactive maps: Designing for human performance. *Hum. Comput. Interact.* **1997**, 12, 93–129.

35. Cohen, P.R.; McGee, D.R. Tangible multimodal interfaces for safety-critical applications. *Commun. ACM* **2004**, *1*, 1–46.

36.   den Os, E.; Boves, L. User behaviour in multimodal interaction. In Proceedings of the HCI International, Las Vegas, NV, USA, 22–27 July 2005. Available online: http://lands.let.ru.nl/literature/boves.2005.2.pdf (accessed on 14 January 2020).

37.   Shi, Y.; Taib, R.; Ruiz, N.; Choi, E.; Chen, F. Multimodal Human-Machine Interface and User Cognitive Load Measurement. *Proc. Int. Fed. Autom. Control* **2007**, *40*, 200–205.

38.   Oviatt, S. User-centered modeling for spoken language and multimodal interfaces. *IEEE Multimed.* **1996**, *4*, 26–35.

39.   Oviatt, S.L. Mutual disambiguation of recognition errors in a multimodal architecture. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Pittsburgh, PA, USA, 15–20 May 1999; pp. 576–583.

40.   Oviatt, S.L. Ten myths of multimodal interaction. *Commun. ACM* **1999**, *11*, 74–81.

41.   Oviatt, S.L.; Coulston, R.; Lunsford, R. When do we interact multimodally? Cognitive load and multimodal communication patterns. In Proceedings of the 6th International Conference on Multimodal interfaces, State College, PA, USA, 13–15 October 2004; pp. 129–136.

42.   Oviatt, S.L.; Coulston, R.; Tomko, S.; Xiao, B.; Lunsford, R.; Wesson, M.; Carmichael, L. Toward a theory of organized multimodal integration patterns during human-computer interaction. In Proceedings of the ICMI 5th international conference on Multimodal interfaces, Vancouver, BC, Canada, 5–7 November 2003; pp. 44–51.

43.   Marusich, L.R.; Bakdash, J.Z.; Onal, E.; Yu, M.S.; Schaffer, J.; O'Donovan, J.; Höllerer, T.; Buchler, N.; Gonzalez, C. Effects of information availability on command-and-control decision making performance, trust, and situation awareness. *Hum. Factors* **2016**, *2*, 301–321.

44.   Connolly, D.W. *Voice Data Entry in Air Traffic Control*; Report N93-72621; National Aviation Facilities Experimental Center: Atlantic City, NJ, USA, 1977.

45.   ICAO. *ATM (Air Traffic Management): Procedures for Air Navigation Services*; DOC 4444 ATM/501; International Civil Aviation Organization (ICAO): Montréal, QC, Canada, 2007.

46.   Helmke, H.; Oualil, Y.; Schulder, M. Quantifying the Benefits of Speech Recognition for an Air Traffic Management Application. In *Konferenz Elektronische Sprachsignalverarbeitung*; 15–17 March 2017; pp. 114–121.

47.   Helmke, H.; Slotty, M.; Poiger, M.; Herrer, D.F.; Ohneiser, O.; Vink, N.; Cerna, A.; Hartikainen, P.; Josefsson, B.; Langr, D.; Lasheras, R.G. Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04. In Proceedings of the IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK, 23–27 September 2018.

48.   Cordero, J.M.; Dorado, M.; de Pablo, J.M. Automated speech recognition in ATC environment. In Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems, London, UK, 29–31 May 2012; pp. 46–53.

49.   Chen, S.; Kopald, H.D.; Elessawy, A.; Levonian, Z.; Tarakan, R.M. Speech inputs to surface safety logic systems. In Proceedings of the IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, 13–17 September 2015.

50.   Chen, S.; Kopald, H.D.; Chong, R.; Wei, Y.; Levonian, Z. Read back error detection using automatic speech recognition. In Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, WA, USA, 26–30 June 2017.

51.   Updegrove, J.A.; Jafer, S. Optimization of Air Traffic Control Training at the Federal Aviation Administration Academy. *Aerospace* **2017**, *4*, 50, doi:10.3390/aerospace4040050.

52.   Helmke, H.; Ohneiser, O.; Buxbaum, J.; Kern, C. Increasing ATM Efficiency with Assistant Based Speech Recognition. In Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, WA, USA, 26–30 June 2017.

53.   Helmke, H.; Rataj, J.; Mühlhausen, T.; Ohneiser, O.; Ehr, H.; Kleinert, M.; Oualil, Y.; Schulder, M. Assistant-Based Speech Recognition for ATM Applications. In Proceedings of the 11th USA/Europe Air Traffic Management Research and Development Seminar (ATM2015), Lisbon, Portugal, 23–26 June 2015.

54.   Traoré, M.; Hurter, C. Exploratory study with eye tracking devices to build interactive systems for air traffic controllers. In Proceedings of the International Conference on Human-Computer Interaction in Aerospace (HCI-Aero'16), Paris, France, 14–16 September 2016; ACM: New York, NY, USA, 2016.

55. Merchant, S.; Schnell, T. Applying Eye Tracking as an Alternative Approach for Activation of Controls and Functions in Aircraft. In Proceedings of the 19th Digital Avionics Systems Conference (DASC), Philadelphia, PA, USA, 7–13 October 2000.

56. Hurter, C.; Lesbordes, R.; Letondal, C.; Vinot, J.L.; Conversy, S. Strip'TIC: Exploring augmented paper strips for air traffic controllers. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 22–26 May 2012; ACM: New York, NY, USA, 2012; pp. 225–232.

57. Alonso, R.; Causse, M.; Vachon, F.; Parise, R.; Dehaise, F.; Terrier, P. Evaluation of head-free eye tracking as an input device for air traffic control. *Ergonomics* **2013**, *2*, 246–255.

58. Westerman, W.C. Hand Tracking, Finger Identification, and Chordic Manipulation on a Multi-Touch Surface. Ph.D. Thesis, University of Delaware, Newark, DE, USA, 1999. Available online: https://resenv.media.mit.edu/classarchive/MAS965/readings/Fingerwork.pdf (accessed on 14 January 2020).

59. Seelmann, P.-E. 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's Thesis, Technische Universität Braunschweig, Braunschweig, Germany, 2015.

60. Jauer, M.-L. Multimodal Controller Working Position, Integration of Automatic Speech Recognition and Multi-Touch Technology, original German title: Multimodaler Fluglotsenarbeitsplatz, Integration von automatischer Spracherkennung und Multi-Touch-Technologie. Bachelor's Thesis, Technische Universität Braunschweig, Braunschweig, Germany, 2014.

61. Prakash, A.; Swathi, R.; Kumar, S.; Ashwin, T.S.; Reddy, G.R.M. Kinect Based Real Time Gesture Recognition Tool for Air Marshallers and Traffic Policemen. In Proceedings of the 2016 IEEE 8th International Conference on Technology for Education (T4E), Mumbai, India, 2–4 December 2016; pp. 34–37.

62. Singh, M.; Mandal, M.; Basu, A. Visual gesture recognition for ground air traffic control using the Radon transform. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2–6 August 2005; pp. 2586–2591.

63. Savery, C.; Hurter, C.; Lesbordes, R.; Cordeil, M.; Graham, T.C.N. When Paper Meets Multi-touch: A Study of Multi-modal Interactions in Air Traffic Control. In *Human-Computer Interaction—INTERACT 2013*; Lecture Notes in Computer Science; Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8119, pp. 196–213.

64. Mertz, C.; Chatty, S.; Vinot, J.-L. Pushing the limits of ATC user interface design beyond S&M interaction: The DigiStrips experience. In Proceedings of the 3rd USA/Europe Air Traffic Management Research and Development Seminar (ATM2000), Naples, Italy, 3–6 June 2000.

65. EUROCONTROL. *E-OCVM Version 3.0 Volume I—European Operational Concept Validation Methodology*; 2010.

66. NASA. Technology Readiness Level Definitions, n.d. Available online: https://www.nasa.gov/pdf/458490main_TRL_Definitions.pdf (accessed on 14 January 2020).

67. SESAR Joint Undertaking. Introduction to the SESAR 2020 Programme Execution, 2015. Available online: https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/jtis/h2020-pr-exec-intro-er-sesar-ju_en.pdf (accessed on 14 January 2020).

68. Nielsen, J. *Usability Engineering*; Academic Press: Boston, MA, USA, 1993.

69. *DIN EN ISO 9241-11:2016. Ergonomics of Human-System-Interaction—Part 11: Usability: Definitions and Concepts*; ISO: Geneva, Switzerland, 2017.

70. Chen, Y.-H.; Germain, C.A.; Rorissa, A. An Analysis of Formally Published Usability and Web Usability Definitions. *Proc. Am. Soc. Inf. Sci. Technol.* **2009**, *46*, 1–18.

71. Shackel, B. The concept of usability. In *Visual Display Terminals: Usability Issues and Health Concerns*; Ennet, J.L.B., Arver, D.C., Andelin, J.S., Smith, M., Eds.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1984; pp. 45–88.

72. Shackel, B. Usability—Context, Framework, Definition, Design and Evaluation. In *Human Factors for Informatics Usability*; Shackel, B., Richardson, S., Eds.; Cambridge University Press: Cambridge, UK, 1991; pp. 21–38.

73. Maguire, M. Methods to support human-centred design. *Int. J. Hum.-Comput. Stud.* **2001**, *55*, 587–634.

74. Weinschenk, S. *Usability: A Business Case, Human Factors International*; White Paper; 2005. Available online: https://humanfactors.com/downloads/whitepapers/business-case.pdf (accessed on 14 January 2020).

75. Seebode, J.; Schaffer, S.; Wechsung, I.; Metze, F. Influence of training on direct and indirect measures for the evaluation of multimodal systems. In Proceedings of the Tenth Annual Conference of the International Speech Communication Association (INTERSPEECH2009), Brighton, UK, 6–10 September 2009.

76. Nielsen, J.; Levy, J. Measuring usability: Preference vs. performance. *Commun. ACM* **1994**, *4*, 66–75.

77. Xu, Y.; Mease, D. Evaluating web search using task completion time. In Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'09), Boston, MA, USA, 19–23 July 2009; ACM: New York, NY, USA, 2009; pp. 676–677.

78. Wechsung, I. What Are Multimodal Systems? Why Do They Need Evaluation? Theoretical Background. In *An Evaluation Framework for Multimodal Interaction*; T-Labs Series in Telecommunication Services; Springer: Cham, Switzerland, 2014; pp. 7–22, doi:10.1007/978-3-319-03810-0_2.

79. Landauer, T.K. Research methods in human-computer interaction. In *Handbook of Human-Computer Interactio*; 1988; pp. 905–928.

80. Virzi, R.A. Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough? *Hum. Factors* **1992**, *4*, 457–468.

81. Nielsen, J. Why You Only Need to Test with 5 Users, 2000. Available online: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (accessed on 14 January 2020).

82. Schmettow, M. Sample size in usability studies. *Commun. ACM* **2012**, *4*, 64–70.

83. Ajzen, I.; Fishbein, M. *Understanding Attitudes and Predicting Social Behavior*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1980.

84. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *3*, 319–340.

85. Davis, F.D.; Bagozzi, R.P.; Warshaw, P.R. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Manag. Sci.* **1989**, *8*, 982–1003.

86. Davis, F.D.; Venkatesh, V. A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *Int. J. Hum.-Comput. Stud.* **1996**, 45, 19–45.

87. Yousafzai, S.Y.; Foxall, G.R.; Pallister, J.G. Technology acceptance: A meta-analysis of the TAM: Part 1. *J. Model. Manag.* **2007**, *3*, 251–280.

88. Kim, H.; Kankanhalli, A. Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. *MIS Q.* **2009**, *3*,567–582.

89. Markus, M.L. Power, politics, and MIS implementation. *Commun. ACM* **1983**, *6*, 430–444.

90. Likert, R.A. Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *140*, 5–55.

91. Davis, F.D. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1985. Available online: https://dspace.mit.edu/bitstream/handle/1721.1/15192/14927137-MIT.pdf (accessed on 14 January 2020).

92. Doll, W.J.; Hendrickson, A.; Deng, X. Using Davis's perceived usefulness and ease-of-use instrument for decision making: A confirmatory and multi-group invariance analysis. *Decis. Sci.* **1998**, *4*, 839–869.

93. Jackson, T.F. System User Acceptance Thru System User Participation. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*; American Medical Informatics Association: Bethesda, MD, USA, 1980; Volume 3, pp. 1715–1721.

94. Lin, W.T.; Shao, B.B.M. The relationship between user participation and system success: A simultaneous contingency approach. *Inf. Manag.* **2000**, *27*, 283–295.

95. Luna, D.R.; Lede, D.A.R.; Otero, C.M.; Risk, M.R.; de Quirós, F.G.B. User-centered design improves the usability of drug-drug interaction alerts: Experimental comparison of interfaces. *J. Biomed. Inform.* **2017**, *66*, 204–213.

96. Kujala, S. User involvement: A review of the benefit and challenges. *Behav. Inf. Technol.* **2003**, *1*, 1–16.

97. König, C.; Hofmann, T.; Bruder, R. Application of the user-centred design process according ISO 9241-210 in air traffic control. *Work* **2012**, *41*, 167–174.

98. DLR Institute of Flight Guidance, TriControl—Multimodal ATC Interaction, 2016. Available online: http://www.dlr.de/fl/Portaldata/14/Resources/dokumente/veroeffentlichungen/TriControl_web.pdf (accessed on 14 January 2020).

99.   Ohneiser, O. *RadarVision—Manual for Controllers, Original German Title: RadarVision—Benutzerhandbuch für Lotsen*; Internal Report 112-2010/54; German Aerospace Center, Institute of Flight Guidance: Braunschweig, Germany, 2010.

100. Brooke, J. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Iindustry*; Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L., Eds.; Taylor & Francis: London, UK, 1996; pp. 189–194.

101. Brooke, J. SUS: A Retrospective. *J. Usability Stud.* **2013**, *2*, 29–40.

102. Sauro, J.; Leis, J.R. When Designing Usability Questionnaires, Does It Hurt to Be Positive? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 2215–2224.

103. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum.-Comput. Interact.* **2008**, *24*, 574–594.

104. Bangor, A.; Kortum, P.T.; Miller, J.T. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* **2009**, *4*, 114–123.

105. Brinkman, W.-P.; Haakma, R.; Bouwhuis, D.G. Theoretical foundation and validity of a component-based usability questionnaire. *Behav. Inf. Technol.* **2009**, *28*, 121–137.

106. Jakobi, J. Prague—A SMGCS Test Report, 2010. Available online: http://emma2.dlr.de/maindoc/2-D631_PRG-TR_V1.0.pdf (accessed on 14 January 2020).

107. Bishop, P.A.; Herron, R.L. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int. J. Exerc. Sci.* **2015**, *3*, 297–302.

108. Burnard, P.; Gill, P.; Stewart, K.; Treasure, E.; Chadwick, B. Analysing and presenting qualitative data. *Br. Dent. J.* **2008**, *8*, 429–432.

109. Nørgaard, M.; Hornbæk, K. What do usability evaluators do in practice? An explorative study of think-aloud testing. In Proceedings of the 6th conference on Designing Interactive systems, University Park, PA, USA, 26–28 June 2006; ACM: New York, NY, USA, 2006; pp. 209–218.

110. Battleson, B.; Booth, A.; Weintrop, J. Usability Testing of an Academic Library Web Site: A Case Study. *J. Acad. Librariansh.* **2001**, *3*, 188–198.