

Article

# Integrating Eye- and Mouse-Tracking with Assistant Based Speech Recognition for Interaction at Controller Working Positions

Oliver Ohneiser <sup>1,2,\*</sup> , Jyothsna Adamala <sup>3</sup> and Ioan-Teodor Salomea <sup>4</sup>

<sup>1</sup> German Aerospace Center (DLR), Institute of Flight Guidance, Lilienthalplatz 7, 38108 Braunschweig, Germany

<sup>2</sup> Institute for Informatics, Clausthal University of Technology, Albrecht-von-Groddeck-Straße 7, 38678 Clausthal-Zellerfeld, Germany

<sup>3</sup> Faculty of Informatics, Automotive Software Engineering, Technische Universität Chemnitz, Straße der Nationen 62, 09111 Chemnitz, Germany; Jyothsna.Adamala@s2017.tu-chemnitz.de

<sup>4</sup> Faculty of Aerospace Engineering, "Politehnica" University of Bucharest, Str. Gh. Polizu No. 1, 1st District, 010737 Bucharest, Romania; Teodor.Salomea@euroavia-bucuresti.ro

\* Correspondence: Oliver.Ohneiser@DLR.de; Tel.: +49-531-295-2566

**Abstract:** Assistant based speech recognition (ABSR) prototypes for air traffic controllers have demonstrated to reduce controller workload and aircraft flight times as a result. However, two aspects of ABSR could enhance benefits, i.e., (1) the predicted controller commands that speech recognition engines use can be more accurate, and (2) the confirmation process of ABSR recognition output, such as callsigns, command types, and values by the controller, can be less intrusive. Both tasks can be supported by unobtrusive eye- and mouse-tracking when using operators' gaze and interaction data. First, probabilities for predicted commands should consider controllers' visual focus on the situation data display. Controllers will more likely give commands to aircraft that they focus on or where there was a mouse interaction on the display. Furthermore, they will more likely give certain command types depending on the characteristics of multiple aircraft being scanned. Second, it can be determined via eye-tracking instead of additional mouse clicks if the displayed ABSR output has been checked by the controller and remains uncorrected for a certain amount of time. Then, the output is assumed to be correct and is usable by other air traffic control systems, e.g., short-term conflict alert. If the ABSR output remains unchecked, an attention guidance functionality triggers different escalation levels to display visual cues. In a one-shot experimental case study with two controllers for the two implemented techniques, (1) command prediction probabilities improved by a factor of four, (2) prediction error rates based on an accuracy metric for three most-probable aircraft decreased by a factor of 25 when combining eye- and mouse-tracking data, and (3) visual confirmation of ABSR output promises to be an alternative for manual confirmation.

**Keywords:** air traffic controller; human machine interaction; multimodality; eye-tracking; mouse-tracking; automatic speech recognition; controller command prediction; attention guidance



**Citation:** Ohneiser, O.; Adamala, J.; Salomea, I.-T. Integrating Eye- and Mouse-Tracking with Assistant Based Speech Recognition for Interaction at Controller Working Positions.

*Aerospace* **2021**, *8*, 245. <https://doi.org/10.3390/aerospace8090245>

Academic Editor:  
Alexei Sharpanskykh

Received: 19 July 2021  
Accepted: 1 September 2021  
Published: 3 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

One central task of air traffic controllers (ATCos) is to issue verbal commands to aircraft pilots via radiotelephony in order to enable a safe, orderly, and expeditious flow of air traffic [1,2]. Usually, ATCos also need to enter this recently instructed command information into an electronic air traffic control (ATC) system such as aircraft radar labels or flight strips. This documentation supports ATCo hearbacks, i.e., to compare pilot's readbacks with ATCo instructions [3] and helps to monitor the aircraft status regarding the issued command characteristics.

If ATCo commands are issued via controller pilot data link communications (CPDLC)—being more common for non-time-critical commands in en-route phase—the content au-

tomatically feeds the ATC system and is uplinked to the aircraft pilot in order to be acknowledged. However, the traditional verbal way of ATCo-pilot communication that is assumed to remain in the medium-term future especially in highly dynamic and time-critical approach domain induces additional workload for the ATCo. This is because the ATCo needs to express the same information content twice—verbally for pilots via radiotelephony using standard phraseology according to ICAO (International Civil Aviation Organization) specifications [4] and manually for the ATC system.

Thus, automatically extracting the relevant command parts of verbal clearances to feed the electronic ATC systems without intense ATCo effort became a highly relevant technological topic in ATC. As a first step, automatic speech recognition (ASR) helps to provide the uttered words of ATC communication in written form. In addition, automatic command extraction from ATC utterances is also needed to understand the meaning of written word sequences. This language understanding task [5] can be heavily supported by using context knowledge about airspace situation, aircraft information, weather, etc. as provided through command predictions by an assistant system and used by an ASR engine.

Such assistant based speech recognition (ABSR) systems have proven to be a lightweight and easy-to-use technology to fulfill the task of ATC command recognition [6]. ABSR systems have also shown to improve air traffic management (ATM) efficiency and save aircraft fuel as ATCos can better guide air traffic with reduced workload [6]. However, ABSR command predictions have varying levels of accuracy, e.g., depending on individual ATCo habits and situations. Thus, it would be beneficial to know what part of the overall situation the ATCo currently processes—cognitively or manually.

Current prototypic ABSR implementations for ATC approach require a manual confirmation of ABSR output or a correction of recognized values, respectively [6]. Confirmation clicks via mouse are even needed if the ABSR system has low error rates [6]. Therefore, ATCos in ABSR studies are open to automatically accept ABSR output after a threshold time. However, this would also mean that sometimes unchecked and potentially erroneous ABSR output would also get automatically accepted.

Benefits of multimodal and more natural interaction at a controller working position (CWP) have already been investigated, i.e., to combine interaction technologies such as speech recognition and eye-tracking with each other to support ATCo tasks [7]. Hence, integrating further unobtrusive sensor data from eye- and mouse-tracking with ABSR and reasonably using these modalities' benefits promises to further improve efficiency of ATCos' CWP interaction.

The four derived research objectives are to (1) collect eye and mouse movement data of ATCos while monitoring radar traffic and prepare raw data for further applications, (2) extract relevant information from aforementioned interaction modalities and develop a framework to integrate the interaction data into an existing ABSR system to improve the overall performance, (3) develop and implement a method to calculate probabilities for predicted ATCo commands based on aircraft level and evaluate their quality, and (4) develop a CWP system to enable unobtrusive (visual) ABSR output confirmation and evaluate its usefulness.

Operator interaction data from eye- and mouse-tracking can support two important steps of ABSR applications as will be shown in this paper: (1) predict more accurate ATCo commands in order to reduce command recognition error rates, (2) check implicit ATCo confirmation of presented ABSR output or escalate attention guidance mechanisms to enforce ABSR output check. These two conceptual enhancements have been implemented, tested, and evaluated. The one-shot experimental case study with two controllers in a human-in-the-loop simulation of an ATC approach scenario at DLR Braunschweig in May 2021 revealed promising results—even if not significant due to the limited number of study subjects—to further refine the integrated use of interaction data: (1) command predictions on aircraft callsign level got more accurate by a factor of four, (2) combination of eye- and mouse-tracking metrics was superior over single modality metrics with an

improvement factor of 25 for prediction error rates, and (3) ABSR output confirmation by ATCos worked feasibly just by using gaze information.

Section 2 outlines related work on eye- and mouse-tracking as well as speech recognition and combinations of modalities relevant for ATC systems. Both, the baseline CWP and our CWP prototype with integrated eye- and mouse-tracking for ABSR output confirmation are described in Section 3. Section 4 explains the concept of assigning individual probabilities to command predictions based on ATCo interaction data. The study setup, methods, and subject data are explained in Section 5. The results of the study as sketched above are presented and discussed per conceptual enhancement in Section 6. Section 7 concludes and discusses the results more generally. Finally, Section 8 outlines future work.

## 2. Related Work on Speech Recognition, Eye-Tracking, and Mouse-Tracking

The following subsections give evidence to the use and benefits of speech recognition, eye-tracking, and mouse-tracking prototypes and applications as well as analyzes how the modalities can be used together and benefit from each other, respectively.

### 2.1. Related Work on Automatic Speech Recognition (ASR)

ASR means to convert speech, i.e., audio signals, into a sequence of words, commonly referred to as transcription. This transcription contains all uttered words and has special transcription rules for spelled letters, truncated and non-understandable words, human noise, and different versions of English or even non-English words [8]. The next important step is the language understanding, i.e., to transform the sequence of words into machine-readable semantic meaning, commonly referred to as annotation.

Speech recognition found its way into daily life as Amazon Alexa, Apple's Siri®, Google Assistant, or Microsoft's Cortana show. ASR activities in ATC [9] and using contextual knowledge to improve ASR began decades ago [10]. The mandatory use of ICAO standard phraseology, which limits the number of words and structures, helps to analyze verbal ATC communication [4]. However, transcription and especially annotation is more complex, because ATC radiotelephony users often deviate from the phraseology. Many European air navigation service providers and air traffic management system providers agreed on an ontology for annotating ATC utterances in a consortium led by DLR to enable better interoperability [11]. This ontology dramatically eases semantic interpretation especially when ATCos or pilots deviate from standard phraseology.

Assistant based speech recognition (ABSR) has proven to be a good approach [12] to achieve low ATC command recognition error rates [6]. In ABSR systems, ASR engines are supported by hypotheses about the next ATC commands, so called ATCo command prediction, that reduce the ASR engine's search space [13]. With this technology, command recognition error rates of below 2% are possible [14]. The command annotations can be used for further applications such as radar label maintenance to reduce ATCo workload [13], workload assessment [15], safety nets [16,17], arrival management planning input [18,19], or ATC simulation and training support [20,21]. The most advanced command prediction techniques base on machine learning and cover all relevant flight phases in the approach, en-route, and tower environment [22–24]. The command prediction error rate of an early implementation for multiple remote tower simulation command predictions was below 10% [25]. An ATC command prediction error rate of even 0.3% has been achieved for simulated Prague approach environment [26].

Another relevant metric is the portion of predicted commands, i.e., the number of predicted commands divided by the total number of commands per aircraft callsign, that an ATCo could theoretically issue. The lower the portion of predicted commands, the less alternatives that an ASR engine needs to choose from. For example, 144 heading commands are modeled as being usually possible with the qualifiers *RIGHT* and *LEFT* for the value range from 005, 010 to 355, 360. For the multiple remote tower environment, a context portion predicted of below 10% was achieved [25].

Currently, besides some statistical approaches, actually issued ATC commands were either predicted or were not predicted at all by an ABSR system, i.e., for comparison reasons we assume that predictions have a probability of one divided by the number of all predicted commands (uniform probability) or of zero. However, information about the certainty of different words and commands can support the ASR engine to choose the correct words [27,28].

## 2.2. Related Work on Eye-Tracking

Eye-tracking is a technology based on sensors to determine a human's gaze point and gaze movements as well as pupil size [29,30]. Most modern eye-trackers emit near-infrared light that is reflected by the eye's pupil and cornea [31]. These reflections can be measured with an infrared camera to derive the human's gaze points and further eye-tracking metrics [32]. Such eye tracking techniques do not distract the people involved because infrared light is invisible to the human eye.

Eye-tracking devices can be mounted on the head or can be worn as glasses with the advantage of free movement for the human user, but with the disadvantage of being more intrusive on the human's body [33]. Other eye-trackers can solely be mounted on a monitor. However, this leads to a restricted range of gaze detection. In a calibration process, the pupils' and corneas' reflection are matched with the screen coordinates that the human would be focusing on.

A number of metrics regarding eye-tracking have been established for further interaction analysis. A gaze point is a single point of gaze measurement that is often recorded with 50–60 Hz. A fixation is a cluster of subsequent gaze points defined through spatial thresholds and timely dwell times, such as 200–300 ms. There are many different algorithms for eye-tracking fixation identification based on spatial and temporal information [34]. Fixations indicate well the human's visual attention [31]. Given the fixation, the dwell time—hereinafter referred to as fixation duration—can also be measured [35]. The rapid eye movement segments between fixations are called saccades. The sequence of fixations and saccades is called scan path and is important to estimate user behavior in analyzing screen content [36]. Analyzing such scan pattern can help to train highly specialist screen users such as ATCos [37,38].

For the purpose of gaze analysis, certain spots of a screen are defined as areas of interest (AoI). An AoI is defined as “physical location, where specific task-related information can be found” [39]. The time spent on an AoI as a sum of fixations can be used to derive the human's attention or situational awareness in a broader view. This data is often presented as colored heat maps of human's gaze points on screen [40].

Eye-tracking is already widely used to analyze human's behavior on websites, e.g., using fixation count and fixation duration to predict customer interest and choices [41,42]. The time-to-first-fixation of an AoI was found to not support customer intention prediction [41].

In another study about eye-tracking based intent prediction with a support vector machine, a customer request prediction accuracy above 75% was achieved almost 2 s before the customer request towards a worker for an ingredient was uttered verbally [43]. Again, the fixation count and fixation duration (initial and in total) were considered. Furthermore, the fixation time was analyzed, i.e., how recent did the fixation happen on an AoI. Support vector machines using visual attention data have also been used successfully to predict human behavior in problem-solving tasks [44]. Hence, eye-tracking data can enable benefits in online applications, but also with offline analysis after recording [45].

Different research prototypes incorporating eye-tracking have already been developed for ATC [46–49]. Eye-tracking data assist to guide human ATC operators' attention via visual cues based on the desired and actual area of attention [50–52]. A combination of eye-tracking and electroencephalography was even used to control vigilance and attention of ATCos [53]. One important advantage of eye-tracking methods for ATCos is the potential to relieve them from tasks that would otherwise have to be done by hand [54].

### 2.3. Related Work on Mouse-Tracking

Mouse-tracking is a cheap and simple hardware-based method to acquire information that can be translated into visual attention later on. Human computer users can move a mouse to position a cursor on screen, can perform clicks with left and right mouse button, and scroll with a mouse wheel if applicable. The main mouse functions are metaphors of humans pointing to things (cursor) or touching things (selection of screen items with clicks) with their fingers or hands. Hence, mouse usage generates a variety of input data for the computer when users select text, hover over icons, or click to start events. Furthermore, this kind of tracking is unintrusive [55].

Mouse-tracking data for user intent prediction can be captured with a relatively low rate of 10 Hz [56]. Mouse cursor trajectories support understanding human decision processes [57,58]. Mouse movement paths seem to be more important than speed and acceleration of mouse movements in order to anticipate user decisions similar to the scan path in eye-tracking [59]. The cognitive processes related to eye- and mouse-tracking are similar as it is assumed in both cases to indicate visual attention [60]. Humans tend to use the mouse cursor for examining screen content, e.g., text reading and highlighting as well as interaction with screen content, but they may also ignore the mouse if it does not seem to be useful [61,62]. When clicking with the mouse, humans follow the mouse cursor even more visually compared to just move the mouse [56]. In more than two-thirds of the cases, the human watches the mouse cursor region on screen after a mouse saccade [63]. In more than 80% of the cases, if screen areas are examined visually, they are also examined with the mouse. Similarly, if they are not examined visually, they are also ignored with the mouse [63].

### 2.4. Multimodal Integration of Different Modalities Related to Human-Machine Interaction

Different approaches combine multiple interaction modalities to be used either independently of each other or to combine the advantages of them.

Eye-tracking can be used to re-assign probabilities of speech recognition hypotheses or to adapt the language model, respectively, by considering human's visual attention leading to significant decrease in word error rate [64]. However, achieved better recognition accuracy with such technique was connected more to the visual field than to the visual focus [65]. Eye-tracking and other non-verbal modalities have been combined to make speech recognition more robust against noise [66]. Eye-tracking was also found to be complementary to speech recognition for affect recognition in a gaming environment's multimodal interface [67] and for tracking reading progress [68].

The multimodal CWP prototype "TriControl" combines speech recognition, eye-tracking, and multi-touch sensing to issue ATCo commands [69]. The three main parts of an ATC command—callsign, command type, and command value—are entered into the ATC system via three different modalities, i.e., by looking at an aircraft radar label for the callsign, performing defined multi-touch gestures for the command type, and by uttering only the command value [70]. These three command parts are put together, confirmed, and sent to the aircraft via data link or electronically read, e.g., by looking at aircraft callsign "SAS818", swiping down for command type "DESCEND", and uttering "four thousand" for a command value of 4000 ft [71]. The possibility to work with different modalities in parallel enables faster and more intuitive interaction especially for approach ATCos [7].

Human-machine interfaces (HMI) that offer multiple modalities are called multimodal HMIs [72–75]. Multimodal HMIs can have several advantages such as robustness [76,77], quick, safe, and reliable use [78,79], individualized use [80], natural and intuitive interaction [81,82], workload reduction [83], and adaptation for certain human needs in environments like system control [84]. Human HMI users often change between multimodal and unimodal use [85,86]. Some tend to prefer multimodal interaction if well-designed [76], others prefer unimodal interaction especially in phases of low cognitive workload [87]. An example HMI for cars also offers speech, gaze, and gestures for system input [88].

Examples of multimodal research prototypes in ATC, e.g., combine gestures with speech recognition [89] or eye-tracking [90]. Additionally, in SESAR (Single European Sky ATM Research Programme) speech recognition and eye-tracking for attention guidance have been investigated and were found to be important future CWP technologies [91,92].

### 3. Description of Controller Working Position Prototype with Integrated Eye- and Mouse-Tracking for ABSR Output Confirmation

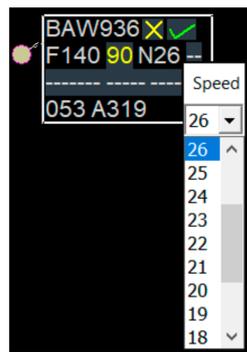
#### 3.1. Description of the Baseline Controller Working Position (Mouse-Click Trigger)

ATCos will be using the same basic CWP setup to evaluate the baseline and our solution system. The baseline includes the common interaction method with using symbols to be clicked in the aircraft radar label. The newly implemented solution system works by just looking or mouse-hovering at the aircraft radar label to start the ABSR output confirmation process. Hence, the majority of ATCos' tasks are the same in baseline and solution run as detailed in Section 5.2. ATCos have to monitor air traffic in approach phase with the given situation data display (see Figure 1).



**Figure 1.** Aircraft radar labels next to aircraft circle icons (containing sequence numbers) flying within Düsseldorf approach airspace shown on DLR's radar display RadarVision [93]. The five shaded label cells in the second and third label lines may depict the last ATCo command value for a certain command type (altitude, speed, direction, rate of altitude change, miscellaneous).

The first label line in any of the labels in Figure 1 indicates the callsign and the weight category in brackets. “medium” is the default weight class category. The second line shows (1) flight level (first letter is “F”) or altitude in hundreds of feet (first letter “A”), (2) the last given or recognized altitude command, (3) the speed in tens of knots (“N”), and (4) the last given or recognized speed command. The third line displays last issued heading/waypoint (“270”/“DL455”) clearances, rate of climb/descent with an arrow if applicable, and any other miscellaneous recently given command content such as an ILS-clearance (“ILS”) or handover to tower (“Twr”). The label example in Figure 2 also shows an optional fourth label line activated by mouse-over function with current heading (“053”) and aircraft type (“A319”).



**Figure 2.** Baseline aircraft radar label with white frame and yellow ABSR output value expecting manual ATCo confirmation through mouse click on green check mark (or rejection on yellow cross) and drop-down menu to change misrecognized or not recognized speed value.

Based on the air traffic situation and the ATCos' situational awareness, ATCos issue commands to aircraft pilots. The primary way to issue commands shall be the acoustic modality, i.e., to press a foot switch (push-to-talk), utter commands/clearances, and release the foot switch again. The recorded verbal utterance is analyzed in the speech recognition process by the ABSR system. The ABSR output is presented as yellow value in one of the five shaded aircraft radar label cells (see yellow flight level "90" in Figure 2). Clicking on one of the five shaded cells will open a drop-down menu to enable manual correction of the ABSR output. The first line of the aircraft radar label also shows a green check mark and a yellow cross to completely accept or reject all shown ABSR output for this aircraft, respectively. The former should ultimately be clicked if all ABSR output shown in the label is correct. All label values will then turn into white. Hence, the ABSR output confirmation by ATCos is triggered by mouse-clicks. In earlier trials with the same configuration, ATCos complained about the need to always click on the check mark given the high command recognition rate of the ABSR system. Furthermore, they need to move the mouse cursor—and thus also their gaze—to a less important area in the corner of the aircraft radar label. This causes additional manual and cognitive workloads. ATCos would rather just see the highlighted ABSR output that enters the ATC system directly if there is no ATCo intervention in a certain amount of time.

### 3.2. Description of the Solution Controller Working Position (Attention Trigger)

Based on the aforementioned ATCo recommendation, we modified the concept of ABSR output confirmation [94]. However, as a safety net, we still want to check if the ATCo at least noticed the ABSR output and did not intervene in a certain amount of time.

Thus, to avoid manual workload for ABSR output confirmation, the visual attention shall be used as a trigger in the confirmation process without the need for mouse clicks. One pre-assumption is that the ATCo has his/her visual attention at the spot he/she is looking at. This might not always be true, e.g., in case of staring at a certain position without presuming anything. However, this is a valid approximation to support ATCos in a visual task [50]. An infrared eye-tracker mounted on the bottom of the situation data display continuously records the ATCos' gaze points. The software module *ModEyeGaze* tries to match these gaze points with relevant objects displayed on the screen. These objects can be aircraft icons, aircraft labels, and airspace points.

The accuracy of eye-tracking is not of utmost importance, i.e., an accuracy of pixels is not required as it is not important to determine if the ATCo is looking at the speed or the altitude field in a label. An accuracy of roughly less than 1 cm is feasible to match the gaze points with displayed objects such as aircraft radar labels given a further visual threshold. Furthermore, a dwell time is defined in order to calculate a fixation on a displayed object. This avoids too many fixations in case the ATCo is just quickly shifting his/her view to the other side of the display. Like in the baseline system, yellow ABSR output values will

appear in the aircraft radar label immediately after the speech recognition process ends (see yellow values in Figure 3).



**Figure 3.** Solution aircraft radar labels with yellow ABSR output expecting attention-based ATCo confirmation and colored label frames in different states; left: light blue frame in saliency level “2” as visual check gaze for ABSR output is pending, right: green frame in saliency level “5” as visual check gaze has confirmed and time for potential manual ASBR output correction is running.

Peripheral cues are used to guide the operator’s attention [95]. More precise, different saliency levels of labels are applied depending on the visual check status by the ATCo to smoothly guide the ATCos’ attention to the relevant spots. All aircraft labels are in the default saliency level *transparent* (“−1”) initially. As soon as yellow ABSR output appears in a label, eye-tracking data analysis will be activated. The layout is as shown in Figure 2 of baseline system, but without the cross and check mark. The saliency level of the label will be escalated further every 5 s if *ModEyeGaze* does not detect an ATCo fixation on a highlighted aircraft radar label.

The label status is switched to saliency level *white* (“0”), i.e., a white label frame will be drawn. Saliency level *yellow* (“1”) with a yellow label frame is activated 5 s after the start of saliency level *white* to get the ATCo’s attention. Accordingly, saliency levels *light blue* (“2”) (see left label of Figure 3) followed by *dark blue* (“3”) are activated later after a gap of 5 s each. Thus, if there was no visual scan of the ABSR output (aircraft radar label) for 25 s after the appearance of the ABSR output value in yellow, the ABSR output will be rejected (*saliency level 4*) and does not enter the ATC system. The label’s saliency level will revert to *transparent* (“−1”) afterwards.

If *ModEyeGaze* detects an ATCo fixation on an aircraft radar label that has at least one unchecked yellow ABSR output value independent of the current saliency level, saliency level *green* (“5”) will be activated, i.e., a green label frame (see right label of Figure 3) will remain until the end of the maximum time for optional correction (10 s). If the correction time has passed, all visible yellow values in the aircraft radar label will enter the ATC system and the label will revert to saliency level *transparent* (“−1”) with all label values displayed in white color.

Eye-tracking as a technology might be more error-prone than manual system operator input especially if ATCos heavily move around with body and head compared to the calibration seating position. Therefore, mouse interaction data with the situation data display is used as a backup. The frequency of mouse usage by the ATCos depends on the CWP interaction design. However, as this data is just used as a backup data input, it is of less importance if the mouse is really used. Accordingly, if the mouse cursor is moved on an aircraft radar label that currently displays yellow ABSR output values and the mouse-over time exceeds a certain threshold time, this is determined as a match as if the ATCo would have looked at the label. Hence, the label frame turns green and counts down the remaining time for optional ABSR output value correction.

As system operators often carry their gaze, i.e., their visual attention, along with the mouse cursor, the gaze- or mouse-over initiated check of the solution system is called “attention triggered”.

#### 4. Description of Command Prediction Rescoring with Integrated Eye- and Mouse-Tracking

The second use case for operator gaze and interaction data is the enhancement of ATCo command prediction quality [96]. The implemented algorithm will be tested on the baseline run (Section 5.1), but also works if the ABSR output confirmation is used as in the solution system explained in Section 5.2. DLR's command hypotheses generator predicts ATCo commands for the speech recognition engine for given timeticks as shown below in Table 1.

**Table 1.** Examples for controller command predictions in ontology format with higher probability for aircraft that recently received ATCo attention.

Aircraft Callsign	Command Type	Second Type	Command Value	Unit	Qualifier	Uniform Probability	Re-Assigned Probability
AFR641P	HEADING		260		RIGHT	0.1	0.02
AFR641P	CLEARED	ILS	RW23R			0.1	0.02
AFR641P	DESCEND		4000	ft		0.1	0.02
BAW936	TRANSITION		DOMUX 23			0.1	0.06
DLH5MA	DESCEND		80	FL		0.1	<u>0.23</u>
DLH5MA	REDUCE		200	kt		0.1	<u>0.23</u>
DLH5MA	INFORMATION	QNH	1013			0.1	<u>0.23</u>
KLM1853	CONTACT		TOWER			0.1	0.03
KLM1853	CONTACT_ FREQUENCY		118.300			0.1	0.03
UAE57	DIRECT_TO		DL455		none	0.1	<u>0.13</u>

In Table 1's example, five different aircraft callsigns are predicted to possibly receive an ATCo command in the near future. For those callsigns different command types and values are reasonable due to their current airspace position and current motion characteristics. Hence, the number of predicted commands per aircraft can vary. In the basic ABSR implementation, no probability values are used, i.e., all predicted commands (here: 10 different ones) are assumed to have the same probability  $P(\text{cmd})_u$  (here 0.1). The basic advantage of this command prediction for the speech recognition engine is to know beforehand about commands that may be uttered (e.g., "AFR641P DESCEND 4000 ft") and to know, which will probably not be uttered (e.g., "KLM1853 DESCEND 4000 ft"). However, there might exist further data that even state which of the predicted commands are more likely to be uttered than others, i.e., to re-assign probabilities for command predictions with higher weightings for some aircraft commands (exemplarily underlined in column "Re-assigned Probability" with  $P(\text{cmd})_{ra}$  of Table 1). From an implementation point of view, the term *assignment* is more correct than *re-assignment*. However, the latter term better emphasizes to compare individualized probabilities against uniform probabilities for command predictions as outlined above.

It is important to note that the re-assignment does not intend to further predict yet unpredicted commands or to delete some predicted commands. Hence, as in the basic implementation, it can still happen that the ATCo issued a command to aircraft callsign "DAL27V", which is not a predicted aircraft callsign in the example of Table 1.

The basic pre-assumption is again: "the visual attention is where the ATCo looks at". However, some derived assumptions need to be made for this concept, i.e., display spots—including aircraft—that get more attention from the ATCo than others will more likely be involved in very near-term future ATC commands that the ATCo will issue. We assume that an ATCo will more likely give a command to an aircraft that he/she currently looks at or recently looked at—maybe even a multiple of times—as compared to an aircraft that was never looked at in the recent past by the ATCo, as determined by eye-tracking and

*ModEyeGaze*. In Table 1's example, we assume that DLH5MA and UAE57 have recently been looked at. Thus, predicted commands that include these aircraft callsigns receive probabilities above the "uniform" probability average for all commands. This implies that the probabilities for all the other aircraft needs to be reduced and re-assigned (AFR641P, BAW936, KLM1853).

Mouse interaction is again used as backup sensor data, i.e., if the ATCo moved the mouse and rested over an aircraft radar label recently or clicked very close by, this is considered to be similar to the visual attention via eye-tracking. For all interaction data stored in a data base, i.e., the combination of eye-tracking recorded with 60 Hz and mouse-interaction data recorded with 10 Hz (except the mouse clicks), different ratios will be tested. The most recent data from the last five to ten seconds for eye-tracking and the most recent data from the last three seconds for mouse-tracking is used in our concept due to expert feedback and initial feasibility testing. Three parameters of the recent past seconds will be considered for re-calculating probabilities: gaze duration on aircraft, gaze counts on aircraft, and mouse movements related to aircraft shown on a radar display.

#### 4.1. Command Probability Calculation Based on ATCo Interaction Data (Aircraft Level)

The calculation of probabilities for command predictions with respect to different aircraft based on ATCo interaction data will be explained in the following. The total command probability  $P(cmd)$  for a single command can be calculated with individual weightages  $W$  for each of the three interaction data metrics that sum up to one:

$$P(cmd) = W_{ETfix_{dur}} \cdot P(cmd)_{ETfix_{dur}} + W_{ETfix_{cnt}} \cdot P(cmd)_{ETfix_{cnt}} + W_{MTint} \cdot P(cmd)_{MTint}. \quad (1)$$

These metrics are called eye-tracking gaze fixation duration ( $ETfix_{dur}$ ), eye-tracking gaze fixation count ( $ETfix_{cnt}$ ), as well as mouse interaction data ( $MTint$ ) and will be explained in Sections 4.2 and 4.3.

#### 4.2. Command Probability Calculation Based on Eye-Tracker Data (Aircraft Level)

The total probability of an aircraft receiving an ATC command in the near future should be extremely high in case the ATCo looked at this aircraft for a long amount of time in the recent past. This mathematical weightage can be best expressed with an exponential function instead of a linear function. Thus, the re-calculation of probability  $P$  per command ( $cmd$ ) for a concrete aircraft ( $A/C_k$ ) based on eye-tracking gaze fixation duration ( $ETfix_{dur}$ ) is given by:

$$P(cmd_{A/C_k})_{ETfix_{dur}} = \frac{e^{dur_{A/C_k}}}{\sum_{i=1}^{\#A/C} (\#cmd_{A/C_i} e^{dur_{A/C_i}})}. \quad (2)$$

The parameter  $dur$  is the time spent on an aircraft during the last five seconds,  $\#cmd_{A/C_i}$  represents the number of predicted commands per aircraft with all aircraft from iterator start  $i = 1$  to the number of considered aircraft ( $\#A/C$ ) being summed up.

The eye-tracking gaze fixation count ( $ETfix_{cnt}$ ) in Equation (3) is considered in a linear way as the number of fixations on an aircraft is not assumed to be as an extreme indicator as the duration for an aircraft to receive the next ATC command. It is calculated with the following equation where  $cnt$  is the number of fixations for the specific aircraft in the last ten seconds:

$$P(cmd_{A/C_k})_{ETfix_{cnt}} = \frac{cnt}{\sum_{i=1}^{\#A/C} (\#cmd_{A/C_i} cnt)}. \quad (3)$$

Both eye-tracking probabilities ( $ET$ ) can be combined to a single probability with an appropriate weight.

#### 4.3. Command Probability Calculation Based on Mouse-Tracker Data and Combination of Interaction Data (Aircraft Level)

Mouse-tracking (MT) data are considered by Euclidian distance between the position of closest aircraft radar icon and position of mouse cursor/click. This closest aircraft

influences the mouse interaction weighting score  $miw$  to be (a) 5 if the aircraft has been visited with the mouse cursor for at least 300 ms or (b) 10 if the ATCo left/right clicked close to this aircraft as a sign of more active interaction with the aircraft's characteristics. The command probability based on mouse interaction data ( $MTint$ ) in Equation (4) is only considered for an aircraft ( $A/C$ ) if  $miw$  is greater than zero, i.e., if any mouse interaction close to the analyzed aircraft has taken place:

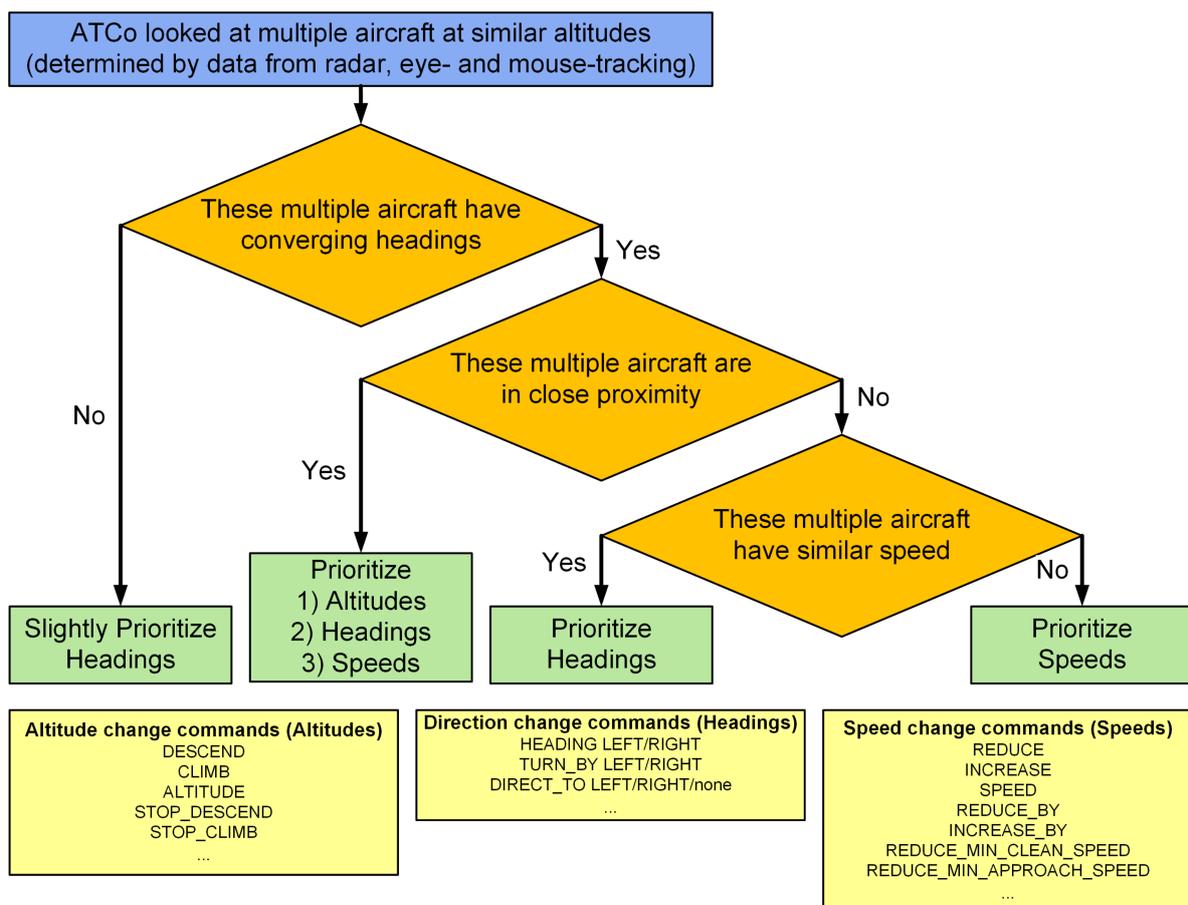
$$P(cmd_{A/C_k})_{MTint} = \frac{e^{miw_{A/C_k}}}{\sum_{i=1}^{\#A/C} (\#cmd_{A/C_i} e^{miw_{A/C_i}})}. \quad (4)$$

Inactive mouse interaction can result from the CWP design or from individual preferences of the ATCo. Unlike ET, positions of aircraft radar labels are not considered for MT as labels may overlap and may be moved away just for readability even if the labels are far away from aircraft icons and contain relevant information why the ATCo looks there.

#### 4.4. Air Traffic Situation Dependent Command Probability Combined with Interaction Data (Command Type Level)

We further assume that scanning different aircraft in the recent past leads to dedicated command types if some of the scanned aircraft have certain characteristics. For example, if the ATCo scans an aircraft close to the runway, the likelihood of a *CONTACT* command to the tower increases. If the ATCo fixes the gaze on a certain waypoint and on an aircraft for which this waypoint has been predicted as a command value, the likelihood for a *DIRECT\_TO* command to this waypoint increases. Furthermore, if an approach ATCo scans two or more aircraft at similar altitudes, the likelihood of commands from the categories of altitude change commands, direction change commands, or speed change commands can be adjusted as shown in Figure 4 based on ATCo feedback. For example, if scanned aircraft in similar altitudes have converging headings and are in close proximity, altitude change commands would be re-assigned with higher probabilities than heading change commands and especially than speed change commands. If these aircraft are not in close proximity, the speed difference might decide about prioritizing heading or speed change commands. Individual air traffic situations require individual decisions about ATC commands as well as individual conflict detection and resolution strategies [97], but slightly different probabilities on command type level can help to predict commands better on average.

If in Table 1's example DLH5MA was recently scanned, having the same altitude and intersecting path with another aircraft, the *DESCEND* command might be re-assigned with higher probability, e.g., 0.39 as compared to 0.15 for each of the *REDUCE* and *INFORMATION QNH* commands.



**Figure 4.** Flow chart to determine priorities for ATC command types based on aircraft scanned by ATCos.

## 5. One-Shot Experimental Case Study with Controllers in Simulation Environment

For a quantitative and qualitative evaluation on how DLR's ABSR application benefits from the use of eye- and mouse-tracking interaction data, relevant data from the simulation trials of a one-shot experimental case study was recorded in log files and data bases. This data comprises of:

- Positions of aircraft icons and aircraft radar labels with their states as shown on the situation data display
- Verbal utterances with automatic transcriptions, annotations, and instruction methods
- Eye gaze data with timeticks and fixation positions/durations
- Mouse interaction data with timeticks, click positions, and movements
- Answers of online questionnaires

### 5.1. Study Setup and Schedule for Evaluation of Eye- and Mouse-Tracking Support for Speech Recognition

In May 2021 we conducted an early interaction study at DLR Braunschweig with two controllers living close by—as COVID-19 restrictions prohibited trials with international ATCos. Hence, there was no scientific sampling and recruitment process. The study subjects were both male, roughly at the same age, wore a face mask (due to Covid-19 hygienic protocol), and spoke English with a German accent being relevant for speech recognition. Furthermore, both subjects wore glasses which is relevant for eye-tracking. One of the participants was an active licensed ATCo for tower and approach and the other participant was a former ATCo trainee for Düsseldorf approach area. Both subjects were not involved in the research activities and received the main part of the study information

only in the briefing session. The complete hardware setup of the prototypic CWP can be seen in Figure 5.



**Figure 5.** Study participant during simulation trials using an eye-tracking supported attention guidance system for assistant based speech recognition.

The subject used a foot switch to enable and disable voice recording (push-to-talk). The voice itself was recorded via the headset. The mouse placed to the right of the keyboard could be used to manually correct ABSR output or give commands via mouse. The leftmost monitor shows the situation data display with aircraft radar data in Düsseldorf approach airspace. The eye-tracker is mounted onto the bottom of this monitor. All other devices were not relevant for the subject's work during the scenario, but to run the simulation. The right monitor presents software module output of the arrival manager, the speech recognition engine, and the air traffic simulator running on the two Linux laptops on the right side of the photograph. The situation data display and the eye-tracking system runs on a Windows laptop (hardly visible below the right monitor). The disinfection material placed on the desk was used before a new operator started working on the CWP prototype to fulfill the hygienic protocol.

The software setup of the human-in-the-loop simulation comprised of an air traffic scenario for Düsseldorf approach (ICAO airport code EDDL). The only active runway was 23R. The duration of the scenario was one hour and included 38 approaching aircraft without considering departures. Seven aircraft were of weight category "heavy", all others were "medium" class aircraft. The participants had to handle the traffic being a "Complete Approach" controller, i.e., combined pickup/feeder ATCo in Europe or combined feeder/final ATCo in the US, respectively. This setup was similar to the earlier AcListant<sup>®</sup> [14,18], AcListant<sup>®</sup>-Strips [13], and TriControl [7,71] trials.

The four-hour-schedule of the study started with a 30-min briefing about the tasks to perform and included an eye-tracking calibration exercise. Two training runs for baseline and solution condition with roughly 20 min each and individual short breaks between

simulation runs followed. The baseline and solution runs themselves lasted up to one hour each—conducted in alternate order for the different participants to avoid bias. During the final half an hour, participants had to fill a questionnaire as well as needed to answer open questions and give comments during a debriefing.

### 5.2. Subjects Tasks and Execution of Simulation Study

The ATCos' task was to issue ATC commands primarily via voice by using the push-to-talk functionality. An example would be the following transcription of words: "lufthansa five mike alfa descend flight level seven zero turn right heading three six zero". If relevant parts of this utterance are correctly recognized by the speech recognition engine, the semantic representation of the utterance as per the agreed ontology, also known as the annotations would be displayed as follows: "DLH5MA DESCEND 70 FL, DLH5MA HEADING 360 RIGHT". These commands are converted to the necessary format for the air traffic simulator which itself changes the motion of the relevant aircraft. Hence, there are no active simulation pilots during the runs (amongst other reasons due to COVID-19 restrictions). All commands recognized by ABSR will be executed by the simulator. In almost all cases, misrecognized commands have not been shown as ABSR output, because they have been invalidated beforehand as not being plausible, due to reasons such as missing a correct callsign or a command value being out of a reasonable range.

Some technical problems of the CWP system that occurred during baseline and solution runs need to be mentioned that probably also affected the rating of the tested features. There was an operating system latency of roughly one second due to a laptop docking station issue that was only found after the trials. With this, there was a slight lag for the output display to appear, i.e., the confirmation saliency level, the ABSR output or the zoomed situation data display region appeared later than expected/theoretically possible. Furthermore, some commands have not been properly forwarded to the traffic simulator, i.e., altitude commands between 4000 and 6000 feet, *DIRECT\_TO*-commands, and some *ILS* clearances were affected. Nevertheless, all traffic could be handled and could be guided to land on the runway. As the flown trajectory did not matter for data analysis, but only the relevant eye- and mouse-tracking data, as well as the given ATC commands, the technical problems mentioned above should not heavily influence the basic conclusions of the simulation runs.

## 6. Results Regarding Effectivity of Eye- and Mouse-Tracking to Support Speech Recognition Applications

Data of two baseline and two solution runs has been recorded. Only the middle 45 min of the runs were analyzed to avoid data of a "slow start" and "scenario fading out". As Table 2 shows, ATCos issued 180 ATC commands per run on an average considering both modalities. Roughly 125 of these 180 ATC commands were recognized from slightly more than 100 speech utterances on an average, i.e., 1.3 ATC commands per speech utterance. The remaining 55 ATC commands were instructed via mouse in roughly 49 mouse issuing occasions, i.e., 1.1 ATC commands per mouse issuing occasion.

**Table 2.** Number (#) of actually issued ATC commands per run and command modality.

Run	# Actually Issued ATC Commands via Mouse	# Actually Issued ATC Commands via Speech	# Actually Issued ATC Commands per Run	# Speech Utterances/Mouse Issuing Occasions per Run
Baseline	88	105	193	154
Solution	22	146	168	144
All	55	125	180	149

In baseline runs, roughly 105 and 88 commands were issued via voice and mouse, respectively. The different types of issued ATC commands—by using both modalities with some misrecognitions—were *ALTITUDE* (36.4%, mainly *DESCEND*), *HEADING* (34%),

CLEARED ILS (13.6%), SPEED (6.6%, mainly REDUCE), CONTACT (6.5%), and others including DIRECT\_TO (3%).

Multiple thousand gaze fixations have been determined by the eye-tracking algorithm per run. A total of 42% of those fixations were on aircraft radar labels, 23% on aircraft radar icons, and 35% on airspace waypoints. In the baseline scenario, on an average more than 6000 mouse movements, around 250 left clicks, and less than ten right clicks on the situation data display have been captured per run.

#### 6.1. Enhancement of Probabilities for Speech Recognition Hypotheses by Eye- and Mouse-Tracking Data

This section compares the re-assigned ATCo command prediction probabilities with the uniform probabilities of the basic ABSR system implementation. The first part of the analysis concentrates on the benefits of re-assigned probabilities for different aircraft callsigns of command predictions while the second part also investigates re-assigned probabilities for different command types of single aircraft command prediction sets.

There are two basic result areas for the analysis. First, a factor showing the improvement in prediction accuracy as compared to the basic ABSR implementation, i.e., if the factor is greater than 1, the enhanced implementation outperforms the basic. Second, a four-field confusion matrix that helps to classify predicted and actually issued commands, i.e., the percentage of correct command predictions can be derived.

##### 6.1.1. Conditions and Metrics for Evaluating Prediction Probabilities on Aircraft Callsign Level

The recorded data is analyzed (1) for three conditions of eye- and mouse-tracking metrics as well as for two combinations of them, (2) for input modalities speech, mouse, and both combined, and (3) for the four simulation runs.

As explained above, the terms baseline and solution are right for the task of non-manual ABSR output check, but may be misleading for the task of analyzing the re-assignment of command prediction probabilities. However, the display appearance was slightly different in the two runs—cross and check mark in the first aircraft radar label line were not shown for solution runs unlike in baseline runs as explained in Section 3.2. Nevertheless, data from baseline and solution runs can loosely be compared with each other for a few special analyses. Therefore, the simulation runs are abbreviated as B (“baseline”) and S (“solution”). Mouse-tracker data only exists for the B runs as mouse-tracking has only been implemented for S runs’ setup; eye-tracker data exists for all runs.

The average improvement factor is calculated as shown in Equation (5) to sketch the enhancement of the probability ( $P$ ) re-assignment ( $ra$ ) concept compared to uniform ( $u$ ) probabilities per command ( $cmd$ ):

$$Improvement\ Factor = \frac{P(cmd)_{ra}}{P(cmd)_u}. \quad (5)$$

Five conditions or condition combinations, respectively, for the re-assignment of prediction probabilities based on aircraft level were analyzed with their influence on the prediction accuracy:

1. Only eye-tracking fixation duration of last 5 s to be considered (ETfix<sub>dur</sub>)
2. Only eye-tracking fixation counts of last 10 s to be considered (ETfix<sub>cnt</sub>)
3. Only mouse-tracking interaction data of last 3 s to be considered (MTint)
4. Combining (1) and (2) with 50% weightage each (ET)
5. Combining (4) with 70% weightage and (3) with 30% weightage (ET+MT).

From Equation (6) and using the definition in Table 3, *Accuracy* is defined as the percentage of correctly predicted ATCo commands. In other words, it is the number of commands predicted with above-average probabilities (compared to uniform average probabilities) which were actually issued plus the number of commands predicted with

average or below-average probabilities which were not issued divided by the number of all predicted commands:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}. \quad (6)$$

**Table 3.** Confusion matrix of ATC commands predicted vs. actually issued commands.

	Command Issued YES	Command Issued NO
Command Predicted YES	True Positive (TP)	False Positive (FP)
Command Predicted NO	False Negative (FN)	True Negative (TN)

More precisely, the following *Accuracy* values always consider Top  $N$  aircraft, e.g., for Top 2 A/C, the two aircraft callsigns that have the highest re-assigned probability compared to the other aircraft. Hence, if the ATCo actually issues a command to one of the two highest-ranked aircraft in terms of prediction probability, it is a TP. If the ATCo issues a command to the third ranked aircraft, it would be a FN. An aircraft is a FP if its callsign was predicted with above-average probability, but is not affected by the ATC command at the timetick it was issued. Finally, a callsign is said to be a TN if the used callsign was predicted with average or below-average probability and was not issued a command by the ATCo. As noted above, gazes on aircraft only influence the command prediction probability of callsigns if commands with the aircraft callsigns have been predicted in the basic implementation, i.e., in 3.2% of the cases aircraft callsigns receive a command that was not predicted. As it was neither predicted in the basic implementation, nor in the enhanced implementation, this has no negative influence on the defined *Accuracy*. Hence, if  $N$  is set to the maximum number of aircraft, *Accuracy* for Top  $N$  will be 100%.

Usually, there is a high one-digit number of aircraft to be considered at the same time as these are the aircraft under ATCo's responsibility. However, commands are only predicted for some of those aircraft as prediction for other aircraft might temporarily not be reasonable due to their motion characteristics. So, for each point in time when the ATCo issues one or multiple commands, there are usually multiple aircraft to be considered. For the four conducted simulation runs, commands have been predicted for 7.8 aircraft on an average at a time. Hence, for 149 prediction timeticks (100 speech utterances plus 49 mouse issuing occasions) almost 1200 aircraft callsigns have been predicted in total per run. Based on experiments, it is thus most reasonable to consider the Top 3 A/C only. Top 3 A/C are selected as shown in Table 4.

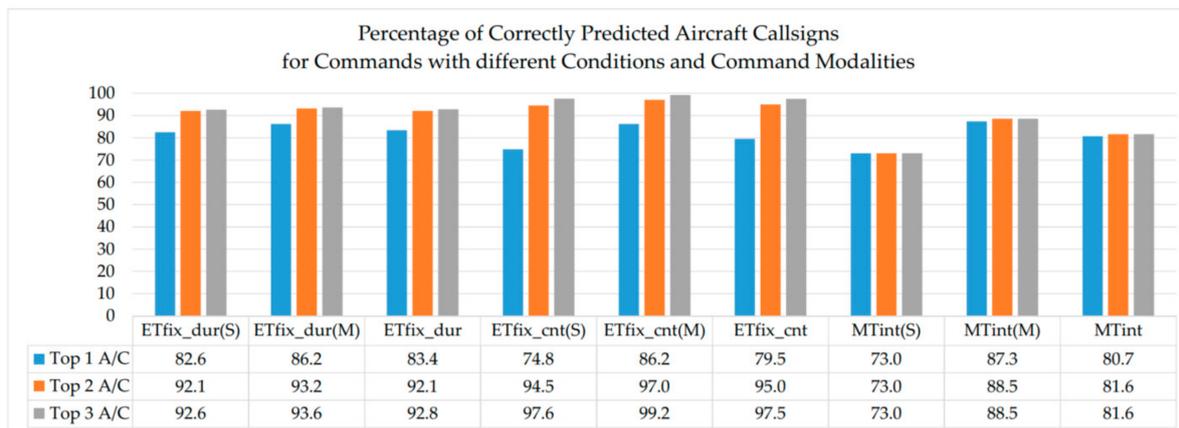
**Table 4.** Example of prediction sets for Top  $N$  A/C based on Table 1.

	$N$ Highest Prediction Probability for Aircraft Callsign	Probability Sum
Top 1 A/C	{DLH5MA}	0.69 <sup>1</sup>
Top 2 A/C	{DLH5MA; UAE57}	0.82 <sup>2</sup>
Top 3 A/C	{DLH5MA; UAE57} <sup>3</sup>	0.82

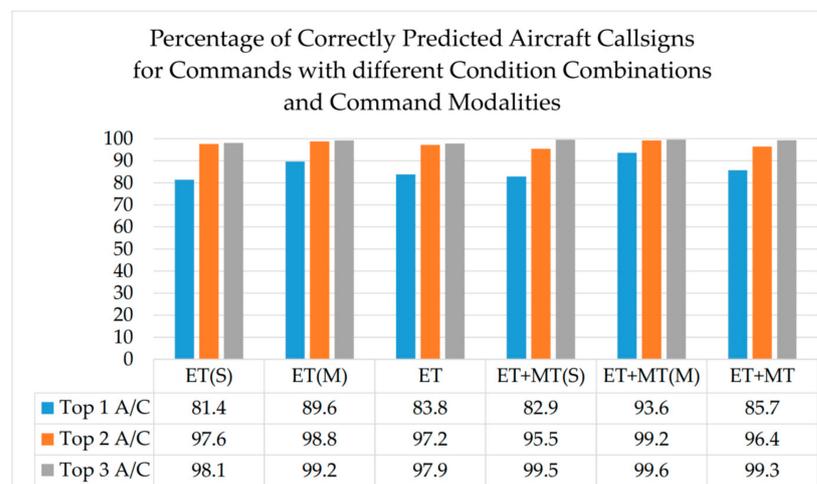
<sup>1</sup>  $3 \times 0.23$  (for the three commands of DLH5MA); <sup>2</sup>  $3 \times 0.23 + 0.13$  (for the three commands of DLH5MA and the one command of UAE57); <sup>3</sup> neither of the three further aircraft {AFR641P; BAW936; KLM1853} is considered for Top 3 as they all have the same overall probability sum of 0.06 in Table 1 and there would be no single choice aircraft.

#### 6.1.2. Accuracy of Aircraft Callsign Prediction for ATC Commands Based on Interaction Data

The percentages of correctly predicted aircraft callsigns for ATC commands based on Top 1/2/3 A/C for the input modalities speech (S), mouse (M), and both combined, considering the five interaction conditions are shown in Figures 6 and 7 for both B runs in average.



**Figure 6.** Correctly predicted aircraft callsigns for ATC commands when considering Top 1/2/3 aircraft for single interaction data conditions per command modality (speech: S; mouse: M; combined).



**Figure 7.** Correctly predicted aircraft callsigns for ATC commands when considering Top 1/2/3 aircraft for combined interaction data conditions per command modality (speech: S; mouse: M; combined).

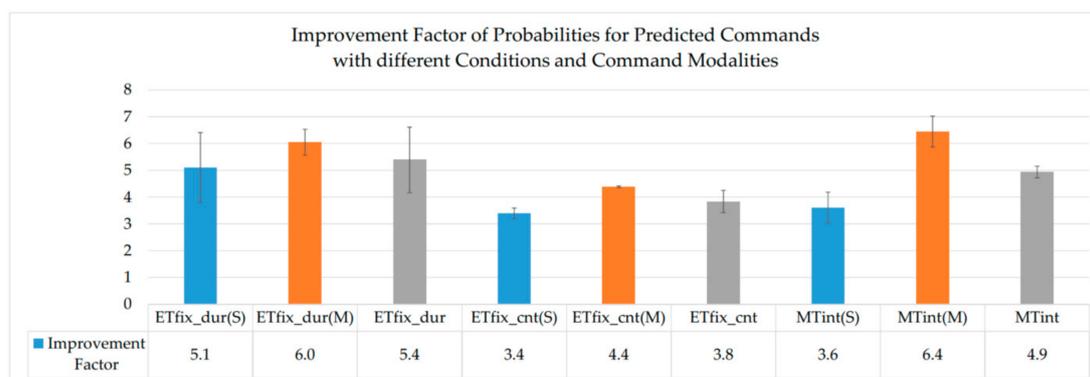
The number of correctly predicted aircraft callsigns increases for the analyzed stand-alone conditions from Top 1 A/C to Top 3 A/C (see Figure 6). The gaze fixation duration metric alone achieves accuracy results above 80% for Top 1 A/C which further increases to around 93% for both Top 2 and Top 3 A/C. The gaze count metric is slightly less accurate in predicting Top 1 A/C as compared to gaze fixation duration metric, but significantly improves the accuracy to around 95% for Top 2 A/C and 98% for Top 3 A/C (see Figure 6). The mouse interaction metric behaves almost in the same for all the three Top A/C categories with accuracies between 73% and 89% (see Figure 6), i.e., the ATCo either has just moved the mouse to the aircraft, which gets the next command or the mouse is not moved at all to that aircraft during the last ten seconds. For all three metrics, aircraft callsigns are predicted more accurately if ATC commands are given via mouse (M) rather than speech (S).

When combining the two eye-tracking metrics or even combining all three interaction metrics, the accuracy of probabilities for aircraft callsign prediction improves significantly (see Figure 7). Independent of the command modality used, from the average values we see that an accuracy rate of 84% and 86% for ET and ET+MT for Top 1 A/C, 97% and 96% for ET and ET+MT for Top 2 A/C, and 98% and 99% for ET and ET+MT for Top 3 A/C was achieved. This implies that the prediction error rates decrease significantly from 16% to 2% (factor of 8 improvement) when Top 3 A/C is predicted as compared to Top 1 A/C for the case when just ET was used. Similarly, when both ET and MT was

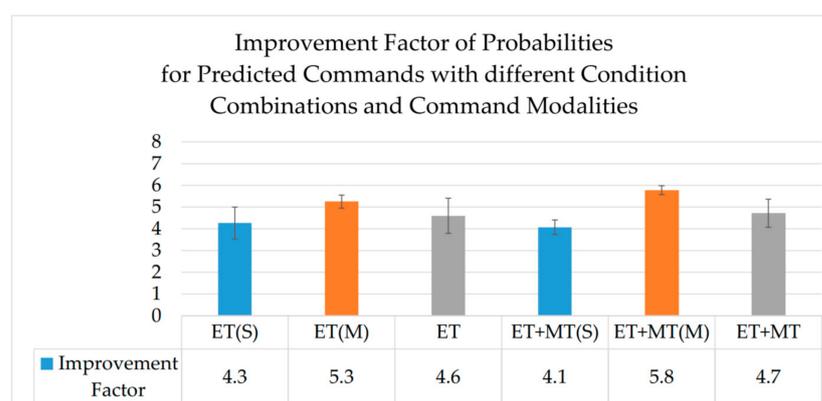
used, the prediction error rates decrease from 14% to 1% (factor of 14 improvement) when Top 3 A/C is predicted as compared to Top 1 A/C. Another impressive result is to compare the prediction error rates for speech modality of the three single modalities for Top 3 A/C of 7.4% (ETfix<sub>dur</sub>), 2.4% (ETfix<sub>cnt</sub>), and 27% (MTint) with the prediction error rate of the combined condition ET+MT(S) of 0.5%—up to a factor of 54 improvement. Overall, it is a factor of 25 improvement when comparing the average prediction error rate of the three single modalities (12.3%) to the combined condition for Top 3 A/C Accuracy.

### 6.1.3. Improvement Factor for Predicted ATC Commands Based on Interaction Data

The improvement factor for all five conditions and command modalities vary between 3.4 and 6.4 as shown in Figures 8 and 9 for both B runs on average. Again, as for the Top A/C analysis, the factor is higher with mouse as command modality. The metrics gaze fixation duration and fixation count achieve improvement factors above 5 and around 4, respectively. The metric mouse interaction is more dependent on the command modality with a factor of 4.9 over all commands. Yet, all the factors illustrated in Figure 8 indicates that the re-assigned probabilities are much better on average as compared to the basic uniform probabilities.



**Figure 8.** Improvement factors for command prediction probabilities for single interaction data conditions per command modality (speech: S; mouse: M; combined) with positive and negative standard deviation of the two average values per run (black lines).



**Figure 9.** Improvement factors for command prediction probabilities for combined interaction data conditions per command modality (speech: S; mouse: M; combined) with positive and negative standard deviation of the two average values per run (black lines).

When combining the eye-tracking metrics and also further integrating the mouse-tracking metric, the average factors for ET and ET+MT are 4.6 and 4.7, respectively.

#### 6.1.4. Detailed Analysis of Specific Results and Discussion on Probability Re-Assignment Quality

Given the above numbers, it is of interest which of the results per condition and per command modality should be interpreted as the core result. As ATCos usually issue commands via speech and the combination of using all three interaction metrics from eye- and mouse-tracking demonstrated to be the most feasible option under the given circumstances, the values for ET+MT(S) should be selected as core results. Thus, an improvement factor of 4.1 (3.7 and 4.4 for the two controllers each per run) is achieved. Furthermore, above 99.5% of aircraft callsigns for ATC commands have been correctly predicted for Top 3 A/C (95.5% for Top 2 A/C and 82.9% for Top 1 A/C). For one ATCo, prediction of Top 3 A/C even reached an accuracy of 100%. For the condition ET+MT(S) with speech command modality, 92% of improvement factors per speech utterance are greater than 1 showing a positive effect of the investigated re-assignment probability implementation.

When correlating Top 1 A/C data from mouse-tracking and eye-tracking, around 66% (two thirds) of predicted aircraft callsigns for ATC commands match, with similar numbers for correct and wrong predictions. When correlating Top 1 A/C data from mouse-tracking and Top 2 A/C data from eye-tracking, 79% of all predicted aircraft callsigns for ATC commands match—83% for correct predictions and 69% for wrong predictions. Hence, there is a slight potential to further filter out wrong predictions by analyzing and comparing single conditions.

The improvement factors for all B-runs analyzed independent of controller, condition, and command modality are always greater than 3 showing a good robustness of the enhanced command prediction probabilities when using ATCo interaction data. The greatest improvement factor for a single run was 7 for one controller in condition with mouse-tracking data only and commands issued via mouse (MTint(M)). If ATCos issue commands via speech, they could basically be looking anywhere. If ATCos issue commands via mouse, they are more or less forced to look at the aircraft radar label and they are definitely forced to move the mouse onto the label to open the intended drop-down menus and select the right values. So, a factor of around 7 seems to indicate the greatest possible factor when considering interaction data. However, the use of mouse-tracking data depends on the CWP and command modality design.

Probabilities of ATC commands derived from interaction data when issuing commands via mouse (ET+MT(M)) and data link can still be used for plausibility checking of command contents. When analyzing all four runs together (2xB, 2xS) for all command modalities and the condition ET, we still achieve 75.1% for Top1 A/C, 90.6% for Top 2 A/C, 93.9% for Top 3 A/C and an improvement factor of 4.1 even if the concept was not intended to be applied on the S-runs.

Some further results for other conditions and modalities are also noteworthy. When considering Top 1 A/C for  $ET_{fix_{dur}}$  in S-runs with commands issued via mouse, there exists no correctly predicted aircraft callsigns. This is a conceptional issue as the commands are only issued after the time for optional manual correction has passed—quite a long time after visually checking the aircraft radar label values inserted via mouse before. The improvement factor and the accuracy increase when the analysis duration is extended, i.e., by looking more into the past to gather interaction data. However, this fact together with the high percentages of B-runs prove the pre-assumptions very well that upcoming ATCo actions are connected to gazes and even non-visual checking is related to hardly any ATCo action concerning a displayed aircraft.

#### 6.1.5. Re-Assigned Prediction Probability Evaluation on Command Type Level

As described in Section 4, the concept of re-assigning prediction probabilities encompasses aircraft callsign level and command type level. However, only aircraft callsign level has been implemented so far. To estimate the further benefits of the command type level, we applied a generalized post-analysis on the command prediction results with re-assigned probabilities. More precisely, we increase the probabilities of command types

that were issued more often and decrease probabilities of command types that were seldom issued. According to the analysis at the beginning of Section 6, we again re-assign the probabilities of the three most often used command types. Thus, for analysis, *DESCEND*, *HEADING*, and *CLEARED ILS* commands have twice as high probability as all other command types for the same aircraft callsign. This reveals an assumed benefit of having different probabilities even for command types.

With this analysis, the improvement factor will further increase by 0.4 when considering different command types for each aircraft callsign. However, it must be mentioned that the analysis approach is just based on statistical incidence, while the concept approach bases on concrete air traffic situations that can be determined via surveillance data. Hence, it is unclear if the improvement factor will in reality be higher or lower than 0.4. Furthermore, it is unclear what the effect on ABSR output will be for command types that occur less frequently, e.g., only less than every tenth command. Though, some of these less frequently occurring command types such as *CONTACT* can be predicted quite reliably in space and time. Hence, it is assumed that a positive influence and an improvement factor increase of more than 0.4 is achievable when implementing the re-assigned probability on command type level.

## 6.2. Using Gazes for Confirmation with Potential Visual Attention Guidance for Speech Recognition Output

In the solution runs 146 ATC commands have been extracted on average from speech utterances. The number of relevant speech utterances is only 123 as often multiple ATC commands were given to aircraft in single utterances. All 123 speech recognition outputs for verbal utterances have been acknowledged via gaze on an aircraft radar label, i.e., the ATCo visually checked one or more at the same time yellow highlighted ABSR output values in a single aircraft radar label. Also, the escalation of saliency levels to enforce the ABSR output check technically worked without any problems. Roughly 120,000 peripheral views on elements at the situation data display have been calculated.

### 6.2.1. Quantitative Questionnaire Results and Discussion

The two subjects rated higher workload for the solution run than for the baseline run, i.e., average Bedford scale workload [98] was 4 for baseline and 7 for solution as well as Raw NASA-TLX scale [99,100] without weighted ratings was 35 for baseline and 51 for solution. The overall score of the system usability scale (SUS) was 77 (range “good”) [101,102]. The ratings for robustness and reliability of the tested system were around the scale mean value. These numbers and the following qualitative feedback should not be generalized given only two study subjects, but can indicate a tendency.

### 6.2.2. Qualitative Questionnaire Results and Discussion

The different frame colors around aircraft radar labels of higher saliency levels seldom appeared for the two subjects as the solution system almost always detected the subjects' gaze at the colored frame in the first saliency level. So, the colors, numbers, and durations of the additional saliency levels could hardly be correctly judged with regards to usefulness. Nevertheless, the eye-tracking based attention guidance for ABSR output was judged to give a medium added value on a scale from very low to very high. Moving and freezing of gazes at a certain aircraft radar label was perceived as physically demanding to some extent. However, the responsiveness of the system given the hardware latency strongly impacted the controlling task in baseline and solution run.

Subjects felt that they had sufficient amount of time to correct the presented ABSR output after the aircraft radar label frame turned green for the confirmation saliency level. The duration for escalating to a higher saliency level should not be changed due to the subjects' ratings. However, the duration of displaying the green aircraft radar label frame in the confirmation saliency level could be reduced. Both subjects voted to decrease the number of different saliency levels. Three different levels are sufficient due to the subjects' opinion. The aircraft radar label frames were found to be unobtrusive, but sometimes

there were too many green frames at the same time, because the ATCo issued many ATC commands in a short amount of time. The maximum number of visible green frames could be reduced to three. The green frames indicate the time to correct the ABSR output after looking at the label. However, the expectation related to a highlighting frame would be that visual attention is required which is not the case. So, it could be a good idea to completely eliminate the green frame when looking away to only let the yellow highlighted ABSR output value remain for a few seconds without an aircraft radar label frame.

After manually clicking check mark and cross in the baseline run subjects felt to have cognitively finished their checking task. This feeling was different for the visual check as the response state, i.e., yellow ABSR output turning white still takes some time as there is still some time remaining for possible correction.

Also, the threshold times for saliency levels could be dependent on the number of highlighted aircraft radar label frames. One subject wished to have check mark and cross even next to the visual ABSR confirmation to be able to return to the default saliency level earlier. Furthermore, parallelly checking ABSR output and pilot readback might be difficult as one or both of them could contain errors and “appear” at the same time. In case of multiple commands in the same transmission or multiple transmissions shortly after each other for the same aircraft it was not clear which elements were already accepted and which were not.

This feedback shows basic feasibility of the visual confirmation concept and implementation without general showstoppers and encourages further advances based on reasonable suggestions.

## 7. Conclusions and Overall Discussion

The four general research objectives have been fulfilled, i.e., (1) eye and mouse movements of ATCos can be recorded and post-processed, (2) relevant information is extracted from such data and integrated into an ABSR system, (3) probabilities for predicted ATCo commands are calculated with good accuracy, and (4) ABSR output can be visually confirmed by ATCos in a CWP system prototype.

Eye- and mouse-tracking were rated to be unobtrusive and important features to easily support ABSR applications with more accurate data and interaction options. Visual confirmation of ABSR output technically worked and confirms that state-of-the-art eye-tracking accuracy is sufficient for applications in various domains and even in the safety-critical ATC domain.

Command prediction probabilities improved by a factor of four on average compared to an existing state-of-research prototype (basic implementation) and included more than 95% of correct aircraft callsigns for Top 2 A/C and even more than 99.5% of correct aircraft callsigns for Top 3 A/C analysis. Thus, Top 2 A/C seems to be sufficient to consider for probability re-assignment even if Top 3 A/C is slightly better. The combination of using all eye-tracking and mouse-tracking metrics together was superior over using some of these metrics alone with an improvement factor for the prediction error rate of 25. This confirms state-of-the-art knowledge that using multiple sensor data is superior to just using single sensor data. To the best of our knowledge no eye- and mouse-tracking based ATC command prediction system or prototype, as well as no visual ASR output confirmation exists in the academic world that could be compared with the results in this paper.

The command predictions support the ABSR engine to reduce command recognition error rates if timely considerable in the search space of the engine. Reduced error rates further enable benefits for speech recognition applications that may lead to reduced workload or increased accuracy of safety net functions. Hence, the concept of visual (and mouse-hover) confirmation should be refined and implementation should be advanced, the concept of re-assigned probabilities based on eye- and mouse-tracking data should be further implemented.

It has to be clearly stated that our one-shot experimental case study without any control group and many possible confounding variables has very low internal validity and

cannot reveal any cause-and-effect relationships. The reported results base on a sample size of just two study subjects and can therefore not be generalized. The reported results might be interpreted as a vague tendency on usefulness of implemented prototypes and indicate that it is worth to move forward with our research from pre-experimental design. Nevertheless, the results presented in this paper tremendously help to design a future broader true experimental design study with randomized groups and clearly defined independent and dependent variables after fixing the reported minor technical issues of the prototypic CWP.

For example, the study design should consider to let all saliency levels appear a number of times to be better judgeable. In addition, the duration of training runs should be extended to reduce the effect of subjects on results with being new and unfamiliar with the elements of the prototypic CWP.

The two controllers had a different professional background, i.e., different number of years of experience as ATCo in approach or tower domain and different experience levels in ATC research. This background and the knowledge about actively participating in a study might have influenced their performance and their reported judgements in a positive or negative way. However, this influencing effect might be bigger for the conceptual element with visual ABSR output confirmation than for the visually nontransparent ATC command prediction rescoreing.

It has also to be noted that the explained pre-assumptions about the connection between visual attention and spot of ATCos' gaze have limitation implications, i.e., the effects of implications are different for different CWPs, ATCos, and other aspects of the working environment. The reported qualitative and quantitative results enable to assess the two implemented techniques in a human-in-the-loop simulation trial with more ATCos in the near future. Then, it can also be determined in detail how much the improved command prediction probabilities help in terms of ASR engine's word error rate, ABSR system's command recognition rate, and further following measures such as ATCo workload when using the system.

All in all, this paper has given first evidence that using further interaction data of a controller working position such as eye-tracking and mouse-tracking can easily enhance existing ATC system prototypes or be integrated in advanced CWP prototypes as demonstrated with functionalities around an Assistant Based Speech Recognition system.

## 8. Outlook on Future Work

The following subsections sketch some future work per each of the two conceptual elements and in general related to CWP interaction.

### 8.1. Outlook on Command Prediction Probability Re-Assignment

Given an improved eye-tracker accuracy, e.g., with advanced devices, it could be checked whether the ATCo looked at, e.g., the label value for current speed of an aircraft. This would lead to an increased likelihood of speed commands for this aircraft or other aircraft being looked at in close timely proximity. The improvement factor for re-assigned ATCo command predictions might be further enhanced if the weighting, e.g., 35%  $ET_{fix_{dur}}$ , 35%  $ET_{fix_{cnt}}$ , and 30%  $MT_{int}$  would be changed dynamically during a simulation run. If it is detected by the mouse-tracker, that the mouse is inactive or the human operator has many eye gaze saccades, the weighting could be adapted.

Legally collecting large amounts of relevant eye- and mouse-tracking data from CWPs—in laboratories or real-life—might be slightly easier than recording radiotelephony utterances due to privacy issues of personal data existing in some countries even if all interaction data could be used in anonymized form to derive patterns and human erroneous behavior. Machine learning on a huge amount of ATC interaction data from eye-tracking, mouse-tracking, and speech recordings could even more automatically individualize re-assigned probabilities for command predictions.

### 8.2. Outlook on ABSR Output Confirmation Mode

Saliency levels should be reduced in their number and re-designed in order to be less intrusive. Taking the existing attention guidance implementation as role model [50], the levels may escalate as follows: The default *transparent* saliency level remains unchanged as well as the first saliency level *white* directly appears with yellow ABSR output values. However, after a few seconds without attention-based trigger, a semi-transparent circle around the aircraft icon should appear. If this visual cue and the white label frame remain undetected, the semi-transparent circle could also receive a flashlight effect for some additional seconds as the highest saliency level. In case the ATCo's attention has been determined to have rested on a highlighted aircraft label, there should be no label frame of any color. The ABSR output value might stay yellow or become another color as visual feedback for checking status for the remaining optional correction time. If the correction time has passed or the highest saliency level duration has passed, all accepted label values turn to white. Furthermore, the optional time for correcting ABSR output should be dependent on the number of aircraft currently under responsibility, i.e., to give the ATCo more time if there are more aircraft to monitor and potential tasks to perform before correcting aircraft radar label input. Also, the time for escalation of saliency levels and the time for optional correction could be made command type specific. In situations of dense air traffic, it might be more important to confirm altitude and heading commands than to confirm *CONTACT* commands.

The feature of visual checking and confirmation via eye gaze could also be applied to other parts of CWPs. One example would be highlighted warnings, e.g., on automatically detected readback errors or medium-term conflict alerts with following escalation and de-escalation via attention guidance mechanisms. Another example is the acknowledgement of the final command in the TriControl prototype via gaze instead of a touch gesture.

### 8.3. Outlook on General Improvements for CWP Interaction

In general, the approximated ATCos' visual attention will be used to assist ATCos in a more convenient way, i.e., giving information at the time and spot that is deemed most reasonable given the current situation. Besides, even further sensors can be included to analyze the ATCos' CWP interaction, e.g., integrate an audio-visual speech recognition system into ABSR.

As a next concrete step, both conceptual techniques will be applied for upcoming ABSR studies in the approach, en-route, and even tower domain.

**Author Contributions:** O.O. was responsible for basic concept, supervision of J.A.'s master's thesis and I.-T.S.'s bachelor's thesis (for both theses including support for concept refinement, literature research, programming, testing, result analysis, etc.), conduction of the study, and concisely writing this article. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data are not publicly available due to included personal data of controllers.

**Acknowledgments:** We like to thank Hartmut Helmke, Shruthi Shetty, Robert Hunger, Michael Finke (all DLR, Germany) for their paper reviews as well as Irina Stefanescu (Technical University Bucharest, Romania) and Norbert Englisch (Technische Universität Chemnitz, Germany) for co-supervising the university theses.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. ICAO. *Air Traffic Services—Air Traffic Control Service, Flight Information Service, Alerting Service, International Civil Aviation Organization (ICAO), Annex 11*; ICAO: Montréal, QC, Canada, 2001.
2. Cardosi, K.M.; Brett, B.; Han, S. *An Analysis of TRACON (Terminal Radar Approach Control) Controller–Pilot Voice Communications, (DOT/FAA/AR-96/66)*; DOT FAA: Washington, DC, USA, 1996.
3. Skaltsas, G.; Rakas, J.; Karlaftis, M.G. An analysis of air traffic controller-pilot miscommunication in the NextGen environment. *J. Air Transp. Manag.* **2013**, *27*, 46–51. [[CrossRef](#)]
4. ICAO; ATM (Air Traffic Management). *Procedures for Air Navigation Services*; International Civil Aviation Organization (ICAO), DOC 4444 ATM/501; ICAO: Montréal, QC, Canada, 2007.
5. Lin, Y. Spoken Instruction Understanding in Air Traffic Control: Challenge, Technique, and Application. *Aerospace* **2021**, *8*, 65. [[CrossRef](#)]
6. 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.
7. Ohneiser, O.; Jauer, M.-L.; Rein, J.R.; Wallace, M. Faster Command Input Using the Multimodal Controller Working Position “TriControl”. *Aerospace* **2018**, *5*, 54. [[CrossRef](#)]
8. Ohneiser, O.; Sarfjoo, S.; Helmke, H.; Shetty, S.; Motlicek, P.; Kleinert, M.; Ehr, H.; Murauskas, Š. Robust Command Recognition for Lithuanian Air Traffic Control Tower Utterances. In Proceedings of the InterSpeech 2021, Brno, Czech Republic, 30 August–3 September 2021.
9. Connolly, D.W. Voice Data Entry in Air Traffic Control. In Proceedings of the Voice Technology for Interactive Real-Time Command/Control Systems Application, N93-72621, Moffett Field, CA, USA, 6–8 December 1977; pp. 171–196.
10. Young, S.R.; Ward, W.H.; Hauptmann, A.G. Layering predictions: Flexible use of dialog expectation in speech recognition. In Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI89), Morgan Kaufmann, Detroit, MI, USA, 20–25 August 1989; pp. 1543–1549.
11. Helmke, H.; Sloty, M.; Poiger, M.; Herrer, D.F.; Ohneiser, O.; Vink, N.; Cerna, A.; Hartikainen, P.; Josefsson, B.; Langr, D.; et al. 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.
12. Rataj, J.; Helmke, H.; Ohneiser, O. AcListant with Continuous Learning: Speech Recognition in Air Traffic Control. In *Air Traffic Management and Systems IV, Selected Papers of the 6th ENRI International Workshop on ATM/CNS (EIWAC2019)*; Springer: Singapore, 2021; pp. 93–109.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Ohneiser, O.; Helmke, H.; Ehr, H.; Gürlük, H.; Hössl, M.; Mühlhausen, T.; Oualil, Y.; Schulder, M.; Schmidt, A.; Khan, A.; et al. Air Traffic Controller Support by Speech Recognition. In *Advances in Human Aspects of Transportation: Part II, Proceedings of the International Conference on Applied Human Factors and Ergonomics (AHFE), Krakow, Poland, 19–23 July 2014*; Stanton, N., Landry, S., Di Bucchianico, G., Vallicelli, A., Eds.; CRC Press: Boca Raton, FL, USA; pp. 492–503.
20. Updegrove, J.A.; Jafer, S. Optimization of Air Traffic Control Training at the Federal Aviation Administration Academy. *Aerospace* **2017**, *4*, 50. [[CrossRef](#)]
21. Schäfer, D. Context-Sensitive Speech Recognition in the Air Traffic Control Simulation. Ph.D. Thesis, University of Armed Forces, Munich, Germany, 2001.
22. Kleinert, M.; Helmke, H.; Siol, G.; Ehr, H.; Finke, M.; Srinivasamurthy, A.; Oualil, Y. Machine Learning of Controller Command Prediction Models from Recorded Radar Data and Controller Speech Utterances. In Proceedings of the 7th SESAR Innovation Days, Belgrade, Serbia, 28–30 November 2017.

23. Helmke, H.; Kleinert, M.; Ohneiser, O.; Ehr, H.; Shetty, S. Machine Learning of Air Traffic Controller Command Extraction Models for Speech Recognition Applications. In Proceedings of the IEEE/AIAA 39th Digital Avionics Systems Conference (DASC), Virtual, 11–16 October 2020.
24. SESAR2020-Exploratory Research Project HAAWAI (Highly Automated Air Traffic Controller Workstations with Artificial Intelligence Integration). Available online: <https://www.hawaii.de> (accessed on 19 August 2021).
25. Ohneiser, O.; Helmke, H.; Shetty, S.; Kleinert, M.; Ehr, H.; Murauskas, Š.; Pagirys, T. Prediction and extraction of tower controller commands for speech recognition applications. *J. Air Transp. Manag.* **2021**, *95*, 102089. [[CrossRef](#)]
26. Kleinert, M.; Helmke, H.; Moos, S.; Hlousek, P.; Windisch, C.; Ohneiser, O.; Ehr, H.; Labreuil, A. Reducing Controller Workload by Automatic Speech Recognition Assisted Radar Label Maintenance. In Proceedings of the 9th SESAR Innovation Days, Athens, Greece, 2–5 December 2019.
27. Nguyen, V.N.; Holone, H. N-best list re-ranking using syntactic score: A solution for improving speech recognition accuracy in air traffic control. In Proceedings of the 16th International Conference on Control, Automation and Systems (ICCAS), Gyeongju, Korea, 16–19 October 2016; pp. 1309–1314.
28. Shore, T.; Faubel, F.; Helmke, H.; Klakow, D. Knowledge-Based Word Lattice Rescoring in a Dynamic Context. In Proceedings of the Inter Speech 2012, Portland, OR, USA, 9–13 September 2012.
29. Punde, P.A.; Jadhav, M.E.; Manza, R.R. A study of Eye Tracking Technology and its applications. In Proceedings of the 1st International Conference on Intelligent Systems and Information Management (ICISIM), Maharashtra, India, 5–6 October 2017; pp. 86–90.
30. Farnsworth, B. What Is Eye Tracking and How Does It Work? Available online: <https://imotions.com/blog/eye-tracking-work/> (accessed on 19 August 2021).
31. Farnsworth, B. 10 Most Used Eye Tracking Metrics and Terms. Available online: <https://imotions.com/blog/10-terms-metrics-eye-tracking/> (accessed on 19 August 2021).
32. Bhattarai, R.; Phothisonothai, M. Eye-Tracking Based Visualizations and Metrics Analysis for Individual Eye Movement Patterns. In Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, 10–12 July 2019; pp. 381–384.
33. Poole, A.; Ball, L.J. Eye tracking in human-computer interaction and usability research: Current status and future prospects. In *Encyclopedia of Human Computer Interaction*; Idea Group Reference: Hershey, PA, USA, 2006; pp. 211–219.
34. Salvucci, D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2000, Palm Beach Gardens, FL, USA, 6–8 November 2000.
35. Scholz, A. Eye Movements, Memory, and Thinking—Tracking Eye Movements to Reveal Memory Processes during Reasoning and Decision-Making. Ph.D. Thesis, Technische Universität Chemnitz, Chemnitz, Germany, 2015.
36. Lorigo, L.; Haridasan, M.; Brynjarsdóttir, H.; Xia, L.; Joachims, T.; Gay, G.; Granka, L.; Pellacini, F.; Pan, B. Eye Tracking and Online Search: Lessons Learned and Challenges Ahead. *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 1041–1052. [[CrossRef](#)]
37. Fraga, R.P.; Kang, Z.; Crutchfield, J.M.; Mandal, S. Visual Search and Conflict Mitigation Strategies Used by Expert en Route Air Traffic Controllers. *Aerospace* **2021**, *8*, 170. [[CrossRef](#)]
38. Kang, Z.; Mandal, S.; Dyer, J. Data Visualization Approaches in Eye Tracking to Support the Learning of Air Traffic Control Operations. In Proceedings of the National Training Aircraft Symposium, Daytona Beach, FL, USA, 14–16 August 2017.
39. Wickens, C.; Hollands, J.; Banbury, S.; Parasuraman, R. *Engineering Psychology and Human Performance*, 4th ed.; Pearson Education: Boston, MA, USA, 2013.
40. Zamani, H.; Abas, A.; Amin, M.K.M. Eye Tracking Application on Emotion Analysis for Marketing Strategy. *J. Telecommun. Electron. Comput. Eng.* **2016**, *8*, 87–91.
41. Goyal, S.; Miyapuram, K.P.; Lahiri, U. Predicting Consumer’s Behavior Using Eye Tracking Data. In Proceedings of the 2nd International Conference on Soft Computing and Machine Intelligence (ISCMI), Hong Kong, China, 23–24 November 2015; pp. 126–129.
42. Sari, J.N.; Nugroho, L.; Santosa, P.; Ferdiana, R. The Measurement of Consumer Interest and Prediction of Product Selection in E-commerce Using Eye Tracking Method. *Int. J. Intell. Eng. Syst.* **2018**, *11*, 30–40. [[CrossRef](#)]
43. Huang, C.-M.; Andrist, S.; Sauppé, A.; Mutlu, B. Using gaze patterns to predict task intent in collaboration. *Front. Psychol.* **2015**, *6*, 1049. [[CrossRef](#)]
44. Eivazi, S.; Bednarik, R. Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction, Palo Alto, CA, USA, 13 February 2011.
45. Duchowski, A.T. A breadth-first survey of eye-tracking applications. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 455–470. [[CrossRef](#)]
46. 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.
47. 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.
48. 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. [[CrossRef](#)] [[PubMed](#)]

49. 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.
50. Ohneiser, O.; Gürlük, H.; Jauer, M.-L.; Szöllösi, Á.; Balló, D. Please have a Look here: Successful Guidance of Air Traffic Controller's Attention. In Proceedings of the 9th SESAR Innovation Days, Athens, Greece, 2–5 December 2019.
51. Rataj, J.; Ohneiser, O.; Marin, G.; Postaru, R. Attention: Target and Actual—The Controller Focus. In Proceedings of the 32nd Congress of the International Council of the Aeronautical Sciences (ICAS), Shanghai, China, 6–10 September 2021.
52. Ohneiser, O.; Jauer, M.-L.; Gürlük, H.; Springborn, H. Attention Guidance Prototype for a Sectorless Air Traffic Management Controller Working Position. In Proceedings of the German Aerospace Congress DLRK, Friedrichshafen, Germany, 4–6 September 2018.
53. Di Flumeri, G.; De Crescenzo, F.; Berberian, B.; Ohneiser, O.; Kraemer, J.; Aricò, P.; Borghini, G.; Babiloni, F.; Bagassi, S.; Piastra, S. Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing with Highly Automated Systems. *Front. Hum. Neurosci.* **2019**, *13*, 1–17. [[CrossRef](#)]
54. 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.
55. Rheem, H.; Verma, V.; Becker, D.V. Use of Mouse-tracking Method to Measure Cognitive Load. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Philadelphia, PA, USA, 1–5 October 2018; Volume 62, pp. 1982–1986.
56. Huang, J.; White, R.; Buscher, G. User see, user point: Gaze and cursor alignment in web search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), Austin, TX, USA, 5–10 May 2012; ACM: New York, NY, USA, 2012; pp. 1341–1350.
57. Zgonnikov, A.; Aleni, A.; Piironen, P.T.; O'Hora, D.; di Bernardo, M. Decision landscapes: Visualizing mouse-tracking data. *R. Soc. Open Sci.* **2017**, *4*, 170482. [[CrossRef](#)] [[PubMed](#)]
58. Calcagni, A.; Lombardi, L.; Sulpizio, S. Analyzing spatial data from mouse tracker methodology: An entropic approach. *Behav. Res.* **2017**, *49*, 2012–2030. [[CrossRef](#)] [[PubMed](#)]
59. Maldonado, M.; Dunbar, E.; Chemla, E. Mouse tracking as a window into decision making. *Behav. Res.* **2019**, *51*, 1085–1101. [[CrossRef](#)] [[PubMed](#)]
60. Krassanakis, V.; Kesidis, A.L. MatMouse: A Mouse Movements Tracking and Analysis Toolbox for Visual Search Experiments. *Multimodal Technol. Interact.* **2020**, *4*, 83. [[CrossRef](#)]
61. Claypool, M.; Le, P.; Wased, M.; Brown, D. Implicit interest indicators. In Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI'01), Santa Fe, NM, USA, 14–17 January 2001; ACM: New York, NY, USA, 2001; pp. 33–40.
62. Rodden, K.; Fu, X.; Aula, A.; Spiro, I. Eye-mouse coordination patterns on web search results pages. In Proceedings of the CHI '08 Extended Abstracts on Human Factors in Computing Systems, Florence, Italy, 5–10 April 2008.
63. Chen, M.C.; Anderson, J.R.; Sohn, M.H. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In Proceedings of the CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01), Seattle, WA, USA, 31 March–5 April 2001; ACM: New York, NY, USA, 2001; pp. 281–282.
64. Cooke, N.; Shen, A.; Russell, M. Exploiting a 'gaze-Lombard effect' to improve ASR performance in acoustically noisy settings. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 1754–1758.
65. Cooke, N.; Russell, M. Gaze-contingent ASR for spontaneous, conversational speech: An evaluation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 4433–4436.
66. Shen, A. The Selective Use of Gaze in Automatic Speech Recognition. Ph.D. Thesis, College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK, 2013.
67. Alhargan, A.; Cooke, N.; Binjammaz, T. Multimodal affect recognition in an interactive gaming environment using eye tracking and speech signals. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17), Glasgow, UK, 13–17 November 2017; ACM: New York, NY, USA; pp. 479–486.
68. Rasmussen, M.; Tan, Z. Fusing eye-gaze and speech recognition for tracking in an automatic reading tutor—A step in the right direction? In Proceedings of the Speech and Language Technology in Education (SLaTE), Grenoble, France, 30 August–1 September 2013.
69. DLR Institute of Flight Guidance. TriControl—Multimodal ATC Interaction. Available online: [http://www.dlr.de/fl/Portaldata/14/Resources/dokumente/veroeffentlichungen/TriControl\\_web.pdf](http://www.dlr.de/fl/Portaldata/14/Resources/dokumente/veroeffentlichungen/TriControl_web.pdf) (accessed on 19 August 2021).
70. 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.
71. Ohneiser, O.; Biella, M.; Sch mugler, A.; Wallace, M. Operational Feasibility Analysis of the Multimodal Controller Working Position "TriControl". *Aerospace* **2020**, *7*, 15. [[CrossRef](#)]
72. Bernsen, N. Multimodality Theory. In *Multimodal User Interfaces. Signals and Communication Technologies*; Tzovaras, D., Ed.; Springer: Berlin/Heidelberg, Germany, 2008.
73. 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.

74. 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.
75. Oviatt, S.L. Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems. *Adv. Comput.* **2002**, *56*, 305–341.
76. Oviatt, S.L. Multimodal interactive maps: Designing for human performance. *Hum. Comput. Interact.* **1997**, *12*, 93–129.
77. Cohen, P.R.; McGee, D.R. Tangible multimodal interfaces for safety-critical applications. *Commun. ACM* **2004**, *1*, 1–46. [[CrossRef](#)]
78. 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.
79. Oviatt, S. User-centered modeling for spoken language and multimodal interfaces. *IEEE Multimed.* **1996**, *4*, 26–35. [[CrossRef](#)]
80. Den Os, E.; Boves, L. User behaviour in multimodal interaction. In Proceedings of the HCI International, Las Vegas, NV, USA, 22–27 July 2005.
81. 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.
82. 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.
83. 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. [[CrossRef](#)]
84. Pentland, A. Perceptual Intelligence. *Commun. ACM* **2000**, *4*, 35–44. [[CrossRef](#)]
85. Oviatt, S.L. Ten myths of multimodal interaction. *Commun. ACM* **1999**, *11*, 74–81. [[CrossRef](#)]
86. 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.
87. 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.
88. Neßelrath, R.; Moniri, M.M.; Feld, M. Combining Speech, Gaze, and Micro-gestures for the Multimodal Control of In-Car Functions. In Proceedings of the 12th International Conference on Intelligent Environments (IE), London, UK, 14–16 September 2016; pp. 190–193.
89. 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.
90. 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.
91. SESAR Joint Undertaking. *European ATM Master Plan—Digitalising Europe's Aviation Infrastructure*; SESAR Joint Undertaking: Brussels, Belgium; Luxembourg, 2020.
92. SESAR2020-Industrial Solution PJ.16-04. Controller Working Position/Human Machine Interface-CWP/HMI. Available online: <https://www.sesarju.eu/projects/cwphmi> (accessed on 19 August 2021).
93. Ohneiser, O. *RadarVision-Manual for Controllers*, Original German Title: *RadarVision-Benutzerhandbuch für Lotsen*; Internal Report 112-2010/54; German Aerospace Center (DLR), Institute of Flight Guidance: Braunschweig, Germany, 2010.
94. Salomea, I.-T. Integration of Eye-Tracking and Assistant Based Speech Recognition for the Interaction at the Controller Working Position. Bachelor's Thesis, "Politehnica" University of Bucharest, Bucharest, Romania, 2021.
95. Wickens, C.D.; McCarley, J.S. *Applied Attention Theory*; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2008.
96. Adamala, J. Integration of Eye Tracker and Assistant Based Speech Recognition at Controller Working Position. Master's Thesis, Technische Universität Chemnitz, Chemnitz, Germany, 2021.
97. Ribeiro, M.; Ellerbroek, J.; Hoekstra, J. Review of Conflict Resolution Methods for Manned and Unmanned Aviation. *Aerospace* **2020**, *7*, 79. [[CrossRef](#)]
98. Roscoe, A.H. Assessing pilot workload in flight. In Proceedings of the AGARD Conference Proceedings Flight Test Techniques, Lisbon, Portugal, 2–5 April 1984.
99. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; North Holland Press: Amsterdam, The Netherlands, 1988; Volume 198.
100. Hart, S.G. Nasa-Task Load Index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society, San Francisco, CA, USA, 16–20 October 2006; Volume 50, pp. 904–908.
101. Brooke, J. SUS-A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B.A., Eds.; Taylor and Francis: London, UK, 1996; pp. 189–194.
102. 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. [[CrossRef](#)]