Fourteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2021)

Embodied Multimodal Interaction with a Portable Mixed Reality-based Digital Tower

Pallavi Mohan, Sameer Alam,
T.N. Mohammed Nadirsha and Nimrod Lilith
Saab-NTU Joint Lab,
Nanyang Technological University Singapore,
Email: {pallavi.mohan | sameeralam | m170013@e | nimrod.lilith}@ntu.edu.sg

Åsa Svensson
Department of Research and Innovation,
LFV Air Navigation Services of Sweden,
Norrköping, Sweden
Email: asa.e.svensson@lfv.se

***Abstract*—A major topic of air traffic research in the past decade has been of digital towers (DTs) that replace physical air traffic control towers by reproducing live camera feeds of out-of-tower (OOT) views using massive immersive spatial displays. With the recent pandemic-induced decline in air traffic, the quest to introduce cost-effective alternative solutions for DTs seems especially urgent. Recent developments in mixed reality (MR)-based visualizations have allowed complex spatial data to be better represented in a stereoscopic virtual space. Moreover, the immersion that MR offers is highly customizable, allowing controllers to maintain their view of the environment along with the OOT view, similar to a physical tower. Additionally, the increasingly robust technical capabilities of self-contained MR-based Head-Mounted Devices (HMDs) (such as Microsoft HoloLens 2) gives rise to the possibility of designing a cohesive and self-contained DT system on such a platform, making the DT system cost-effective, portable and interactive. In this paper, we have developed a simulated Digital Tower environment in mixed reality using Microsoft Hololens 2, with intuitive embodied and multimodal interactive capabilities. The system incorporates several visual features from real-world air traffic control environment and visualizes them in mixed reality with added interactive capabilities for aircrafts, gates and runways. We then assess the preliminary system with experienced air traffic controllers and our findings show that such a mixed reality representations of air traffic control is accepted by controllers as a useful way to visualize information on the OOT view. We also find that the visual and interactive affordances provided by the system shows promising potential to augment, and possibly replace present digital tower set-ups as a self-contained, interactive and portable DT system.**

***Keywords*—digital tower, remote tower, airport management, mixed reality, multimodal interaction, eye tracking.**

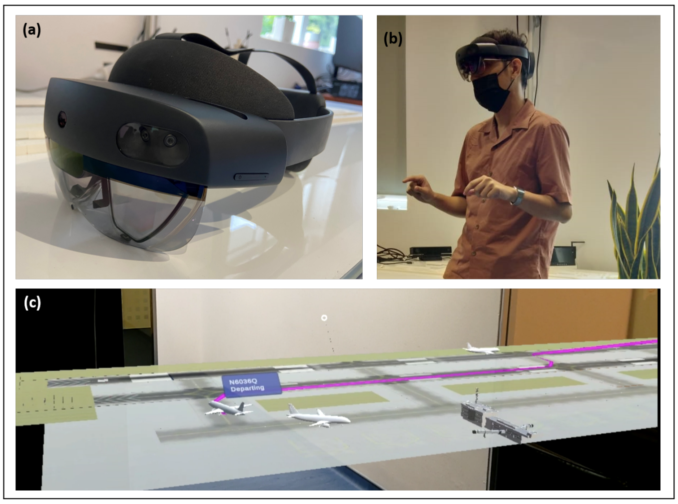

Figure 1: (a) Microsoft HoloLens 2 (b) An Air Traffic Controller wearing the HoloLens 2 (c) The DT simulation in Mixed Reality with overlaid aircraft billboards.

I. Introduction

With recent advances in mixed reality (MR) systems and rendering, it is increasingly possible to replicate a complex real-world system into a high-fidelity holographic representation in mixed reality. Such representations offer numerous advantages over conventional 2D, 3D and even virtual reality (VR) representations, primarily owing to the property of see-through immersion that MR offers. Particularly in contrast with VR environments, MR allows users to maintain their view of the physical world while viewing realistic immersive content. Air traffic control, a spatially complex system that facilitates safety-critical operations by air traffic controllers, is one pertinent area that benefits from a digital representation [1]. Research in the past decade has focused on Digital Towers (DTs) that replace physical towers. One innovation of such digital tower environments is the significantly altered visualization of live camera Out-Of-Tower (OOT) view feeds by using immersive screens in a Cave Automatic Virtual Environment (CAVE)-like set-up that is not co-located with the physical tower. While DTs have made it convenient for multiple aerodromes to be controlled from a single, remote location, they still require the use of multiple expensive spatial displays that can span entire walls and even rooms. With the current pandemic situation causing a sharp decline in air traffic, the quest to introduce cost-effective alternative solutions for air traffic control seems more urgent than ever. Nonetheless, the air traffic controller community has been hesitant to adopt such digital remote set-ups due to the unfamiliarity of the environment [2]. Novel forms of visualization such as MR- and VR-based visualizations allow complex spatial data to be better represented in a stereoscopic virtual space [2] [3]. Moreover, the increasingly robust technical capabilities of self-contained MR-based Head-Mounted Devices (HMDs) such as Microsoft HoloLens 2 (Fig. 1(a) and (b)) give rise to the possibility of a cohesive DT system that is cost-effective, portable and interactive. This paper, therefore, proposes an interactive MR-based DT system that supports instinctual multimodal eye gaze- and hand gesture-based interactions with individual aircraft, gates and runways (shown in Fig. 1(c)), with the possibility of potentially replacing entire massive spatial displays and peripheral systems in a remote DT set-up.

Most DT systems today are either passive or require only menu-driven point-and-click interactions with visual elements onscreen. In contrast, the proposed MR-based system integrates multimodal embodied interactive capabilities for air traffic controllers to intuitively interact with the system for selective information retrieval. Furthermore, to maintain controller performance, interactions with the system employ natural, embodied and multimodal techniques that aim to improve situational awareness in such a safety-critical setting.

Since controllers are familiar with interacting with air traffic control systems by hand-based inputs, the proposed system employs natural hand gesture-based interactions. However, in such a task-critical environment, relying on hand gestures alone may induce fatigue due to the repetitive nature of actions and the wide spatially complex area on which they will be performed. To mitigate this effect, an additional input modality is employed to assist the hand gesture. In particular, the controller's eye gaze location is utilized to lend spatial context awareness to their hand gesture, eliminating the need to perform convoluted gestures to execute simple primary input actions rapidly (such as quickly switching between two different aircraft routes) and enabling quick, natural hand gestures to perform necessary actions. This is intended to reduce input action performance time and effort, thus freeing the air traffic controller's cognitive resources for situation awareness. Therefore, the proposed MR-based digital tower system incorporates embodied multimodal interaction that is rapid and intuitive as a part of its system design, with the additional goal of smoothing the transition and increasing air traffic controller acceptance of the new system. By employing such a self-contained mobile DT system, operators can aim to completely replace massive spatial displays, thus reducing operational costs as well. The findings from this paper, therefore, can be used to guide the design of a large-scale interactive MR DT system in the future that displays live OOT views with augmented information on a self-contained holographic device.

The paper's key contributions can be summarized as follows:

1) We have partially replicated a DT system in an MR environment using Microsoft HoloLens 2 by building a novel 3D simulation system of the OOT view in a mixed reality space;
2) We have incorporated embodied interaction capabilities into this system by defining hand gestures for selective information retrieval by controllers;
3) We have integrated multimodality into the interaction design by leveraging information of the controller's eye gaze to provide spatial context awareness in the MR environment; and
4) The system is preliminarily assessed for acceptance by experienced air traffic controllers, who found it to be highly intuitive and visually effective for information retrieval.

The following sections present the background of this research, the design of this system, an early controller acceptance assessment, results, discussions and future directions.

II. Related Work

Digital towers have been a major topic in air traffic control (ATC) research in recent years, with emerging hardware giving rise to novel ways to visualize the air traffic control system and new dimensions of research in the field [4] [5]. Prototypes of ATC systems have been developed to facilitate research on interactive interfaces, such as in NAVSIM, an air traffic simulator designed to test mouse-based controller interactions [6].

In recent years, Single European Sky ATM Research (SESAR) has undertaken a landmark project that leveraged existing AR/VR tools to develop a synthetic vision system for air traffic control, called RETINA [7]. The RETINA system aimed to reduce the head down time of air traffic controllers while maintaining their performance by utilizing HMDs to augment information that controllers would normally obtain by looking down at additional computer displays. However, while the overall workload was found to be reduced when compared to conventional systems, the physical workload was compromised as a result of the unfamiliar environment for controllers. Moreover, the system offered a passive OOT view with static displays of information. This led to visual clutter due to an 'always-on' model of overlaid information. Additionally, at the time of implementation, the AR display technology was not mature enough to be deployed in a real-world setting and suffered from several disadvantages, such as visual clutter etc.

In contrast, this paper envisions an ATC on-demand service anywhere - it proposes a self-contained Mixed Reality DT system with see-through immersion which can be operated from any location. In addition to displaying a simulation of the OOT view inside the Mixed Reality headset, interactive capabilities will be added to the system, wherein controllers can use natural, embodied, and multimodal interactions to perform input actions such as highlighting aircraft routes, and view flight information quickly and intuitively.

There have been recent works that explore holographic representations of air traffic using MR-based devices. MR offers several affordances that make it a suitable medium for air traffic control representation. While it offers the wide Field-Of-View (FOV) inherent in stereoscopic visualizations, it is also essentially a see-through environment, unlike an immersive VR environment. This means that air traffic controllers can visualize the aerodrome better in such a set-up without impeding their view of other personnel in the room, allowing greater collaboration and communication. The self-contained nature of an HMD further allows individual air traffic controllers to be offered additional augmented

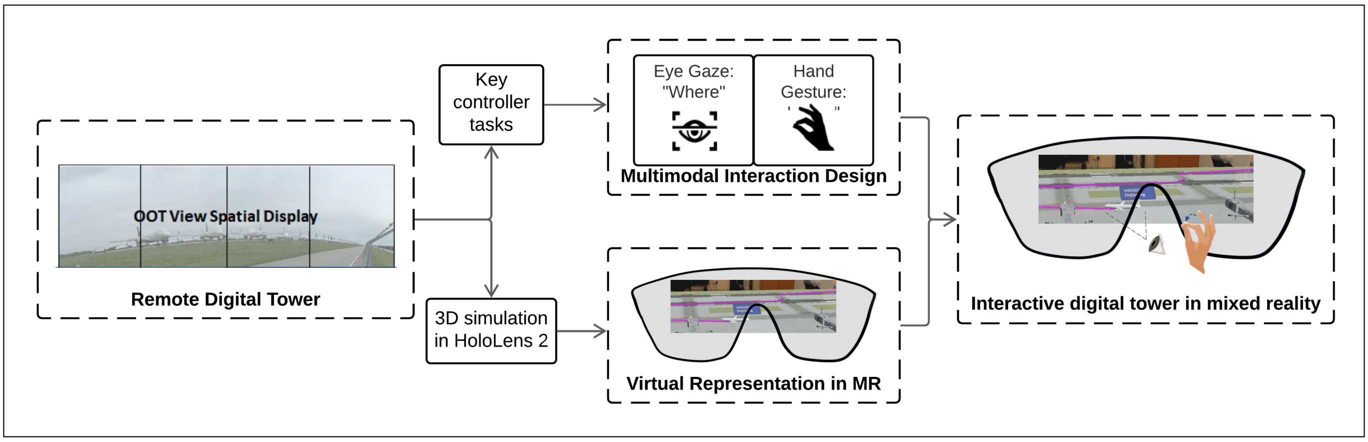

Figure 2: Concept of the interactive MR-based DT system.

information and visuals specific to their authority level and role, within their own device. There have been recent works that propose holographic air traffic control representations [8] [9]. In particular, the system proposed by Han et al. in 2019 is a preliminary 4D MR system using gestural and voice interaction through HoloLens 1, to control both air and ground traffic. However, the interaction capabilities of this system are primarily limited to voice commands for menu-driven input selection [10]. Voice-based interaction poses a risk when interacting with an ATC system as controllers use voice as a primary mode of communication with pilots, and any additional interaction that requires voice as a primary input modality can potentially interfere with this critical communication channel. As a result, we exclude voice as an interaction modality in our design to allow controllers to reserve voice as a channel used solely to communicate with pilots without the secondary function of acting as an input modality to the system.

In the post-pandemic era, reducing the operational and maintenance costs of entire DT systems is crucial in ensuring a resurgence of the aviation industry, which is still re-emerging from the effects of plummeting air traffic. This further motivates the possibility of cost-effective alternatives such as a self-contained MR-based DT system replacing entire spatial displays. Such MR systems cost significantly less than conventional spatial displays. In addition to the cost savings brought about by the replacement of spatial displays, the proposed system also has the potential to render the following redundant: separate computer displays for radar and weather data, interaction devices with the system, and even a dedicated digital tower room.

However, the survey of prevailing research in such DTs reveals a general hesitance to adopt an entirely new technical ecosystem for a safety-critical environment such as air traffic control [2]. Moreover, a finding by Chang and Yeh states that controllers are seen to perform ATC tasks more effectively when ATC systems incorporate changes that consider better interaction [11]. As a result, this research further posits the ability of better visualization and interaction design to smooth this digital transformation for air traffic controllers.

III. The Digital Tower System in Mixed Reality

This paper proposes a digital tower system in a mixed reality space to realize the advantages of a self-contained portable and interactive system such as Microsoft HoloLens 2. Hence, a 3D simulation of an OOT view is designed and developed as a mixed reality application on the HoloLens headset, with added interactive capabilities. This is designed as a simulated prototype of a digital tower system that contains the OOT view and information overlays to facilitate an early assessment of the interactive features of the system and their acceptance by air traffic controllers. A concept of the system is depicted in Fig. 2.

A. Air Traffic Visualization in Mixed Reality

This paper identifies two primary components of an air traffic control tower system that can be visualized in a mixed reality space:

1) the OOT view, either via a physical window or a digital screen; and
2) additional information, which primarily includes aircraft data in the form of flight strips.

Our prototype features a simulation of these two primary components, and integrates multimodal interactive capabilities to both these components. An overview of the various components that make up this system is detailed in Fig. 3.

1) Hardware: Microsoft HoloLens 2 is an untethered holographic computer that contains a dedicated Holographic Processing Unit (HPU) [12]. It contains several integrated hardware sensors that are used for various tracking purposes such as 6-DoF positional head tracking, hand tracking and gesture recognition, and eye gaze tracking.

2) Software: The system is built using Unity3D and the Mixed Reality Toolkit (MRTK) Software Development Kit (SDK) provided by Microsoft for developing applications on HoloLens 2 [13] [14]. The airport and aircraft models are constructed in 3D using purchased assets from Unity. The airport has a minimal set-up consisting of 1 terminal, 2 runways, taxiways, 2 gates and 5 aircraft in different phases.

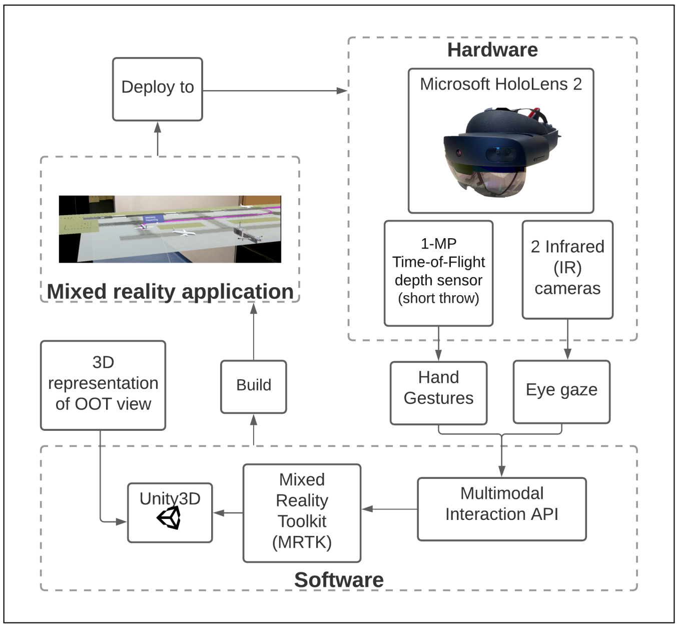

Figure 3: The design of the hardware and software components of the system.

B. Billboard Design

To facilitate seamless information overlays, we attach 'billboards' to each aircraft. Billboards are labels containing aircraft information, such as their callsign, that attach themselves to the aircraft. Co-locating aircraft billboards with the relevant aircraft allows controllers to switch between information retrieval and visual inspection seamlessly. This is in contrast with a conventional tower set-up, where the OOT view is physically placed above information displays and flight strips containing detailed flight information. The spatial target context is informed by the controller's eye gaze, and this information is used to selectively display only information relevant to the controller's current visual focal area. By spatially anchoring billboards containing selective context-aware information, the system aims to achieve: 1) reduced visual clutter; and 2) reduced head-down time for controllers. The preliminary controller acceptance assessment conducted on this system further investigates controllers' perceptions of the effects of such a design.

This system implements two billboard models based on a hierarchical level of abstraction. The design of the billboards followed an iterative model based on several rounds of design and informal discussions with air traffic controllers. Since an aircraft billboard was conceived as a conceptual equivalent of the conventional flight strip, the various components of a standard flight strip were taken as guidelines to design the holographic billboard. Furthermore, from literature, it is often seen that a repercussion of overlaying data is visual clutter; thus the design needs to display sufficient flight data to inform a controller's decision without contributing to visual clutter [7].

To overcome this, we propose a hierarchical model of representing flight data on billboards. In this, each aircraft billboard works at two levels of abstraction and is activated by two different input modalities, as explained:

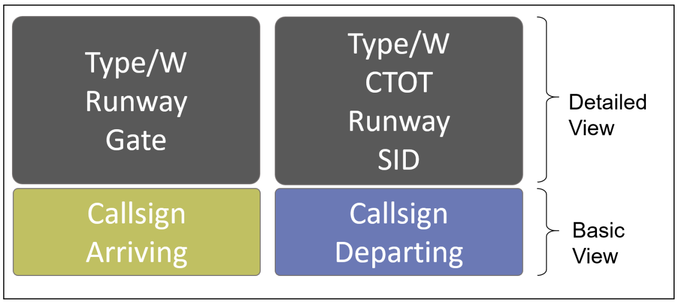

Figure 4: Billboard designs of arriving and departing aircraft.

1) Basic View: This view is activated by the controller's eye gaze location. The basic billboard view is colored according to standard arrival and departure flight strip colors. Hence, billboards of arriving aircraft are colored yellow, while billboards of departing aircraft are colored blue. This billboard contains minimal details; these include the aircraft's callsign and whether it is an arriving or departing aircraft.

2) Detailed View: This view is activated by the controller performing an explicit but simple 'air tap' hand gesture; details of the gesture will be explained in Section IV. The organization of the billboard is as seen in Fig. 4

IV. Interaction Design

The system is designed to contain several types of visual information. However, displaying all available information at all times on a single display would cause added visual clutter, detracting from the OOT view. To avoid this, we propose the addition of interactivity to the system. Controllers can interactively choose to display or hide information according to their preference or situation. Thus, we analyse some common input actions that controllers may perform and map them to various interactive gestures. The analysis of these input actions is described in Section IV-A while the interaction design is detailed in Section IV-B.

A. Functionality Analysis

After a literature survey, several possible input actions and features were identified to be included in the system. This led to a discussion with an air traffic controller in which we identified only the necessary input actions to represent a prototype of a full-fledged DT system to be developed. The following key actions were identified to be included as functionalities in the mixed reality DT system:

1) View basic aircraft billboard;
2) View detailed aircraft billboard;
3) View route of an aircraft (with basic billboard);
4) View distances between 2 aircraft (with basic billboard);
5) View departing flights on/approaching runway entry point; and
6) View arriving flights approaching gate.

These functionalities were then implemented and mapped to interactive gestures defined by an interaction model developed for this system, as detailed in Table I. This is explained further in Section IV-B.

B. Multimodal Interaction

Embodied interaction (the use of various parts of one's body as input elements) has been shown to be more intuitive than indirect interaction methods such as point-and-click [15] [16]; directly manipulating an onscreen object also improves immersion. We want our DT system to have enhanced immersion to allow controllers to focus on the OOT view without distractions from auxiliary menu-based interfaces. Moreover, combining different methods of embodied interaction (multimodal interaction) takes human-machine interaction one step closer to simulating real-world interaction. For an air traffic control system, the interaction design should minimize the effort required for controllers to perform primary and critical actions on the system. Directly manipulating data and objects (such as pointing directly to an aircraft to highlight it, or gazing at a specific focal area to view additional information) is preferable to conventional menu-driven interactions that occlude the OOT view and decrease situation awareness.

Furthermore, presenting a novel DT system such as this would pose the challenge of transitioning air traffic controllers to a new and unfamiliar system. This leads one to the question of whether the answer to a smooth transition lies in intuitive and thoughtful interaction design incorporated in the system.

C. The Where-What Interaction Model

The functionalities implemented require a carefully designed interaction model that leverages the gesture recognition capabilities of Microsoft HoloLens 2. The primary input modality offered by HoloLens 2 is hand gesture-based. Directly manipulating objects, or holograms, with the user's hand is well-supported in the system, and this forms the basis of our interaction model. In addition to supporting hand gesture recognition, HoloLens 2 also provides smooth eye gaze tracking capabilities. This led to the final design of our interaction model for air traffic control in mixed reality - the Where-What interaction model, shown in Fig. 5.

This model conceptualizes a Spatial Context-Action metaphor wherein the spatial context is defined by the user's eye gaze location, or visual focal area ('where' the target is), and the action to be performed is defined by an explicit hand gesture ('what' action the target should perform). This also derives from the gaze-and-commit model of interaction provided by MRTK, where a target is gazed at to focus on it, and its selection is determined by an explicit hand gesture. In our interaction model, however, target selection is done by the eye gaze, and the hand gesture may specify explicit actions for the selected target to do, such as selective information retrieval. This is designed with the aim of reducing visual clutter while introducing new information for the controller to visualize, since eye gaze-based target selection would focus the controller's visual focus area to that of the target aircraft.

Such a model incorporates a multimodal interaction format, wherein two different input modalities specify the spatial context and action simultaneously. In our system in particular, this model is adapted to provide simultaneous locating of the target by using eye gaze, and the action to be performed by using explicit hand gestures. Therefore it incorporates two

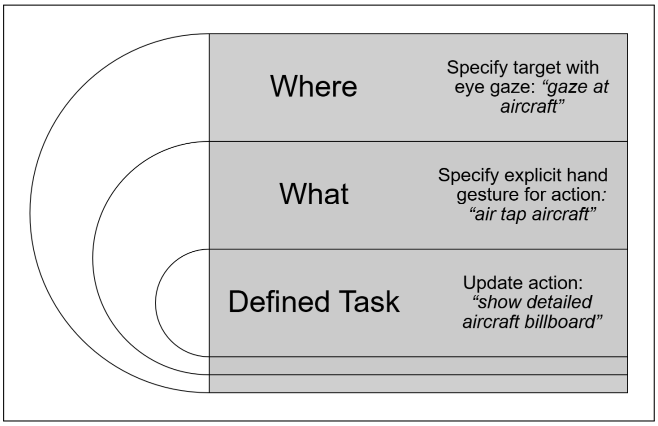

Figure 5: Incorporating the Where-What interaction model in the system's interactive targets.

embodied (human body-based) interactions, namely, eye gaze and hand gestures.

As a result, the detailed interactive functionalities of the system after incorporating the multimodal What-Where interaction model in aircraft targets, and hand gesture interaction in gates and runways, is presented in Table I. The interactive targets are aircraft, gates and runways, with various levels of information embedded in them. These can then be progressively activated using various multimodal gestures. The visualizations of the interactive targets mapped to their respective activation gestures are shown in Fig. 6.

V. Preliminary Assessment

A preliminary assessment of acceptance of the MR-based system was conducted with three licensed air traffic controllers with aerodrome control ratings. The aim of the assessment was to get initial feedback and insight into controllers' perceived usefulness of such a system as a viable replacement or augmentation of a remote DT system. It also aimed to assess the effectiveness of the interactions in the system. Since this was a preliminary assessment of the MR-based DT system using three air traffic controllers, subjective methods were chosen: a think-aloud protocol while the participant interacted with the system, and a post-experience interview session with each participant where they answered questions on the usability of the system, and verbalized their general feedback and insights on the system.

A. Participants and Procedure

There were three participants recruited for this preliminary assessment of the prototype. All participants are either former or present licensed air traffic controllers with aerodrome control ratings.

C1, the first air traffic controller, is aged 60 and has 25 years of experience as an air traffic controller, and is retired. C2, the second air traffic controller, is aged 35 and has 8.5 years of experience as an air traffic controller. C3, the third air traffic controller, is aged 28 and has 4 years of experience as an air traffic controller. All controllers were unfamiliar with the workings of Microsoft HoloLens 2 and navigating

TABLE I. Interaction Functionalities

Target	Selection gesture	Information	Explicit selection gesture	Action performed
Aircraft	Eye gaze	Basic billboard	None	View basic aircraft billboard
		Detailed billboard	Air tap+hold	View detailed aircraft billboard
		Route	Double tap	Highlight route of an aircraft (with basic billboard).
Multiple aircraft	Eye gaze	Relative distances	Air tap both	Highlight distances between 2 aircraft (with basic billboard)
Runway entry	None	Departing flight	Air tap target runway	Highlight departing flights on/approaching selected runway
Gate	None	Arriving flight	Air tap target gate	Highlight arriving flights approaching selected gate

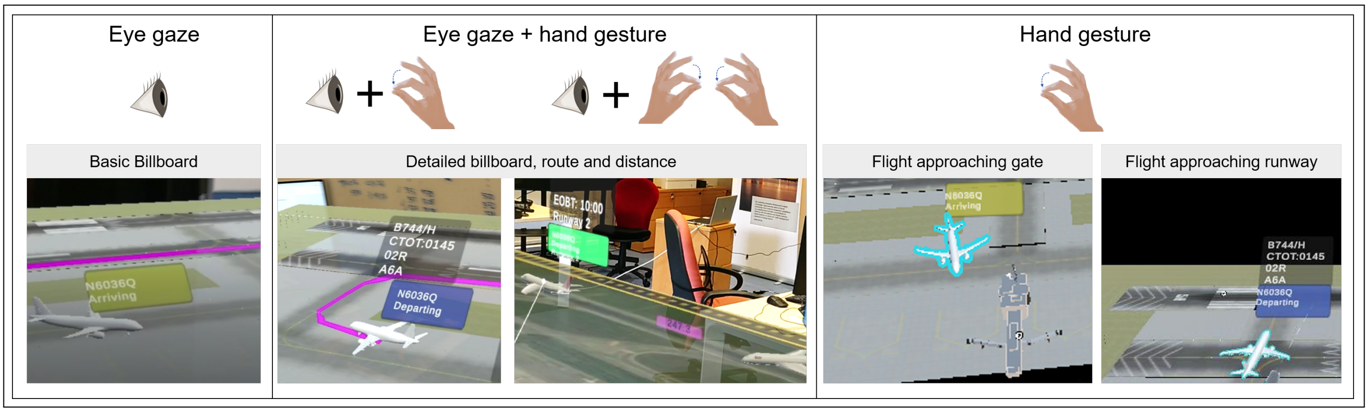

Figure 6: Multimodal gestures and the activated interactive affordances of the system.

in mixed reality in general, so they were initially briefed on the concepts of mixed reality and Microsoft HoloLens 2.

After they strapped on the device and underwent an eye gaze calibration, they were given some free time to explore the system and try out its interactions. Thereafter, they received a briefing on the DT system's features in mixed reality, the interaction model in use, as well as an overview of the functionalities of the system. They were informed of the entities that could be interacted with - the aircraft, gates and runway entry points - and were also informed of how to interact with each target to get further information if needed. Post the briefing, the holographic application was started for each participant. The participants were asked to explore the 3D simulation of the airport and its visual features initially, and were asked to interact with available targets to visualize information. Simultaneously, they were asked to follow a think-aloud protocol to verbalize their experience, thoughts and general impressions at every step while interacting with the system. They were given minimal guidance by the investigators, who also recorded and noted the participants' thoughts.

After the participant used the system for ten minutes, they were allowed to remove the headset to take a break if needed, and the investigators started the post-experience interviews and discussion. The interviews followed a semi-structured format, and consisted broadly of three topics, detailed below.

1) System trust, confidence and understanding: Conducted via a structured interview in which participants gave verbal ratings. The questions in this interview were formulated to summarize the controllers' overall perception of the interactive mixed reality system for air traffic control. They were asked to rate the system in terms of their understanding, confidence and trust in the system on a three-point Likert Scale (1-Low, 2-Medium, and 3-High). For the participant, understanding is defined to be their perceived understanding of the system and how it works. Trust is defined to be their belief in the system to perform their tasks reliably and consistently. Confidence is defined as the ability of a system to offer functionalities and interactions that create a perception of reliability and transparency [17].

2) System Usability: Conducted via a structured interview based on the System Usability Scale (SUS), where participants give verbal ratings. The questions posed in this interview was a derivation of the SUS questionnaire where participants rate each usability category on a Likert scale from 1-5. In our version, participants were asked these questions as a part of a structured interview process. Moreover, with the low number of air traffic controllers available as participants, subjective responses would provide more conclusive findings than quantitative data.

3) Air traffic control applicability: Conducted via a structured interview where participants answered open-ended questions regarding their perception of the system's usefulness in the air traffic control domain. This interview dealt with issues specific to the air traffic control domain. The following questions were asked:

1) Did you think the system will enable you to visualize information without compromising situation awareness?
2) How likely are you to use this system for low density air traffic control?

Figure 7: User perception of system trust, confidence and understanding with their mean ratings.

3) How likely are you to use this system for medium to high density air traffic control?
4) What are some situations in which this system will be particularly useful?
5) What are some situations in which this system will be less useful?

These questions were asked to get controllers' perception on the usefulness of the assessed system in the ATC domain.

4) General feedback session: Here, participants would elaborate on their earlier responses and provide general insights and comments on the system, their perceived reception of it among controllers, possible future extensions and suggestions to improve it.

All responses by the participants were recorded as audio clips and answers, and general insights were also noted down.

B. Ethics Statement

All subjects provided informed consent to take part in this study, which was approved by the Nanyang Technological University Institutional Review Board (IRB-2021-01-028).

C. Results

1) System trust, confidence and understanding: The participants' responses to the questions of system trust, confidence and understanding are summarised in the graph in Fig. 7.

2) System usability: The participants' responses to the SUS-derived questions in the interview (Yes, No, Maybe) are summarized in the stacked graph in Fig. 8, which depicts the division of responses (colored differently) for each question.

3) Air traffic control applicability: All controllers reported that they perceived the system to help them visualize information better without compromising situation awareness. Two controllers reported to be very likely to use the mixed reality system in low density aerodrome environments, while one controller believed that low density environments would have less chance of visual clutter and more time to gather information, hence such a system with overlaid information would be less useful. Two controllers revealed a hesitance to use the system in medium to high density environments owing to the unfamiliarity of the system, while the third controller believed that such situations would benefit the most from selective access to overlaid information on screen and reduced visual clutter. Regarding the question of situations where the system would prove particularly useful, the controllers had varied responses. C1: "Times when you would benefit from not looking down"; C2: "Nighttime operations would greatly benefit from clear visuals such as the billboards"; C3: "In times of bad weather, we would benefit from the clear visuals, overlaid information and flight billboards". However, the controllers also considered situations where the system would be less useful. For example, two controllers mentioned emergency situations as exceptional scenarios where any additional information (including a basic billboard) would pose a risk to their heightened situational awareness. In such cases, C2 suggested introducing an 'emergency mode' of operation in the system with minimal visual overlay and distractions. One controller also believed that good weather and low density environments would render the extra information the system provides to be potentially unnecessary due to operations being less time-critical.

4) General feedback: After the two structured interviews, the participants were free to give general feedback and impressions of their experience using the system. We found that their responses to this had several thematic similarities with their think-aloud responses. We have grouped their subjective comments for both into four main categories: visuals, interactions, system usability and learnability. Some key repeated comments for each category have been detailed in Fig. 9.

VI. Discussion

In general, all three controllers had similar responses in terms of system complexity, functionality integration, and confidence.

Although all controllers were completely unfamiliar with using a mixed reality headset, all reported feeling highly confident while using the system. All controllers also felt that their understanding of the system was slightly affected by the novelty and unfamiliarity of the mixed reality platform and device in use, and thus reported a moderate understanding of the system. The question of trust evoked different responses, however: the retired controller with 25 years of experience, having been familiar with system outages and technical unreliability in their career, stressed on the need for a backup system in case the HoloLens 2 suffered a hardware failure. The younger controllers found it easier to trust the system to be reliable and functional even with safety-critical operations.

The next part of the assessment addressed the system's usability through a modified version of the standard SUS evaluation. While we cannot statistically conclude that the system was found highly usable, the subjective responses to the various components of system usability seem to indicate that all participants found the system reasonably usable. The various questions in SUS can be seen as addressing primarily the usability and learnability of the system. While usability refers to ease of use of the system, learnability refers to the amount of training necessary to surmount the unfamiliarity of a novel system or interaction technique. From Fig. 8, it can be

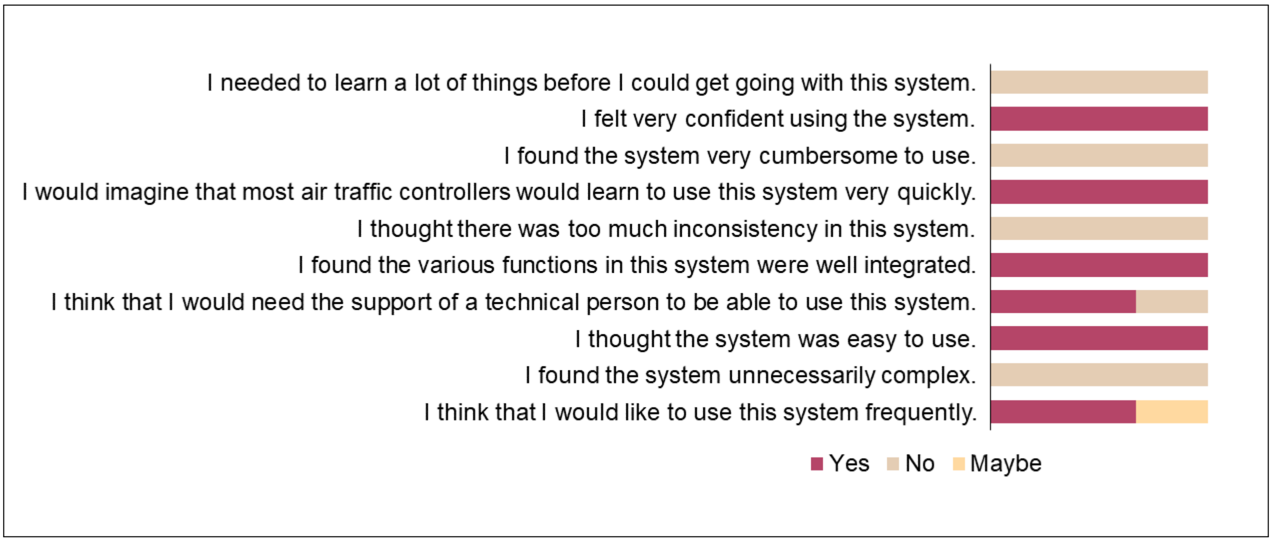

Figure 8: User perception of system usability based on standard SUS questions, showing the division of responses for each question.

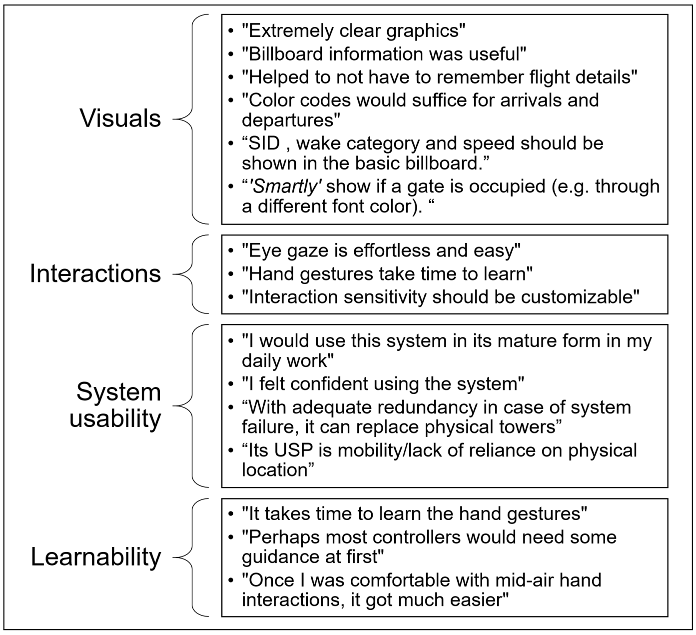

Figure 9: Thematic analysis of the think-aloud and subjective feedback from controllers

seen that all participants have rated the usability of the system favourably. Of particular interest are the varying answers obtained from the participants on the subject of learnability ("I would need the support of a technical person to be able to use this system"). One controller, C2, acknowledged the role that familiarity with using new technology would play in using this system, and concluded that not all controllers would find it easy to navigate the novel mixed reality platform. In particular, C2 reported that the hand interactions require limited practice before they are perceived as natural.

Such a response is consistent with the participants' think-aloud responses as well as their feedback during the general interview, as evinced in Fig. 9 regarding learnability, particularly with respect to hand gesture interactions. This would suggest that such a novel system would benefit from a training session to be given to its users prior to its integration into the ATC domain.

A. *Developed Guidelines*

Taking into account the various comments on learnability and usability of the interactions from the controllers, we form a few general guidelines with which the interaction design of this system will align.

- Since eye gaze interactions were preferable to hand interactions, the most commonly used interactive features will be mapped to eye gaze gestures.
- Hand interactions will be reserved as explicit gestures for seldom-used operations, such as gate and runway interactions.
- Controller preferences for level of visual information vary. Hence controller-specific interaction and visual customization capabilities will be added, so that the system can adapt to each controller's preference. Additionally, the system can also intelligently learn and adapt to each controller's interaction patterns in the future.

The assessment conducted, although limited in nature, is effective in guiding the design of the large-scale system that is under development. While we acknowledge its limited nature, this is seen as an early controller acceptance assessment of a prototype of a highly interactive digital tower system in mixed reality.

The assessment has gleaned several useful insights into controllers' perceptions of our system, and they seem promising enough to imply that this novel digital tower system could possibly be a viable alternative to the present immersive digital tower environments. The system can be especially useful in situations such as nighttime operations and low visibility conditions, as understood from their comments. The controllers' favorable feedback of the multimodal interaction model would further imply that adding natural and embodied multimodal interactions can enhance an existing digital tower system.

VII. Conclusion and Future Work

This paper presented a novel mixed reality-based digital tower that is self-contained within Microsoft HoloLens 2. Furthermore, the system incorporates multimodal interaction capabilities using a novel What-Where interaction model that utilizes eye gaze for spatial target locating and basic flight billboard activation, and hand gestures for explicit information activation. This model is used to provide interactive capabilities to aircraft, runways and gates for controllers to obtain a conformal and highly interactive representation of all available information on a single device. This serves the purpose of reducing head-down time for controllers and enhancing their visualization without contributing to visual clutter, which can lead to improved situational awareness.

The preliminary assessment conducted with three air traffic controllers serves to guide future design and enhancement of this system, since this is an initial prototype of a larger, live OOT-view system in HoloLens 2 that is currently under development. Nevertheless, controllers reported confidence in using the system to achieve their goals and also perceived the system to be usable in an ATC domain. While the three controllers perceived the multimodal interaction model to be initially unfamiliar, they reported the interactions to be comfortable and intuitive after acquiring familiarity. This suggests a learning curve in interacting with a system, especially for users unfamiliar with mid-air interactions on wearable devices. Appropriate training sessions can mitigate this issue.

This prototype of the system will be developed into a fully-functional DT system in the future with the following visual functionalities:

- A live OOT view from cameras depicted as video feeds in mixed reality;
- Aircrafts, runways and gates identified in the live video through vision-based techniques as interactive targets; and
- Individually customizable controller preferences for flexible visual layouts and information abstraction levels.

These functionalities will be integrated with the existing interaction methods and followed with a system evaluation including more controllers.

In conclusion, the preliminary assessment of this prototype system suggests that after limited training, a self-contained mixed reality-based digital tower system that has embodied, multimodal interaction capabilities offers numerous advantages over present-day DT systems. These include reduced operational costs, customizable views, enhanced information visualization and interactivity, and the results indicate that controller confidence will not be affected in operating a novel system on an unfamiliar platform.

Acknowledgement

This work was conducted under the Saab-NTU Joint Lab with support from Saab AB (publ) and the Air Traffic Management Research Institute (ATMRI).

References

[1] G. Rottermanner, V. A. de Jesus Oliveira, P. Lechner, P. Graf, M. Kreiger, M. Wagner, M. Iber, C.-H. Rokitansky, K. Eschbacher, V. Grantz, V. Settgast, and P. Judmaier, "Design and evaluation of a tool to support air traffic control with 2d and 3d visualizations," pp. 885–892, Institute of Electrical and Electronics Engineers (IEEE), 6 2020.

[2] Y. Lee, S. Marks, and A. M. Connor, "An evaluation of the effectiveness of virtual reality in air traffic control," pp. 7–17, Association for Computing Machinery (ACM), 2 2020.

[3] M. Bourgois, M. Cooper, V. Duong, J. Hjalmarsson, M. Lange, and A. Ynnerman, "Interactive and immersive 3d visualization for atc," *Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2005*, pp. 197–203, 2005.

[4] L. Pfeiffert, N. H. Müller, and P. Rosenthal, "A survey of visual and interactive methods for air traffic control data."

[5] C. Jiang, "Recent advances for smart air traffic management: An overview," vol. 572 LNEE, pp. 415–421, Springer, 2020.

[6] G. Rottermanner, V. Settgast, P. Judmaier, K. Eschbacher, and C.-H. Rokitansky, "Vast: A high-fidelity prototype for future air traffic control scenarios," 2019.

[7] A. Ross and G. . Crida, "Retina project conclusions," 2018.

[8] Han, Shah, and Lee, "Holographic mixed reality system for air traffic control and management," *Applied Sciences*, vol. 9, p. 3370, 8 2019.

[9] B. Sikorski, P. Leoncini, and C. Luongo, "A glasses-based holographic tabletop for collaborative monitoring of aerial missions," vol. 12242 LNCS, pp. 343–360, 2020.

[10] S. H. H. Shah, K. Han, and J. W. Lee, "Interaction paradigms for air traffic control and management in mixed reality," vol. 11575 LNCS, pp. 547–556, Springer Verlag, 2019.

[11] Y.-H. Chang and C.-H. Yeh, "Human performance interfaces in air traffic control," *Applied Ergonomics*, vol. 41, no. 1, pp. 123–129, 2010.

[12] "Hololens 2 hardware — microsoft docs,"

[13] "Unity development for hololens - mixed reality — microsoft docs,"

[14] "Introducing mrtk for unity - mixed reality — microsoft docs,"

[15] N. H. Dardas and M. Alhaj, "Hand gesture interaction with a 3d virtual environment," *The Research Bulletin of Jordan ACM*, vol. 11, pp. 86–94, 2011.

[16] W. Lin, L. Du, C. Harris-Adamson, A. Barr, and D. Rempel, "Design of hand gestures for manipulating objects in virtual reality," 2017.

[17] F. Corno, E. Guercio, L. De Russis, and E. Gargiulo, "Designing for user confidence in intelligent environments," *Journal of Reliable Intelligent Environments*, vol. 1, pp. 11–21, 7 2015.