Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility

Maxime Reynal, Jean-Paul Imbert, Pietro Aricò, Jérôme Toupillier, Gianluca Borghini, Christophe Hurter

Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility

Article *in* International Journal of Human-Computer Studies · April 2019

DOI: 10.1016/j.ijhcs.2019.04.001

6 authors, including:

Maxime Reynal
Ecole Nationale de l'Aviation Civile

Jean-Paul Imbert
Ecole Nationale de l'Aviation Civile

Jérôme TOUPILLIER
Ecole Nationale de l'Aviation Civile

Gianluca Borghini
Sapienza University of Rome

Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility

Maxime Reynal [a,*], Jean-Paul Imbert [b], Pietro Aricò [d], Jérôme Toupillier [c], Gianluca Borghini [d], Christophe Hurter [a]

[a] Data, Economics and Information Visualization lab (DEVI), French Civil Aviation University (ENAC), Toulouse, France
[b] Aeronautical Computer Human Machine Interaction lab (ACHIL), French Civil Aviation University (ENAC), Toulouse, France
[c] Air Traffic Management department, French Civil Aviation University (ENAC), Toulouse, France
[d] Molecular Medicine department, Sapienza University, Rome, Italy

ABSTRACT

In an effort to simplify human resource management and reduce costs, control towers are now more and more designed to not be implanted directly on the airport but remotely. This concept, known as Remote Control Tower, offers a "digital" working context because the view on the runways is broadcast remotely via cameras, which are located on the physical airport. This offers researchers and engineers the possibility to develop novel interaction techniques. But this technology relies on the sense of sight, which is largely used to give the operator information and interaction, and which is now becoming overloaded. In this paper, we focus on the design and the testing of new interaction forms that rely on the human senses of hearing and touch. More precisely, our study aims at quantifying the contribution of a multimodal interaction technique based on spatial sound and vibrotactile feedback to improve aircraft location. Applied to Remote Tower environment, the final purpose is to enhance Air Traffic Controller's perception and increase safety. Three different interaction modalities have been compared by involving 22 Air Traffic Controllers in a simulated environment. The experimental task consisted in locating aircraft in different airspace positions by using the senses of hearing and touch through two visibility conditions. In the first modality (spatial sound only), the sound sources (e.g. aircraft) had the same amplification factor. In the second modality (called Audio Focus), the amplification factor of the sound sources located along the participant's head sagittal axis was increased, while the intensity of the sound sources located outside this axis was decreased. In the last modality, Audio Focus was coupled with vibrotactile feedback to indicate in addition the vertical positions of aircraft. Behavioral (i.e. accuracy and response times measurements) and subjective (i.e. questionnaires) results showed significantly higher performance in poor visibility when using Audio Focus interaction. In particular, interactive spatial sound gave the participants notably higher accuracy in degraded visibility compared to spatial sound only. This result was even better when coupled with vibrotactile feedback. Meanwhile, response times were significantly longer when using Audio Focus modality (coupled with vibrotactile feedback or not), while remaining acceptably short. This study can be seen as the initial step in the development of a novel interaction technique that uses sound as a means of location when the sense of sight alone is not enough.

KEYWORDS

Interactive spatial sound; Vibrotactile feedback; Multisensory experience; Air Traffic Control; Remote Towers

1. INTRODUCTION

For some airports with very low commercial traffic density (e.g. approximately two flights per day), it is difficult to provide a constant Air Traffic Control (ATC) service. From an economic point of view, this kind of airport is rarely profitable. One solution could be to centralize ATC into centers in which traffic would be remotely controlled and human resources would be brought together. Air Navigation Authorities and laboratories have already contributed information and recommendations for further developments in this direction (Calvo, 2014; Braathen, 2011; Fürstenau et al., 2009), and different solutions are already in development across Europe and worldwide (Nene, 2016). The present study falls in this research field in the aim of enhancing Air Traffic Controllers' (ATCo) performance and consequently, safety.

These ATC centers would be composed of several rooms called Remote Control Towers (or simply Remote Towers). The main difference between a physical control tower and a remote one is that in a Remote Tower ATCos would only have a dematerialized view of the airport which is controlled, radar and radio facilities. In a physical control tower, other natural stimuli are often provided to them explicitly or implicitly. For example, the sound of an engine starting on the parking area carries information about aircraft location and direction, and that the pilot will soon contact them to ask for taxi instructions. If the tower is tall, its oscillations can inform of gusts of wind. Therefore, some implied information which could be crucial for ATCos at a specific moment could be lost in a remote control environment. However, and precisely because this specific ATC context is dematerialized, the Remote Tower concept offers new possibilities of interaction and feedback. Regarding what is happening at the distant airport, the previous information could be synthetically retrieved or further augmented using spatial sound or vibrotactile feedback, as long as an acceptable level of realism is retained.

One of the recurring problems in approach and local ATC is the visual location of aircraft. In particular conditions, especially when visibility is poor (e.g. heavy fog, or simply of loss of video signal in the remote control context) aircraft detection could be difficult or even impossible using only the visual sense. In these circumstances, ATCos today no longer have access to other means of mentally representing aircraft locations using out-of-date tools such as the goniometer. In addition, some low-traffic areas are not equipped with radars. We know that human perception of sound sources in space is reasonably accurate: the smallest minimum audible angle (MAA) of the human ear is about 1-2 degrees (Mills, 1958). However, one significant type of location error in space occurs when sound sources are almost aligned on the same azimuth (a phenomenon known as "cone of confusion" described by Hermann and Hunt (2011)). The interaction modality presented in this paper has been designed taking these factors into account in order to assist the location of aircraft under poor visibility conditions. We use sound interaction based on hearing and touch channels as information vectors to augment this selection process. In degraded visibility conditions, the contribution could be twofold: to enhance the users' immersion with spatial sound, while increasing their performance with the use of a new interaction technique acting on this spatialized sound. Moreover, this could decrease visual channel bandwidth, which is often overloaded, especially for those professions generating a high mental load such as ATC (Sklar and Starter, 1999; Mélan and Galy, 2011).

In this paper we investigate the following research question: *How can user situational awareness in the aircraft location process be improved?* As a sub-level question, we have applied our research to the specific domain of ATC where we investigated how multiple stimuli may improve perceived information. In particular, an interaction based on head position acting on spatial sounds is presented and studied, as well as its coupling to vibrotactile feedback. Results from subjective and behavioral data helped us to quantify the positive contribution of such interaction modalities. Neurophysiological assessments were also studied in Aricò et al. (2018) from the same experiment.

The paper is structured as follows. Section 2 presents a brief overview of Human-Computer Interaction (HCI) research related to our study, namely HCI in a control tower context, interactive systems based on audio and precisely on spatial sound, haptics, and multisensory HCI considerations. Section 3 presents our interaction modalities and each sensory channel they rely on. Section 4 explains our experimental protocol. The results are related in section 5. They are discussed in section 6, in which we end with a conclusion and give the research perspectives.

2. RELATED WORKS

In this section we introduce some works related to HCI in Air Traffic Management (ATM), interactive systems on sound, and the introduction of haptics and vibrotactile feedback to provide information to the user. Examples in relation to our study are also given.

2.1. Existing augmentations in Control Tower environment

Existing studies and publications on HCI for remote control context bring for the most part increases related to the sense of vision (Cordeil et al., 2016; Hurter et al., 2012; Van Schaik et al., 2010). Sound-based interactive solutions already exist in the context of Remote Towers but they are quite few. To give an example, an innovative method of sound spatialization using binaural stereo in order to discriminate the communications of enroute ATCos is reported in (Guldenschuh and Sontacchi, 2009). Elsewhere, multiple Remote Tower (RT) context (i.e. when multiple airports are controlled from a single remote control room) has been little studied in scientific literature due to the fact that it is still a rather recent field. A first study was published in 2010 to demonstrate the feasibility of multiple RT, at least to control two small airports simultaneously (Papenfuss and Friedrich, 2016; Moehlenbrink and Papenfuss, 2011). Nevertheless, and to the best of our knowledge, no study has reported the introduction of spatial sound to give information to ATCos in an ATM environment.

2.2. Existing interactive systems acting on sound

An interactive system acting on sound that has similarities with the one presented in this paper was published by Bolt (1981). The goal was to focus user’s attention on sounds (about 20 simultaneous images and sounds) while facing a wall of screens operating simultaneously and broadcasting different images. The sounds emanating from these televisions were amplified according to the user’s gaze ("eyes-as-output"). The desire to amplify the sound *towards* the user has been identified several times in recent literature. In particular, we can mention the OverHear system (Smith et al., 2005) that provides a method for remote sound source amplification based on the user gaze direction using directional microphones. We can also address the AuraMirror tool (Skaburskis et al., 2003) which makes it possible to inform the user graphically of his attention by superimposing on his vision a particular shape (e.g. colored "bubbles") around the concerned interlocutors in a multi-speaker situation. However, these solutions are based on eye gaze information and not on user head orientation.

Several works related to 3D-sound interaction have been performed. A concept presented in (Savidis et al., 1996) resembles in some aspects the research of the present paper. In this study, the user is surrounded by interactive sound sources organized in a "ring" topology. They can select specific sound sources with 3D-pointing, gestures and speech recognition inputs. The goal is to provide a way to explore the auditory scene – provided using Head-Related Transfer Function (Brungart and Simpsons, 2001; Cheng and Wakefield, 2001; Wenzel et al., 1993) – with the use of direct manipulation (Hutchins et al., 1985) via a "ring" metaphor mapping of a structured environment.

2.3. Exiting haptics and vibrotactile HCIs

Touch-based, or more generally haptic-based HCIs have been studied for many years. One can quote the PHANTOM device (Massie and Salisbury, 1994), which is a haptic device in the form of a mechanical arm with fine-tuned force feedback. The user who manipulates this device can feel forces materializing the collision with the handled virtual objects, increasing precision in their interaction experience. Another similar project called GROPE was also developed (Brooks et al., 1990). More recently, Ultrahaptics devices (Carter et al., 2013) or more generally mid-air ultrasonic haptic feedback (Wilson et al., 2014) have taken the form of a matrix of ultrasonic transmitters. Using intersections between the emitted ultrasonic waves, the objective is to make the user feel mid-air 3-dimensional shapes. Currently, spatial resolution (i.e. the number of cross-points materializing the shape) is still rather weak. Another device called TeslaTouch (Bau et al., 2010), which relies on the electrovibration principle combined with an interactive display and touch input, enables the user to feel virtual elements through touch. FeelTact can also be cited (Esposito and Lenay, 2011), which is a bracelet used to transmit information to its carrier using vibrations. A large number of vibratory patterns can be created, which provides a rich means of dialogue that can be useful for disabled people.

Numerous studies and systems have been presented in recent years to communicate spatial information to users via vibrotactile feedback. In the context of driving, several studies use vibrotactile patterns to indicate directions or obstacles to be avoided to the driver (Petermeijer et al., 2017; Schwalk et al., 2015; Meng et al., 2015; Gray et al., 2014; Jensen et al., 2011; Ho et al., 2005). More generally, others use vibrotactile feedback to manage the allocation of user attention (Sklar and Sarter, 1999). In the musical field, some use vibrotactile feedback to study the perception of dissonance (Fontana et al., 2016). There have also been many studies in the aeronautical field. For example, a system has been developed with vibrotactile feedback to indicate vibrations to pilots through a belt (Van Erp et al., 2005; Van Erp et al., 2004), or tactile cues have been given to helicopter pilots to help them perform hover maneuvers (Raj et al., 2000).

3. IMPROVING SPATIAL SOUND SOURCE LOCATION USING HEARING AND TOUCH

In this section we present the interaction and feedback techniques that have been investigated in this study, namely spatial sound (3D), Audio Focus interaction (AF), and AF interaction coupled with vibrotactile feedback (AF+V). As said in (Loftin, 2003), our work might belong to the category of "enactive interfaces", since they are part of "*those that helps users communicate a form of knowledge based on the active use of [...] the body*". In addition, "*enactive knowledge is stored in the form of motor responses and acquired by the act of doing*", which is what AF interaction modality asks the user to do. Here, the term "modality" is used as a characteristic of an interaction mode. More precisely, the 3 modalities we investigate respectively rely on spatial sound, head position acting on spatial sound, and the latter coupled with haptic feedback. The term "acting" is here used to designate information provided by the user to the system (act of doing).

3.1. Auditory channel: “Audio Focus” interaction

Following discussions and interviews with professional ATCos, it appeared that a key issue in the control tower is the lack of visual cues on aircraft when meteorological conditions induce poor or no visibility. When weather is good, i.e. when ATCos have a view on the aircraft, they naturally move their heads toward them to visually pinpoint their location. This natural behavior has no sense when visibility is poor. Audio Focus interaction modality is designed to give the ATCo a way to reproduce this behavior in these circumstances, but relying on the auditory channel rather than visual one.

The AF interaction principle is based on the high correlation between head orientation and visual attention (Stiefelhagen et al., 2001). It relies on spatial sound sources (Figure 1). These are engine sounds coming from small types of aircraft. They are linked to each static position of aircraft in the airport vicinity. Participant head orientation and position have been retrieved using a Microsoft HoloLens device. Sound sources (e.g. aircraft) are selected along the head sagittal axis (± 7 degrees): the gain of sound sources located along the head sagittal axis is increased, while the gain of those located away from this axis is decreased. As on average the minimum time required to locate a sound is approximately 100 milliseconds (Vliegen and Opstal, 2001), sound sources are played continuously. Finally, the distance between the concerned aircraft and the user point of view is also mapped into the gain of the sound sources (louder when the aircraft is closer to the user). Typically, the final goal of this interaction modality is to help locate the related aircraft in a 3-dimensional environment. The AF interaction modality can be qualified with "spatial filtering" (Andéol et al., 2017): the relative sound levels between sound sources on and away from the head sagittal axis is adjusted in order to let the participants "play" with the sounds they are hearing by changing their point of view.

3.2. Touch channel: vibrotactile feedback

Humans are good at locating sounds along the horizontal axis (Mills, 1958), and AF interaction relies on this aptitude. Haptics, and more precisely vibrotactile feedback, has been added to this modality with the goal of mapping the vertical axis by giving cues on the vertical position of the aircraft. Vibrotactile feedback is here used to support AF interaction, especially in poor visibility conditions. They are activated only for sound sources located on the head sagittal axis.

The aim is to increase the vertical selectivity by adding information notifying the participant if the aircraft they are aiming at (i.e. on their head sagittal axis) is in the air or on the ground. In the same way as (Petermeijer et al., 2017) and (Sklar and Sarter, 1999) the vibrations are here for the purpose of unloading the visual sense by presenting spatial information through the sense of touch. To present this feedback to participants, two vibrotactile transducers have been fixed on a wooden chair (see section 4.9), one under the chair seat, and the other behind its back. Providing the user's head is oriented toward an aircraft, the first vibrates if the aircraft is located on the ground (Down), and the second vibrates if the aircraft is located in the air (Up). AF+V interaction gives the system the same input as AF interaction (e.g. user head orientation), but giving the user another input (e.g. amplification of sound, as for AF interaction, but here coupled with vibrotactile feedback).

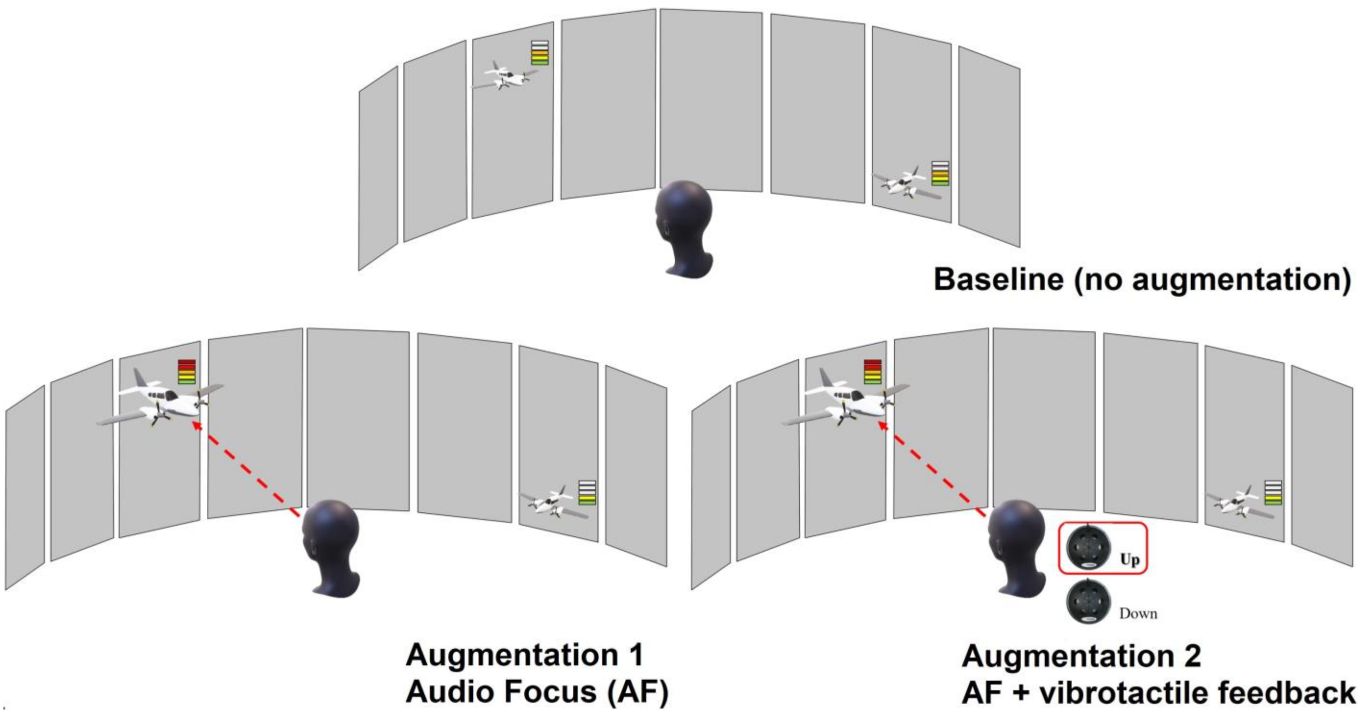

Figure 1. A representation of 3D (top), AF (bottom left), and AF+V (bottom right) modalities. Aircraft sizes, associated with gradually colored bars materialized the related sound volumes.

3.3. Visual channel

The experiment took place in a 3D environment generated with Flight Gear open flight simulator. In good visibility conditions, participants were able to see the aircraft through FG graphics: visual cues on aircraft positions are implicitly given. No aircraft can be seen in degraded visibility (i.e. in foggy conditions). The 3 interaction modalities have been tested in good and poor visibility conditions (Figure 2).

4. MATERIALS AND METHOD

4.1. Participants

Twenty-two French tower ATCos volunteers from different airports took part in the experiment (8 females, 14 males). The mean age was 40.68 years (*SD* = 8). Their professional experience was varied as their mean number of effective years in a control tower was 10.48 (*SD* = 6.87), but this aspect did not affect the experiment (see section 4.3). Eleven controllers formed group *A* and the remainder formed group *B* (see section 4.10). Since all of ATCos are subject to medical tests as part of their professional requirements, none of them had hearing issues. In particular, they reported no attenuation in their auditory bandwidth nor imbalance between the two ears.

4.2. Ethics

All participants were informed beforehand by a scientific officer about the objectives of the study, its methodology, duration, constraints, and foreseeable risks. They were entirely free to refuse to participate in the study and to terminate it at any moment without incurring any prejudice. They were informed of the anonymous nature of the data recorded (and consequently, of the impossibility of destroying the data in a targeted way, if they wanted to). All the participants signed a Consent Form to make it clear that they agreed with the conditions of the experiment. A local ethical committee approved this experiment before its execution.

4.3. Nature of the task

The experimental task concerned location. The participants were asked to give information about their perception of aircraft location in the airport vicinity thanks to their hearing and touch senses, in two visibility conditions and for each of the three interaction modalities. In good visibility conditions, aircraft were visible; in poor visibility condition, they were not. Since the simulation asked the participants to perform an ATM-like task, few constraints have been imposed in order to avoid any confusing effect.

4.4. Experimental conditions

Three different aspects were manipulated during the experiment: interaction type, difficulty level, and visibility conditions. A pretest phase helped to quantify these experimental conditions. The first one is the Modality factor, which is the current feedback and/or interaction type (Figure 1) and which can be *Spatial sound only* (baseline, called *3D*), *Audio Focus* modality (called *AF*), or *Audio Focus modality coupled with vibrotactile feedback* (called *AF+V*). Spatial sound only is here considered as a baseline since sounds are naturally spatialized when they are audible in a physical control tower. In a remote one, these sounds have to be generated to build a realistic environment, as close as possible to the real one. The number of simultaneous sound sources represents the Difficulty factor, which could be *Easy* (1 engine sound), *Medium* (2 simultaneous engine sounds coming from 2 different sound sources having separated positions in the airport vicinity), or *Hard* (3 simultaneous engine sounds coming from 3 different sound sources having separated positions in the airport vicinity). The last factor, named Visibility, is the meteorological or visibility condition (Figure 2), which is *Good* visibility (all aircraft are visible), or *Poor* visibility (fog, no aircraft is visible).

4.5. Hypotheses

We expected participants to locate aircraft more accurately using AF modality in poor visibility conditions, even more so when coupled with vibrotactile feedback. Therefore, three hypotheses have been made. When placed under poor visibility conditions, we expected that participants could locate aircraft more precisely (i.e. greater accuracy):

– H_1: AF *Vs*. 3D: when they are using AF modality compared to 3D modality;

– H_2: AF+V *Vs*. 3D: when they are using AF+V modality compared to 3D modality;

– H_3: AF+V *Vs*. AF: when they are using AF+V modality compared to AF modality.

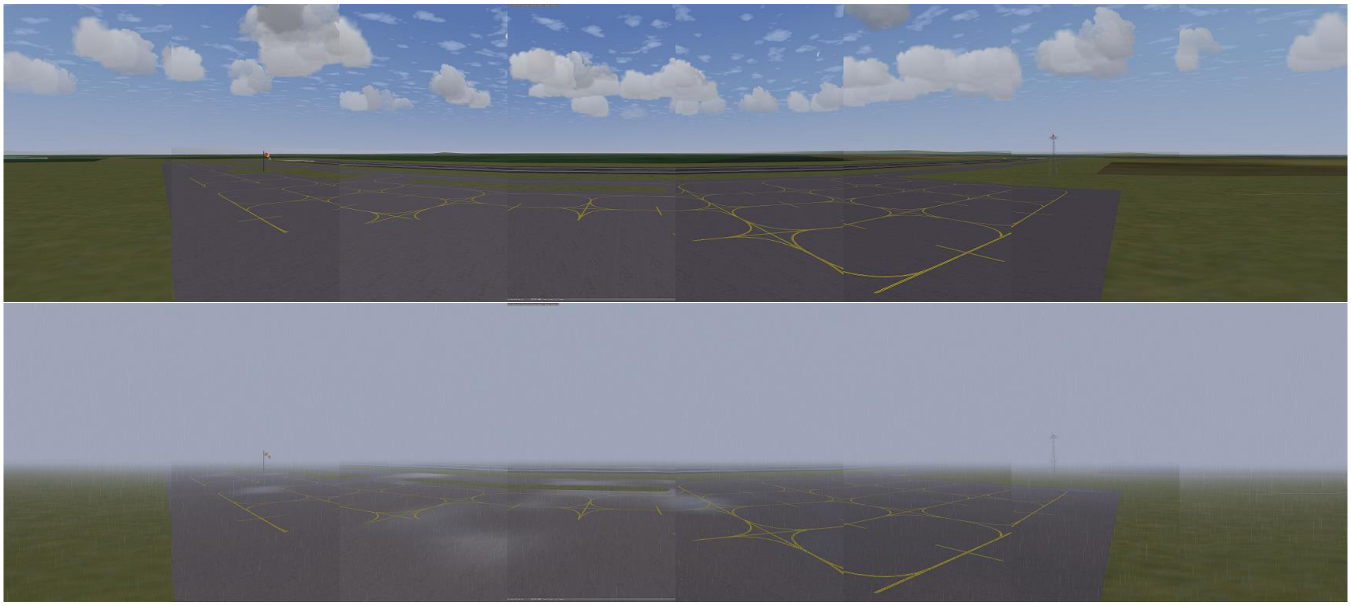

Figure 2. Screenshots of the Flight Gear displays used during the experiment. The two meteorological situations are good visibility (top), and poor visibility (bottom) with fog. No aircraft was visible in this last condition.

4.6. Airport discrimination using selectable areas

At an airport, aircraft follow a runway circuit. For this experiment, and considering statistical simplifications, Muret airport (located near Toulouse, South West of France) has been separated into five distinct areas. From the control tower point of view, which is located in front of the runway, we wanted to discriminate the right, the left, the space located in front of the control tower, and a more distant one. Regarding this approach and how spatial sounds could be manipulated, the choices we made for these five areas are the following (Figure 3): *Take-Off* and *Crosswind* legs, *Downwind* leg, *Base and Final* leg, *Runway, west* part, *Runway, east* part. Sound sources were placed inside each area. An area could not contain more than one aircraft at a time.

4.7. Answering HMI

Multiple static aircraft combinations have been displayed through Flight Gear according to these five areas. All the participants heard the same configurations, in a random order for each Modality, Difficulty level and Visibility conditions. By clicking on the corresponding area(s) through a specific Human-Machine Interface (HMI), participants were able to indicate the origin of the sound(s) they heard. They were seated in front of the separation between the west and east parts of the runway.

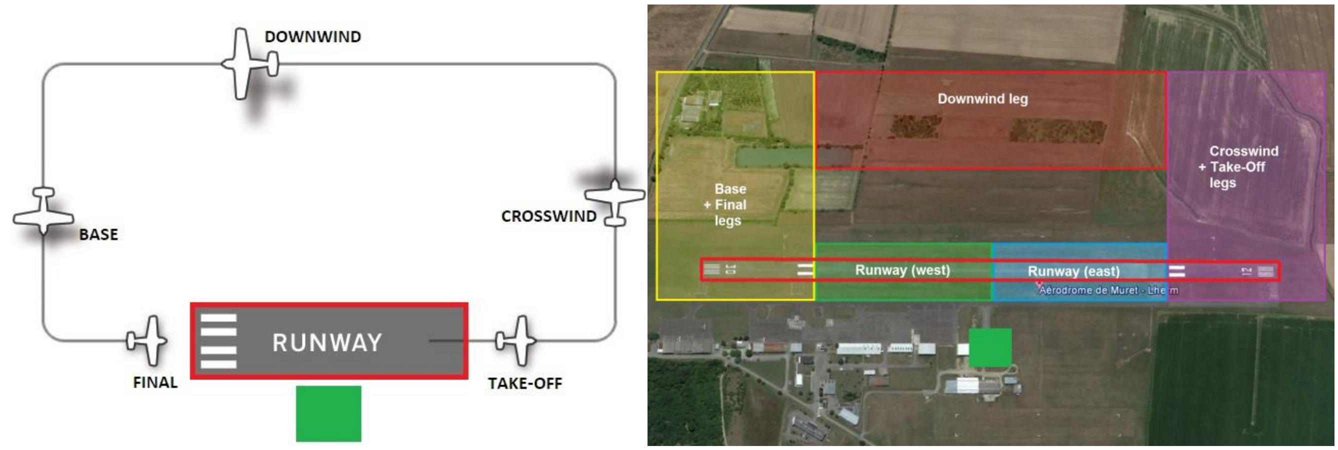

Figure 3. Runway circuit and airport description. On the left, a diagram of a typical runway circuit, which separates the four legs composing an approach segment: crosswind, downwind, final leg and then the runway. On the right, each area is represented on a satellite view of Muret airport. On the two images, the runway is highlighted in red lines, and the position of the control tower is represented with a plain green square.

An answering HMI (Figure 4) was used to register the participants' answers. The distance between the center of this HMI and each of its buttons was constant to minimize and standardize their movements while answering. In addition, every button was the same color in order to not influence their answers. Buttons had two states: selected (dark gray) or not selected (light gray). A validation button was positioned at the bottom of the window to validate their answers. It was displayed in green, and was positioned at the bottom right for right-handed participants, or at the bottom left for left-handed ones. This HMI was displayed on a tablet, which was used by the participants during the experiment.

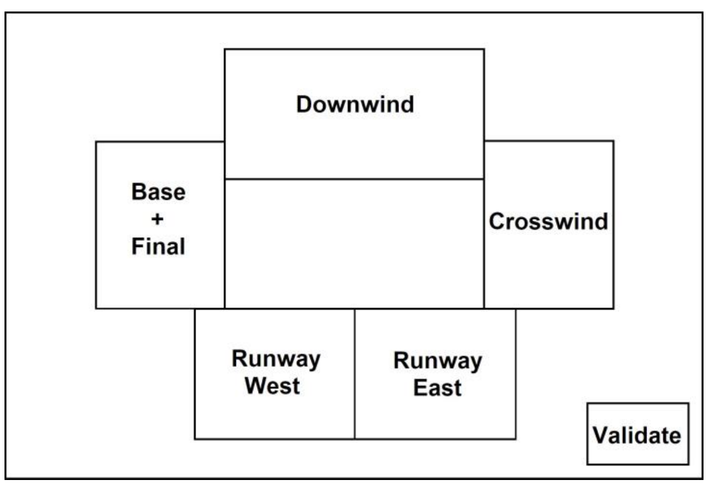

Figure 4. The general organization of the HMI which was used to collect participants' answers. The five buttons in the center corresponds to the 5 areas discriminating the airport environment.

4.8. Metrics

Participants' performance was measured using two different scores (dependent variables). Accuracy in the task quantifies the number of right answers, i.e. the number of correctly located aircraft among the five possibilities offered by the experimental design. The Response time is the time taken by the participants to locate aircraft, from the moment when a new combination was displayed to the moment when the Validate button on the answering HMI was pressed.

4.9. Experimental setup

The setup was composed of: 8 UHD Iiyama Prolite X4071 screens, an Alienware Area51 computer equipped with two NVidia GeForce GTX 1080 graphic cards, and a wooden chair on which two Clark Synthesis T239 Gold tactile transducers have been attached (Fontana et al., 2016) (one behind the back to code the Up and another one under the seat to code the Down). Since their sound quality is good enough to spread engine sounds and their use allows the installation to be non-individual (as opposed to the use of binaural sound through headset, for example), spatial sound was relayed using the screen speakers. This solution provides a physical spatial sound due to the physical positions of the eight speakers. As we said previously, the head orientation has been retrieved using a Microsoft HoloLens mixed reality headset. Its visual augmentation facilities were not used here and the participants were asked to use it with the glasses raised upon their head.

The 3D environment was created using Flight Gear open flight simulator (Figures 2 and 5). The different software modules for the augmentations (spatial sound, AF and vibrotactile feedback) were written in C# language using Microsoft .Net framework 4.6 and Direct Sound library. Network communications were developed using ENAC Ivy bus technologies (Chatty, 2003), which provides a high level means of communication using string messages and a regular expression binding mechanism.

Figure 5. A photograph of the sandbox while a participant is answering.

4.10. Combinatorial of the trials

There were three possibilities for the modality used (3D, AF or AF+V), two possibilities for the visibility conditions (Good and Poor), and three difficulty levels (Easy, Medium and Hard). The number of combinations for 5 airport area among 1, 2 or 3 simultaneous aircraft (difficulty level) is given by binomial coefficients[1]: $5C1$ = 5 (difficulty level 1), $5C2$ = 10 (difficulty level 2) and $5C3$ = 10 (difficulty level 3).

Some of these combinations are redundant. For example, for a given participant it is not mandatory to test combination 1.2*a* and combination 1.2*b* (Figure 6), because of their symmetrical positions: the first one is used to

[1] *nCk* is the binomial coefficient to compute "*k* among *n*".

test the perception of an aircraft located on the left, while the second is used to test the perception of an aircraft located on the right. This is why two sets of trials have been created, in which the same numbers of combinations per difficulty level were selected in a balanced way. Set *A* contained the combinations 1.1, 1.2*a*, 1.2*b*, 1.3*a*, 1.3*b*, 2.1, 2.3*a*, 2.4*a*, 2.5*a*, 2.6*a*, 3.1, 3.3*a*, 3.4*a*, 3.5*a* and 3.6*a*, and set *B* contained the combinations 1.1, 1.2*a*, 1.2*b*, 1.3a, 1.3*b*, 2.2, 2.3*b*, 2.4*b*, 2.5*b*, 2.6*b*, 3.2, 3.3*b*, 3.4*b*, 3.5*b* and 3.6*b*.

Finally, there were 15 combinations to be presented to each participant (from set *A* or from set *B*) which contained the combinations to be presented for each of the 3 levels of difficulty, within 2 visibility conditions and 3 modality conditions. This mean 90 trials to test all the conditions. From a statistical point of view, we decided to present this full set of trials four times to each participant, which means a total of 360 trials for each participant.

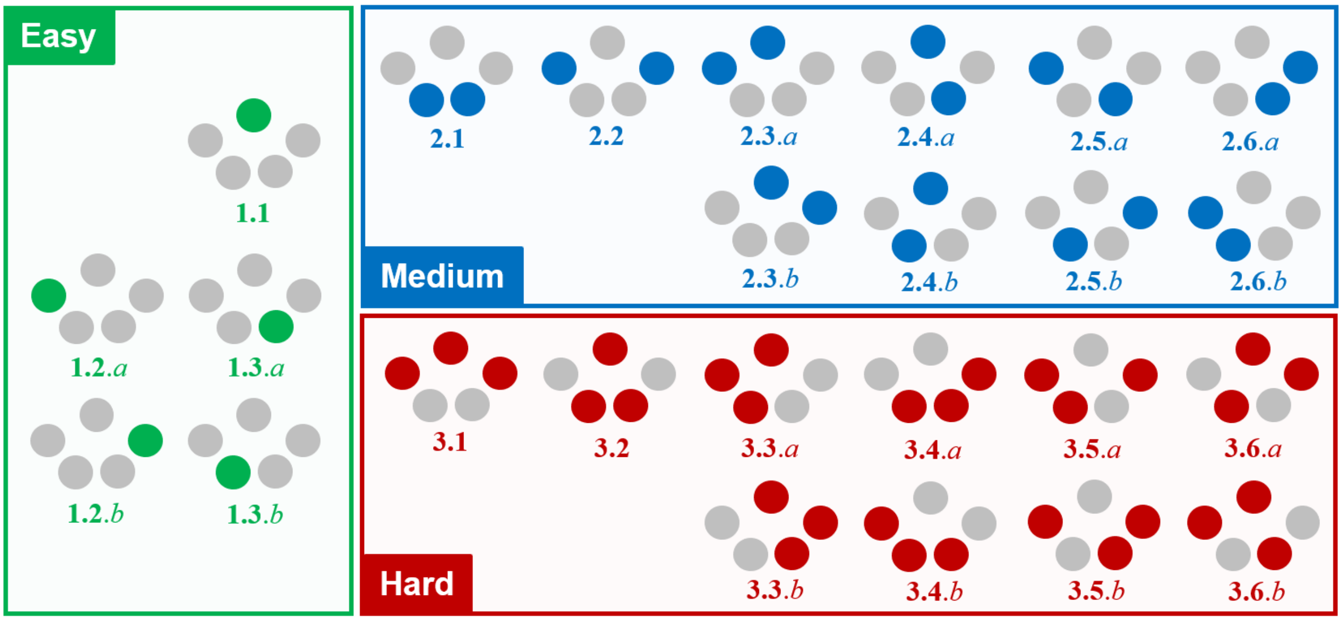

Figure 6. The combination used to place the sound sources during the experiment. The dots correspond to the 5 areas. Grey ones denote unused ones. Difficulty level 1 (Easy) is in green (1.*x*.*y*), difficulty level 2 (Medium) is in blue (2.*x*.*y*), and difficulty level 3 (Hard) is in red (3.*x*.*y*). For a given combination, *a* and *b* designate the possible symmetrical combinations.

4.11. Experiment organization

The experiment started with a training phase which consisted of a presentation of the stimuli (spatial engine sounds) in each different combination. To do so, two experimental blocks have been made: one under good visibility conditions, and another one under poor visibility conditions. Each of these two experimental blocks contained three other sub-blocks, one for each of the three modalities tested. The goal was to accustom the participants to "play" with the sounds by moving their head like if they had to look for something (but here, the search is done audibly and not visually). Each of these sub-blocks contained three trials (1 aircraft, then 2 and finally 3). At the end, all the potential aircraft combinations were presented. If needed, the training blocs were launched another time. Training was stopped once the participants acquired confidence with all the considered experimental cases.

The experiment started after this first training phase. It was divided into blocks for each modality. Each of these blocks contained two sub-blocks for good and poor visibility conditions. Within these two sub-blocks, the fifteen combinations were randomly presented to each participant, following the set *A* or *B* that had been randomly assigned to them. The distribution of the two visibility conditions sub-blocks was also randomized. Finally, the modality blocks were randomly ordered too. The three modality blocks were presented four times to each participant, with a five-minute break in the middle of the experiment.

At the end the participants completed an online questionnaire in order to give their feelings about the usability, fatigue, performance felt, and also perceived workload (see section 5.2).

4.12. Data analysis

4.12.1. Behavioral Data

The two behavioral variables which were acquired correspond to the two variables described in section 4.8: *Accuracy* which is defined by a percentage of correct aircraft localization by the participants for the five airport areas, and *Response time*, in milliseconds. The Accuracy variable was normalized using Arcsine transform (Wilson et al., 2013), and Response time using Log transform (Robert and Casella, 2004). For each variable, a two-way Analysis of Variance (ANOVA) with repeated measures (CI = .95) 3 × 2 (Modality [3D, AF, AF+V] × Visibility [Good, Poor]) was conducted and Tukey's HSD was used for post-hoc analysis.

Since there are not enough observations per participant to apply 3 level interaction analysis (i.e. 3 × 3 × 2 (Modality [3D, AF, AF+V] × Difficulty [1, 2, 3] × Visibility [Good, Poor]), which means 18 observations per participant among 22 participants), difficulty levels were averaged for each Modality and Visibility condition.

4.12.2. Questionnaires

After the experiment, participants were immediately asked to fill in an online questionnaire which was divided into five main parts. The first part contained General questions where participants were asked to submit their identification (ID number, gender, age, etc), to answer questions concerning failure to locate an aircraft, the preferred modality during the experiment, their performance scores from a general, subjective and qualitative point of view, and their opinion about the different modalities. The second part addressed the Usability aspects: participants were asked to give a score out of 7 (1 for easy, 7 for hard to use) for each of the 3 modalities. The third part addressed Fatigue aspects: as for Usability, participants were asked to give a score out of 7 (1 for extremely difficult to locate, 7 for easy to locate) for each Modality × Visibility × Difficulty level combinations. The fourth part was the NASA Task Load Index (NASA-TLX) questionnaires (Hart, 2006; Hart and Staveland, 1988), one for each modality. The fifth and last part was used for free remarks and suggestions for improvements, and consisted in a single question "*If you think of anything in relation to the experiment itself or more generally to the deported and/or augmented towers, or if you have any ideas for improvement in relation to the modalities that have been proposed to you, you are kindly invited to explain it here before completing this questionnaire*".

These results have been analyzed using descriptive explanations for the General questions (see section 5.2.1). One-way ANOVA with repeated measures (CI = .95) with *Modality* factor [3D, AF, AF+V] and *Usability score*, *NASA-TLX Mental demand score*, *NASA-TLX Physical demand score*, *NASA-TLX Temporal demand score*, *NASA-TLX Performance score*, *NASA-TLX Effort score* and *NASA-TLX Frustration score* dependent variables implemented as within factors. A two-way ANOVA with repeated measures (CI = .95) 3 × 2 (Modality [3D, AF, AF+V] × Visibility [Good, Poor]) with *Fatigue score* dependent variable implemented as within factor was conducted. All the values were normalized using Arcsine transform, and Tukey's HSD was used for post-hoc analysis.

5. RESULTS

5.1. Behavioral results

5.1.1. Accuracy

The analysis revealed main effects for Modality and Visibility factors and a Modality × Visibility interaction. Detailed results are reported in Table 1.

Variable	**ddl**	***F***	***p***	η^2_p
Modality	2, 42	116.93	< .0001	.85
Visibility	1, 21	409.93	< .0001	.95
Modality × Visibility	2, 42	215.52	< .0001	.91

Table 1. Results from 3 × 2 (Modality [3D, AF, AF+V] × Visibility [Good, Poor]) ANOVA with Accuracy dependent variable implemented as within factor.

Tukey's HSD post-hoc analysis showed the following main results (Figure 7). For the Modality factor main effect: participants were more accurate using AF+V modality (*M* = 1.36, *SD* ± .02) than using AF modality (*M* = 1.19, *SD* ± .01, $p < .001$) and 3D modality (*M* = 1.11, *SD* ± .02, $p < .001$). They were also more accurate using AF modality than using 3D modality ($p < .001$). For the Modality × Visibility interaction: in good visibility conditions, results were more accurate using AF modality (*M* = 1.47, *SD* = .02) than using AF+V modality (*M* = 1.43, *SD* = .02, $p < .001$) and 3D modality (*M* = 1.43, *SD* = .02, $p < .001$). In poor visibility conditions, participants gave more accurate answers using AF+V modality (*M* = 1.28, *SD* ± .03) than using AF modality (*M* = .9, *SD* ± .02, $p < .001$) and 3D modality (*M* = .78, *SD* ± .02, $p < .001$). Participants were also more accurate using AF modality than using 3D modality ($p < .001$).

5.1.2. Response time

The analysis revealed main effects for Modality and Visibility factor and a Modality × Visibility interaction. Detailed results are reported in Table 2.

Variable	**ddl**	***F***	***p***	η^2_p
Modality	2, 42	51.86	< .0001	.71
Visibility	1, 21	16.55	< .001	.44
Modality × Visibility	2, 42	18.43	< .0001	.47

Table 2. Results from 3 × 2 (Modality [3D, AF, AF+V] × Visibility [Good, Poor]) ANOVA with Response time dependent variable implemented as within factor.

Tukey's HSD post-hoc analysis showed the following main results (see Figure 7). For the Modality factor main effect: participants were faster using AF+V modality (*M* = 3.88, *SD* ± .02) than using AF modality (*M* = 3.93, *SD* ± .02, $p < .01$), but slower than using 3D modality (*M* = 3.8, *SD* ± .01, $p < .001$). They were also slower using AF modality than using 3D modality ($p < .001$). For the Modality × Visibility interaction, in good visibility conditions, participants were slower using AF+V modality (*M* = 3.85, *SD* ± .02) than using 3D modality (*M* = 3.81, *SD* ± .02, $p < .05$). They were also slower using AF modality (*M* = 3.89, *SD* ± .02) than using 3D modality ($p < .001$). In poor visibility conditions, participants were faster using AF+V modality (*M* = 3.92, *SD* ± .02) than using AF modality (*M* = 3.97, *SD* ± .02, $p < .01$), but slower than using 3D modality (*M* = 3.79, *SD* ± .02, $p < .001$). They were also slower using AF modality than using 3D modality ($p < .001$).

5.2. Results from questionnaires

5.2.1. General questions

Results from the General questions section of the questionnaires are summarized as following. To the question "*In a real working context, have you ever failed to locate an airplane?*", 81% of the participant answered Yes. The ways they used to detect their errors were the use of ATC tools, the help of their instructor, observations from the pilot or simply by visual scanning. To the question "*Which modality did you prefer to use during this experiment? (3D, AF or AF+V)*", 95.7% of the participants answered AF+V, while 4.3% of them answered AF. To the question "*In general, during the experiment, how many airplanes have you been unable to locate: No airplanes, A small number of airplanes, Some airplanes, Many airplanes, or a very large number of airplanes?*", 78.3% of the participants answered Some airplane, 17.4% answered Many airplanes, and 4.3% answered A small number of airplanes. To the question "*Do you think that spatial sound can be a help or a hindrance to locate airplanes?*" From 1, meaning a hindrance, to 7, meaning a help, the mean answer was 5.2 (*SD* = 1.87). To the question "*Do you think that Audio Focus interaction can be a help or a hindrance to locate airplanes?*" From 1, meaning a hindrance, to 7, meaning a help, the mean answer was 4.68 (*SD* = 1.87). To the question "*Do you think that Audio Focus interaction coupled with vibrotactile feedback can be a help or a hindrance to locate airplanes?*" From 1, meaning a hindrance, to 7, meaning a help, the mean answer was 5.82 (*SD* = 1.71).

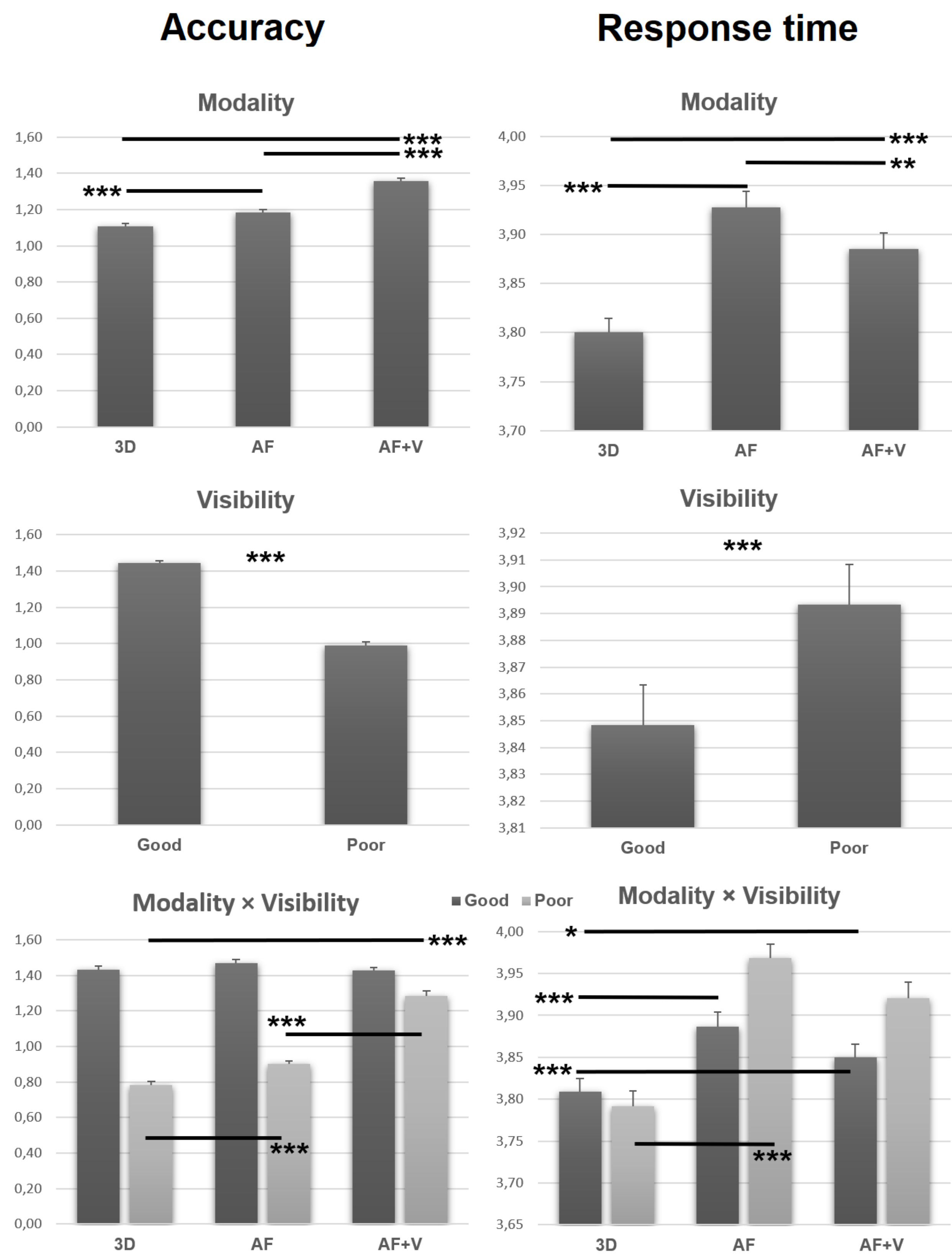

Figure 7. Results from inferential analysis on behavioral data. Left column is for Accuracy dependent variable (Arcsine normalization), and right one is for Response time (Log normalization). Error bars are standard errors.

5.2.2. Usability

A main effect was found for the Modality factor. Detailed Usability results are reported in Table 3.

Variable	ddl	*F*	*p*	η^2_p
Modality	2, 42	11.36	< .001	.35

Table 3. Results from one-way ANOVA with Usability score dependent variable implemented as within factor.

Tukey's HSD post-hoc analysis revealed that AF+V modality was considered more usable by the participants (M = 1.28, SD ± .04) than AF modality (M = .98, SD ± .04, $p < .001$) and 3D modality (M = .96, SD ± .08, $p < .001$). No significant difference was found between AF and 3D modalities (Figure 8).

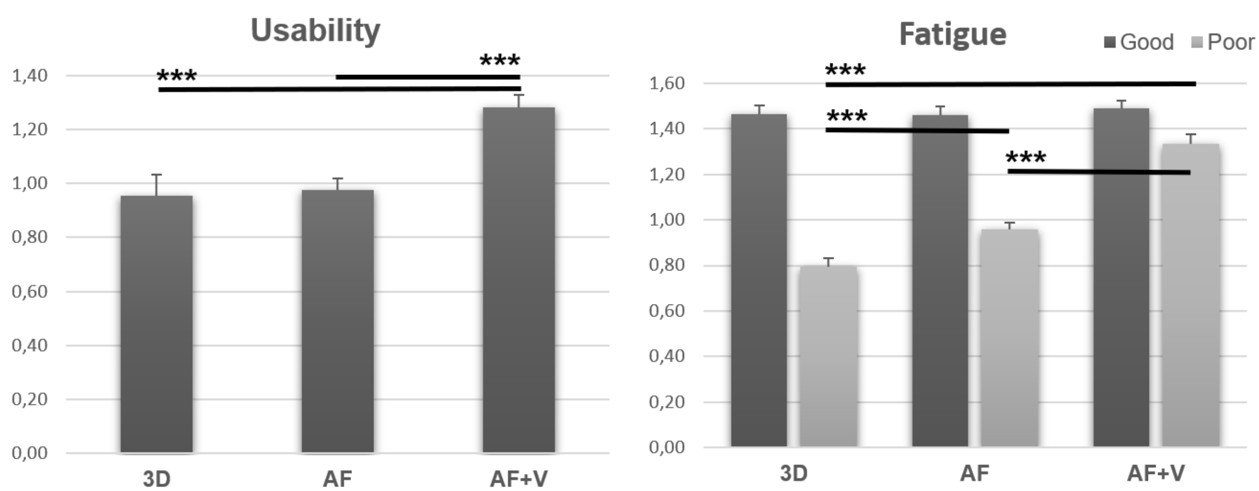

Figure 8. Results from inferential analysis on questionnaire data. Error bars are standard errors.

5.2.3. Fatigue

Main effects for Modality and Visibility factors and a significant Modality × Visibility interaction were found. Detailed Fatigue results are reported in Table 4.

Variable	ddl	*F*	*p*	η^2_p
Modality	2, 42	60.23	< .0001	.74
Visibility	1, 21	138.09	< .0001	.87
Modality × Visibility	2, 42	67.23	< .0001	.76

Table 4. Results from 3 × 2 (Modality [3D, AF, AF+V] × Visibility [Good, Poor]) ANOVA with Fatigue score dependent variable implemented as within factor.

Tukey's HSD post-analysis revealed the following main results (Figure 8). For the Modality factor main effect: participants have felt less fatigue using AF+V modality (M = 1.41, SD ± .03) than using AF modality (M = 1.21, SD ± .02, $p < .001$) and 3D modality (M = 1.13, SD ± .02, $p < .001$). They also perceived less fatigue using AF modality than 3D modality ($p < .05$). For the Modality × Visibility interaction, no significant results were found in good visibility conditions. However, in poor visibility conditions, participants perceived less fatigue using AF+V modality (*M* = 1.33, *SD* ± .04) than AF modality (*M* = .96, *SD* ± .03, $p < .001$) and 3D modality (*M* = .8, *SD* ± .04, $p < .001$). They also perceived less fatigue using AF modality than 3D modality ($p < .001$).

5.2.4. NASA-TLX

Results are reported for each NASA-TLX parts in Table 5 and Figure 9. For the Mental demand part, a main effect was found on the Modality factor [$F(2, 42) = 8.94$, $p < .001$, $\eta^2 = .3$]. Tukey's HSD post-hoc analysis revealed that participants felt that AF+V required a smaller mental demand (*M* = .86, *SD* ± .04) than 3D modality (*M* = 1.03, *SD* ± .06, $p < .001$). Also, AF modality required a smaller mental effort (*M* = .93, *SD* = .04) than 3D one ($p < .05$). No significant difference was found between AF+V and AF modalities. Concerning the Performance part, a main effect was found on the Modality factor [$F(2, 42) = 55.97$, $p < .0001$, $\eta^2 = .73$]. Tukey's HSD post-hoc analysis revealed that participants felt more efficient using AF+V modality (*M* = 1.31, *SD* ± .05) than AF modality (*M* = .95, *SD* ± .03, $p < .001$) and 3D modality (*M* = .8, *SD* ± .03, $p < .001$). They also felt more efficient using AF modality than 3D modality ($p < .05$). For the Effort part, a main effect was found on the Modality factor [$F(2, 42) = 13.01$, $p < .0001$, $\eta^2 = .38$]. Tukey's HSD post-hoc analysis revealed that AF+V modality less effort from the required the participants (*M* = .78, *SD* ± .04) than AF modality (*M* = .99, *SD* ± .03, $p < .001$) and 3D modality (*M* = 1.02, *SD* ± .05, $p < .001$). No significant difference was found between AF and 3D modalities. Finally for the Frustration section, a main effect was found on the Modality factor [$F(2, 42) = 24.53$, $p < .0001$, $\eta^2 = .54$]. Tukey's HSD post-hoc analysis revealed that participants felt less frustrated using AF+V modality (*M* = .57, *SD* ± .04) than AF modality (*M* = .86, *SD* ± .05, $p < .001$) and 3D modality (*M* = .94, *SD* ± .06, $p < .001$). No significant difference was found between AF and 3D modalities. No significant results were found concerning Physical demand and Temporal demand parts.

NASA-TLX	**Variable**	**ddl**	F	p	η^2_p
Mental demand	Modality	2, 42	8.94	< .001	.3
Physical demand	Modality	2, 42	1.4	.26	.06
Temporal demand	Modality	2, 42	2.68	.08	.11
Performance	Modality	2, 42	55.97	< .0001	.73
Effort	Modality	2, 42	13.01	< .0001	.38
Frustration	Modality	2, 42	24.53	< .0001	.54

Table 5. Results from ANOVAs with each NASA-TLX sections implemented as within factor.

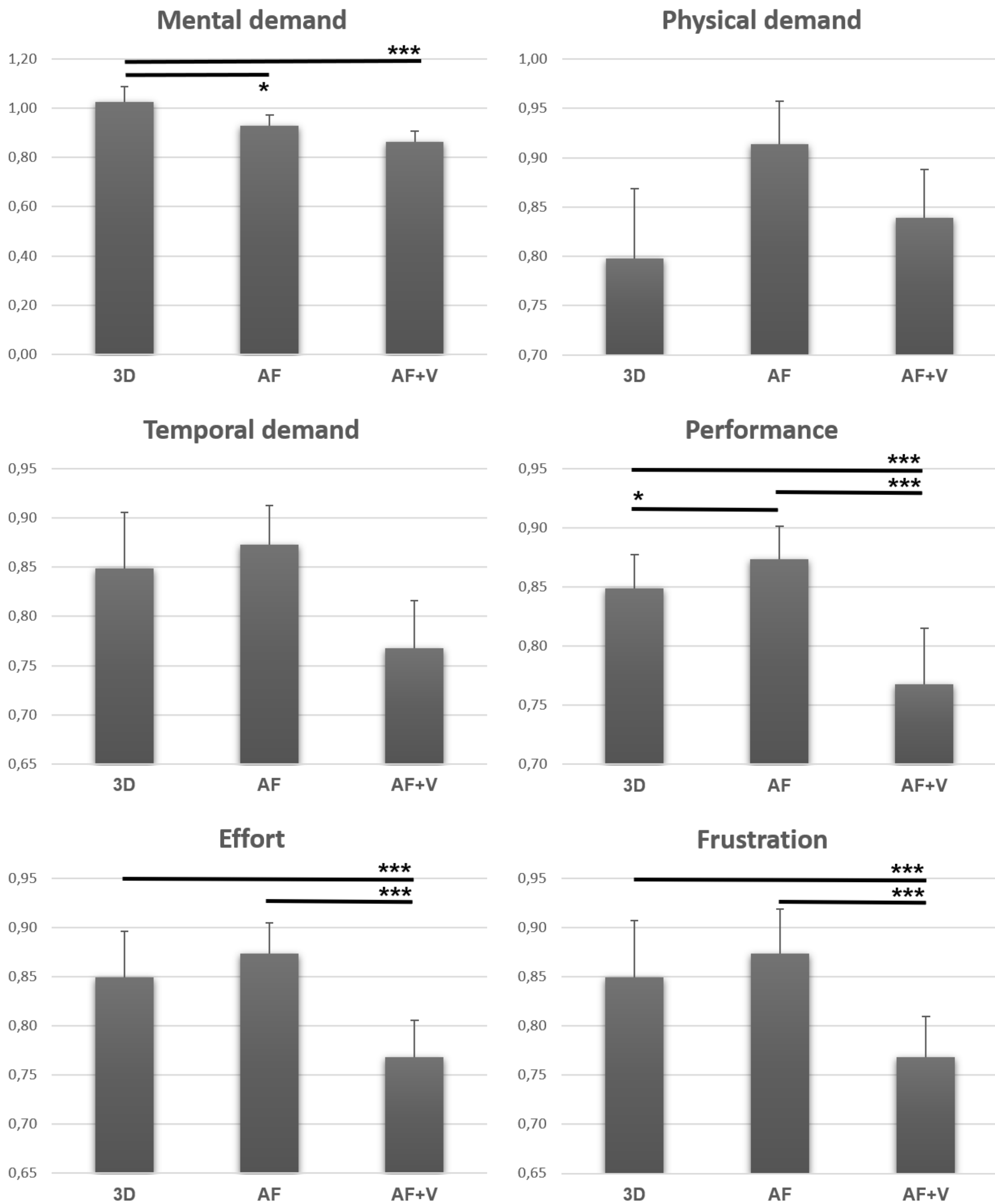

Figure 9. Results from inferential analysis on NASA-TLX data. Error bars are standard errors.

5.3. Summary of the results

To clarify our findings, we propose in this section a simplified view of the results mentioned previously. Table 6 summarizes the main results found under poor visibility conditions.

Variable	**Simplified result**	**Comments**
Accuracy	AF+V > AF > 3D	More accurate
	AF+V > AF	Faster
Response Time	AF+V < 3D	Slower
	AF < 3D	Slower
Usability	AF+V > AF > 3D	More usable
Fatigue	AF+V > AF > 3D	Require less fatigue
NASA-TLX Mental demand	AF+V > 3D	Smaller mental demand
NASA-TLX Performance	AF+V > AF > 3D	More efficient
NASA-TLX Effort	AF+V > AF > 3D	Less effort
NASA-TLX Frustration	AF+V > AF > 3D	Less frustrating

Table 6. Summary of the results in poor visibility conditions. Right column gives an indication how the symbols > and < should be read.

6. DISCUSSION

6.1. General discussion

In this study we proposed two new interaction modalities, which enable ATCos to detect where aircraft are located in a remote control environment using the senses of hearing and/or touch, especially when the visibility conditions are poor and do not allow them the see the aircraft. The Audio Focus interaction modality was designed to enhance the selectivity process of the human ear in a spatial sound environment. Vibrotactile feedback has been added to the Audio Focus interaction modality for further improvement. The results are in line with our first hypothesis: perceived locations of aircraft were significantly more accurate when visibility was poor using AF and AF+V modalities compared to the 3D one. However, this was achieved with relatively longer response times. More precisely, response times were slightly higher with AF interaction in poor visibility conditions: whereas ATCos take about 6.3 seconds on average to locate aircraft with only spatial sound, they took about 9.4 seconds using AF interaction, and about 8.5 seconds when AF was coupled with vibrotactile feedback, which is relatively small regarding the ATCos tasks. This additional time could be explained by the fact that this kind of interaction requires the users to make head movements to locate sound sources, while only spatial sound is almost instantaneous. This can be viewed as negligible when this result is confronted with the actual benefit in terms of accuracy: there was a 49 % location accuracy using 3D modality, 61 % using AF modality and finally, 90 % using AF+V modality.

Subjective results revealed a clear preference for AF modality coupled with vibrotactile feedback. Using the latter, participants perceived it as easier to locate aircraft when visibility was poor than with only spatial sound. Likewise, perceived fatigue was reduced when AF modality was coupled with vibrations. However, participants felt that spatial sound only necessitated a lower physical demand compared to the other two modalities, which is easily understandable since AF interaction requires the users to make movements, which is not the case for 3D modality. Mental demand was lower when AF modality was coupled with vibrotactile feedback. Also, AF interaction was perceived as more efficient than only spatial sound. When AF interaction was coupled with vibrotactile feedback, it was perceived as the least frustrating and the least demanding in terms of effort to be provided for the task.

All these results suggest that a feature such as AF interaction could be useful in a remote control context. Indeed, according to the participants, 81 % of them had failed at least once during their previous working experience to locate an airplane. When it occurred, the solutions they had chosen to solve this problem were multiple. Notwithstanding, AF interaction modality coupled with vibrations can be seen as an interactive, natural and intuitive feature. There was a clear preference to use this modality (95.7 %) which suggests that the AF concept is generally seen as useful, especially when coupled with vibrotactile feedback. The participants found it helpful to locate sound sources with an average score of 5.82 out of 7.

Technologies, for example collaborative maps in military field or electronic strips in ATM (Hurter et al., 2012), are not always well welcomed by their end-users as they represent a significant change compared to those they are meant to replace or simplify (Cohen and McGee, 2004). This is not the case of AF interaction, which seems to be acceptable for the participants who tested it. AF interaction was designed with the consideration that ATCos often search for aircraft they are controlling in their field of view. From this perspective, the AF modality does not ask the users to deeply modify their habits as they already move their head when searching for aircraft through the window. Moreover, we know that HMI which supports the use of several modalities (for the present case, stimuli) fosters mutual disambiguation and reduces the number of errors (Cohen and McGee, 2004; Oviatt, 1999). Our results confirm this tendency with the combined use of spatial sounds and vibrotactile feedback.

As a result, additional perceptual information that could be provided by spatial audio should be studied for remote operations in ATM. In our study, this was considered as a baseline (3D modality). However, adding sound information to a context (e.g. ATC) where operators are not always used to dealing with it, could be seen as uncomfortable. A solution to this potential drawback could be to integrate it as an on-demand feature.

6.2. Limitations

These subjective results have to be interpreted with care because of the way the questionnaire were delivered. They were not completed during or immediately after each condition but immediately after the whole experiment itself. We wanted the participants to stand back sufficiently from the different interaction modalities before asking them to give scores to each experienced modality. In other words, we did not want the participants to re-evaluate the rating scales afterwards because of the introduction of a new interaction modality. We are aware that there is indeed a bias linked to possible forgetting after experiencing the different conditions. In fact, it may cause the loss of recency effect. However, this bias should impact all the conditions in an equivalent way because of the randomization of the modality presentation. Actually, the participants had the same number of trials, difficulty levels, visual conditions and modalities, but in a different order from one another.

6.3. Conclusion and perspectives

The achieved results suggested that Audio Focus is an interaction form that is well perceived and demonstrates promising behavioral results in terms of accuracy in the perception of aircraft location in the airport vicinity under poor visibility conditions. It was proven that interactive spatial sound coupled with vibrotactile feedback improves the sound source selectivity process. Nevertheless, a further study could be performed in a setting closer to operational constraints in order to confirm these findings for the ATC field. Besides, such a technique can easily be exported beyond ATM. In particular, fields that can be cited are visual handicap, Virtual/Augmented Reality (V/AR) or video games. Some studies in visual handicap have already suggested other interaction concepts close to the Audio Focus interaction modality (Munatt and Werber, 1994; Criespien et al., 1994). Since the sensory channels brought into play here are those of hearing and touch, the AF interaction modality could be useful for people with visual impairments who are in situations that require them to develop a mental map of their environment or simply to interact with spatially structured HMIs. For the same purpose, this kind of interaction could also be integrated into V/AR environments. Finally, Audio Focus could also be a playful concept of interaction in the field of video games.

This research opens new perspectives to enhance HCI quality and efficiency in ATC with the final aim of improving user performance and increasing safety.

ACKNOWLEDGEMENTS AND FUNDINGS

We would like to thank all the air traffic controllers who took part in this experiment as well as all the people involved in the project. This work was conducted in the frame of the MOTO (embodied reMOte TOwer) project, and was financed by the European Commission (Horizon2020 project Sesar-06-2015, GA n. 699379).

REFERENCES

Andéol, G., Suied, C., Scannella, S., & Dehais, F. (2017). The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers. *Journal of the Association for Research in Otolaryngology, 18(3)*, 457-464. https://doi.org/10.1007/s10162-016-0611-7.

Aricò, P., Reynal, M., Imbert, J. P., Hurter, C., Borghini, G., Di Flumeri, G., ... & Pozzi, S. (2018, July). Human-machine interaction assessment by neurophysiological measures: a study on professional air traffic controllers. In the *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4619-4622). IEEE.

Bau, O., Poupyrev, I., Israr, A., & Harrison, C. (2010, October). TeslaTouch: electrovibration for touch surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 283-292). ACM. https://doi.org/10.1145/1866029.1866074.

Bolt, R. A. (1981, August). Gaze-orchestrated dynamic windows. In *ACM SIGGRAPH Computer Graphics* (Vol. 15, No. 3, pp. 109-119). ACM. https://doi.org/10.1145/800224.806796.

Braathen, S. (2011). Air transport services in remote regions. International Transport Forum Discussion Paper.

Brooks Jr, F. P., Ouh-Young, M., Batter, J. J., & Jerome Kilpatrick, P. (1990, September). Project GROPEHaptic displays for scientific visualization. In *ACM SIGGraph computer graphics* (Vol. 24, No. 4, pp. 177-185). ACM. https://doi.org/10.1145/97879.97899.

Brungart, D. S., & Simpson, B. D. (2001). Auditory localization of nearby sources in a virtual audio display. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* (pp. 107-110). IEEE. https://doi.org/10.1109/ASPAA.2001.969554.

Calvo, J. A. (2014). SESAR Solution Regulatory Overview – Single Airport Remote Tower (Technical report). SESAR H2020.

Carter, T., Seah, S. A., Long, B., Drinkwater, B., & Subramanian, S. (2013, October). UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (pp. 505-514). ACM. https://doi.org/10.1145/2501988.2502018.

Chatty, S. (2003). The Ivy Software Bus. http://www.eei.cena.fr/products/ivy/documentation/ivy.pdf. Last accessed September 15, 2018.

Cheng, C. I., & Wakefield, G. H. (1999, September). Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. In *Audio Engineering Society Convention 107*. Audio Engineering Society.

Cohen, P. R., & McGee, D. R. (2004). Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1), 41-46.

Cordeil, M., Dwyer, T., & Hurter, C. (2016, November). Immersive solutions for future air traffic control and management. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces* (pp. 25-31). ACM. https://doi.org/10.1145/3009939.3009944.

Crispien, K., Würz, W., & Weber, G. (1994, September). Using spatial audio for the enhanced presentation of synthesised speech within screen-readers for blind computer users. In *International Conference on Computers for Handicapped Persons* (pp. 144-153). Springer, Berlin, Heidelberg.

Esposito, N., & Lenay, C. (2011, November). FeelTact: rich tactile feedback for mobile gaming. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (p. 71). ACM.. https://doi.org/10.1145/2071423.2071511.

Fontana, F., Camponogara, I., Cesari, P., Vallicella, M., & Ruzzenente, M. (2016). An exploration on whole-body and foot-based vibrotactile sensitivity to melodic consonance. *Proc. of SMC*.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuß, A., & Kaltenhäuser, S. (2009). Steps towards the virtual tower: remote airport traffic control center (RAiCe). *Reconstruction*, 1(2), 14.

Gray, R., Ho, C., & Spence, C. (2014). A comparison of different informative vibrotactile forward collision warnings: does the warning need to be linked to the collision event?. *PloS one*, *9*(1), e87070. https://doi.org/10.1371/journal.pone.0087070.

Guldenschuh, M., & Sontacchi, A. Application of transaural focused sound reproduction. In *6th Eurocontrol INO-Workshop 2009*.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Hermann, T., & Hunt, A. (2011). *The sonification handbook* (pp. 399-425). J. G. Neuhoff (Ed.). Berlin, Germany: Logos Verlag.

Ho, C., Tan, H. Z., & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(6), 397-412.

Hurter, C., Lesbordes, R., Letondal, C., Vinot, J. L., & Conversy, S. (2012, May). Strip'TIC: exploring augmented paper strips for air traffic controllers. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 225-232). ACM. https://doi.org/10.1145/2254556.2254598.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human-computer interaction*, 1(4), 311-338. https://doi.org/10.1207/s15327051hci0104_2.

Jensen, M. J., Tolbert, A. M., Wagner, J. R., Switzer, F. S., & Finn, J. W. (2011). A customizable automotive steering system with a haptic feedback control strategy for obstacle avoidance notification. *IEEE Transactions on vehicular technology, 60*(9), 4208-4216. https://doi.org/10.1109/TVT.2011.2172472.

Loftin, R. B. (2003). Multisensory perception: Beyond the visual in visualization. *Computing in Science & Engineering, 5*(4), 56-58. https://doi.org/10.1109/MCISE.2003.1208644.

Massie, T. H., & Salisbury, J. K. (1994, November). The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems* (Vol. 55, No. 1, pp. 295-300).

Meng, F., Gray, R., Ho, C., Ahtamad, M., & Spence, C. (2015). Dynamic vibrotactile signals for forward collision avoidance warning systems. *Human factors, 57*(2), 329-346. https://doi.org/10.1177/0018720814542651.

Mélan, C., & Galy, E. (2011). Recall performance in air traffic controllers across the 24-hr day: influence of alertness and task demands on recall strategies (pp. 35-54). InTech.

Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America, 30(4)*, 237–246.

Moehlenbrink, C., & Papenfuss, A. (2011, September). ATC-monitoring when one controller operates two airports: Research for remote tower centres. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 76-80). Sage CA: Los Angeles, CA: Sage Publications.

Nene, V. (2016). Remote Tower Research in the United States. In *Virtual and Remote Control Tower* (pp. 279-312). Springer, Cham. https://doi.org/10.1007/978-3-319-28719-5_13.

Oviatt, S. (1999, May). Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 576-583). ACM. https://doi.org/10.1145/302979.303163.

Papenfuss, A., & Friedrich, M. (2016, September). Head up only—a design concept to enable multiple remote tower operations. In *Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th* (pp. 1-10). IEEE. https://doi.org/10.1109/DASC.2016.7777948.

Petermeijer, S. M., Cieler, S., & De Winter, J. C. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis & Prevention, 99*, 218-227. https://doi.org/10.1016/j.aap.2016.12.001.

Raj, A. K., Kass, S. J., & Perry, J. F. (2000, July). Vibrotactile displays for improving spatial awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 1, pp. 181-184). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/154193120004400148.

Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4757-4145-2.

Savidis, A., Stephanidis, C., Korte, A., Crispien, K., & Fellbaum, K. (1996, April). A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. In *Proceedings of the second annual ACM conference on Assistive technologies* (pp. 117-123). ACM. https://doi.org/10.1145/228347.228366.

Schwalk, M., Kalogerakis, N., & Maier, T. (2015). Driver support by a vibrotactile seat matrix–Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. *Procedia Manufacturing, 3*, 2466-2473. https://doi.org/10.1016/j.promfg.2015.07.507

Skaburskis, A. W., Shell, J. S., Vertegaal, R., & Dickie, C. (2003, April). Auramirror: artistically visualizing attention. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 946-947). ACM. https://doi.org/10.1145/765891.766086.

Sklar, A. E., & Sarter, N. B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human factors, 41*(4), 543-552. https://doi.org/10.1518/001872099779656716.

Smith, D., Donald, M., Chen, D., Cheng, D., Sohn, C., Mamuji, A., ... & Vertegaal, R. (2005, April). OverHear: augmenting attention in remote social gatherings through computer-mediated hearing. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1801-1804). ACM. https://doi.org/10.1145/1056808.1057026.

Stiefelhagen, R., Yang, J., & Waibel, A. (2001, November). Estimating focus of attention based on gaze and sound. In *Proceedings of the 2001 workshop on Perceptive user interfaces* (pp. 1-9). ACM. https://doi.org/10.1145/971478.971505.

Van Erp, J. B. F., Jansen, C., Dobbins, T., & Van Veen, H. A. H. C. (2004, June). Vibrotactile waypoint navigation at sea and in the air: two case studies. In *Proceedings of EuroHaptics* (pp. 166-173).

Van Erp, J. B., Van Veen, H. A., Jansen, C., & Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP), 2*(2), 106-117. https://doi.org/10.1145/1060581.1060585.

Van Schaik, F. J., Roessingh, J. J. M., Lindqvist, G., & Fält, K. (2010). Assessment of visual cues by tower controllers, with implications for a Remote Tower Control Centre.

Vliegen, J., & Van Opstal, A. J. (2004). The influence of duration and level on human sound localization. *The Journal of the Acoustical Society of America, 115*(4), 1705-1713. https://doi.org/10.1121/1.1687423.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America, 94*(1), 111-123. https://doi.org/10.1121/1.407089.

Wilson, E., Underwood, M., Puckrin, O., & Letto, K. & Doyle, R. & Caravan, H., & Camus, S. & Bassett, K. (2013). The arcsine transformation: has the time come for retirement. https://www.mun.ca/biology/dschneider/b7932/B7932Final10Dec2010.pdf. Last accessed September 15, 2018.

Wilson, G., Carter, T., Subramanian, S., & Brewster, S. A. (2014, April). Perception of ultrasonic haptic feedback on the hand: localisation and apparent motion. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1133-1142). ACM. https://doi.org/10.1145/2556288.2557033.