



HAL
open science

Non-visual interaction concepts: considering hearing, haptics and kinesthetics for an augmented remote tower environment

Maxime Reynal

► **To cite this version:**

Maxime Reynal. Non-visual interaction concepts: considering hearing, haptics and kinesthetics for an augmented remote tower environment. Human-Computer Interaction [cs.HC]. ISAE-SUPAERO, 2019. English. NNT: . tel-02919969v2

HAL Id: tel-02919969

<https://theses.hal.science/tel-02919969v2>

Submitted on 28 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)*

Présentée et soutenue le *08/11/2019* par :

Maxime Reynal

**Concepts d'Interaction Non Visuelle:
Prise en compte de l'ouïe, du toucher et de la proprioception pour un
environnement de tour de contrôle déportée.**

ANDREAS KERREN
SÉBASTIEN TREMBLAY
MICKAËL CAUSSE
NADINE ROUILLON COUTURE
CHRISTOPHE HURTER
JEAN-PAUL IMBERT

JURY
Professor
Professeur
Professeur
Professeur d'Université
Professeur
Ingénieur de recherche

Rapporteur
Rapporteur
Président du jury
Membre du jury
Directeur de thèse
Co-directeur de thèse

École doctorale et spécialité :

EDSYS : Informatique 4200018

Unité de Recherche :

UR Laboratoire de Recherche ENAC

Directeur(s) de Thèse :

Christophe Hurter et Jean-Paul Imbert

Rapporteurs :

Andreas Kerren et Sébastien Tremblay

Non-Visual Interaction Concepts

Considering hearing, haptics and kinesthetics for an augmented remote tower environment

A dissertation submitted in partial fulfillment
of the requirements for the Degree of

Doctor of Philosophy

Author:
Maxime Reynal

November 2019

National Higher French Institute of Aeronautics and Space (ISAE-Supaéro)
Held at the French Civil Aviation School (ENAC)
In the department of Data, Economy and Visualization (DEVI)



**Université
de Toulouse**

Abstract

In an effort to simplify human resource management and reduce operational costs, control towers are now increasingly designed to not be implanted directly on the airport but remotely. This concept, known as *remote tower*, offers a “digital” working context: the view on the runways is broadcast remotely using cameras located on site. Furthermore, this concept could be enhanced to the control of several airports simultaneously from one remote tower facility, by only one air traffic controller (*multiple remote tower*). These concepts offer designers the possibility to develop novel interaction forms. However, the most part of the current augmentations rely on sight, which is largely used and, therefore, is sometimes becoming overloaded. In this Ph.D. work, the design and the evaluation of new interaction techniques that rely on non-visual human senses have been considered (e.g. hearing, touch and proprioception). Two experimental campaigns have been led to address specific use cases. These use cases have been identified during the design process by involving experts from the field, appearing relevant to controllers due to the criticality of the situation they define. These situations are a) poor visibility (heavy fog conditions, loss of video signal in remote context), b) unauthorized movements on ground (when pilots move their aircraft without having been previously cleared), c) runway incursion (which occurs when an aircraft crosses the holding point to enter the runway while another one is about to land), and d) how to deal with multiple calls associated to distinct radio frequencies coming from multiple airports. The first experimental campaign aimed at quantifying the contribution of a multimodal interaction technique based on spatial sound, kinesthetic interaction and vibrotactile feedback to address the first use case of poor visibility conditions. The purpose was to enhance controllers’ perception and increase overall level of safety, by providing them a novel way to locate aircraft when they are deprived of their sight. 22 controllers have been involved in a laboratory task within a simulated environment. Objective and subjective results showed significantly higher performance in poor visibility using interactive spatial sound coupled with vibrotactile feedback, which gave the participants notably higher accuracy in degraded visibility. Meanwhile, response times were significantly longer while remaining acceptably short considering the temporal aspect of the task. The goal of the second experimental campaign was to evaluate 3 other interaction modalities and feedback addressing 3 other critical situations, namely unauthorized movements on ground, runway incursion and calls from a secondary airport. We considered interactive spatial sound, tactile stimulation and body movements to design 3 different interaction techniques and feedback. 16 controllers’ participated in an ecological experiment in which they were asked to control 1 or 2 airport(s) (Single Vs. Multiple operations), with augmentations activated or not. Having no neat results regarding the interaction modalities into multiple remote tower operations, behavioral results shown a significant increase in overall participants’ performance when augmentation modalities were activated in single remote control tower operations. The first campaign was the initial step in the development of a novel interaction technique that uses sound as a precise means of location. These two campaigns constituted the first steps for considering non-visual multimodal augmentations into remote tower operations.

Keywords: Human-Computer Interaction, Haptics, Kinesthetics Interaction, Sonification, Remote Control Tower

Résumé

Afin de simplifier la gestion des ressources humaines et de réduire les coûts d'exploitation, certaines tours de contrôle sont désormais conçues pour ne pas être implantées directement sur l'aéroport. Ce concept, connu sous le nom de *tour de contrôle distante* (*remote tower*), offre un contexte de travail "digital" : la vue sur les pistes est diffusée via des caméras situées sur le terrain distant. Ce concept pourrait également être étendu au contrôle simultanés de plusieurs aéroports à partir d'une seule salle de contrôle, par un contrôleur seul (tour de contrôle distante *multiple*). Ces notions nouvelles offrent aux concepteurs la possibilité de développer des formes d'interaction novatrices. Cependant, la plupart des augmentations actuelles reposent sur la vue, qui est largement utilisée et, par conséquent, parfois surchargée. Nous nous sommes ainsi concentrés sur la conception et l'évaluation de nouvelles techniques d'interaction faisant appel aux sens non visuels, plus particulièrement l'ouïe, le toucher et la proprioception. Deux campagnes expérimentales ont été menées. Durant les processus de conception, nous avons identifié, avec l'aide d'experts du domaine, certaines situations pertinentes pour les contrôleurs aériens en raison de leur criticité: a) la mauvaise visibilité (brouillard épais, perte de signal vidéo), b) les mouvements non autorisés au sol (lorsque les pilotes déplacent leur appareil sans y avoir été préalablement autorisés), c) l'incursion de piste (lorsqu'un avion traverse le point d'attente afin d'entrer sur la piste alors qu'un autre, simultanément, s'apprête à atterrir) et d) le cas des communications radio simultanées provenant de plusieurs aéroports distants. La première campagne expérimentale visait à quantifier la contribution d'une technique d'interaction basée sur le son spatial, l'interaction kinesthésique et des stimuli vibrotactiles, afin de proposer une solution au cas de perte de visibilité sur le terrain contrôlé. L'objectif était d'améliorer la perception de contrôleurs et d'accroître le niveau général de sécurité, en leur offrant un moyen différent pour localiser les appareils. 22 contrôleurs ont été impliqués dans une tâche de laboratoire en environnement simulé. Des résultats objectifs et subjectifs ont montré une précision significativement plus élevée en cas de visibilité dégradée lorsque la modalité d'interaction testée était activée. Parallèlement, les temps de réponse étaient significativement plus longs relativement courts par rapport à la temporalité de la tâche. L'objectif de la seconde campagne expérimentale, quant à elle, était d'évaluer 3 autres modalités d'interaction visant à proposer des solutions à 3 autres situations critiques : les mouvements non autorisés au sol, les incursions de piste et les appels provenant d'un aéroport secondaire contrôlé. Le son spatial interactif, la stimulation tactile et les mouvements du corps ont été pris en compte pour la conception de 3 autres techniques interactives. 16 contrôleurs aériens ont participé à une expérience écologique dans laquelle ils ont contrôlé 1 ou 2 aéroport(s), avec ou sans augmentation. Les résultats comportementaux ont montré une augmentation significative de la performance globale des participants lorsque les modalités d'augmentation étaient activées pour un seul aéroport. La première campagne a été la première étape dans le développement d'une nouvelle technique d'interaction qui utilise le son interactif comme moyen de localisation lorsque la vue seule ne suffit pas. Ces deux campagnes ont constitué les premières étapes de la prise en compte des augmentations multimodales non visuelles dans les contextes des tours de contrôles déportées Simples et Multiples.

Mots-clefs: Interaction Homme-Machine, Interaction haptique, Interaction kinesthésique, Sonification, Tour de Contrôle Déportée

Remerciements

Ce doctorat, longtemps envisagé, n'aurait pas pu être possible sans la contribution, la collaboration ou simplement la présence de certaines personnes. Je souhaite ici les remercier.

Je remercie tout d'abord mon encadrant **Christophe**, sans qui je n'aurais pas eu cette opportunité et qui a trouvé le temps de m'aider à une certaine période relativement chargée du projet. Un grand merci ensuite à mon autre encadrant, **Jean-Paul**, qui a su me guider durant ces 3 années. Ton implication et tes conseils sur les plans organisationnel, technique, méthodologique, rédactionnel, ont contribué à me former un peu plus au *métier* de chercheur.

De même, a big thank to **Pietrò** from La Sapienza University; sometimes, you took the role of my hypothetical third advisor – See you soon around the world! I would also like to thank all the members of my jury, for their hard work on reading this document, for giving me constructive feedback and for the reviewers for correcting and grading it.

Je tiens ensuite à remercier l'équipe ACHIL, ainsi que le personnel de l'ENAC : **Charles** pour avoir partagé avec moi quelques galères technico-temporelles (ainsi que l'architecture Romaine de nuit), **Nadine** pour les conseils fort utiles en statistiques, tout autant que les remarques lors des soutenances blanches, **Mathieu** pour le support technique nécessaire (sans oublier **Philippe**), et **Railane** pour ses qualités managériales. Merci également aux personnels techniques que j'ai souvent embêté, **Serge** et **David**.

Je remercie par ailleurs tous les contrôleurs aériens impliqués dans le projet, sans qui ce travail n'aurait pas pu aboutir tant leur expertise fut nécessaire. En premier lieu **Jérôme**, merci pour ton implication, et désolé pour le temps de publication ! **Christophe Thomas** pour l'intérêt certain que tu portes à ce que nous tentons d'entreprendre; **Rémi**, **Julie** et tous ceux qui ont participé de près ou de loin aux phases de conception et aux expériences.

Par ailleurs : aux partenaires de labeur. Ils m'ont accompagnés durant ces 3 années : **Moc-tar** a.k.a. "Le Bâ****", on l'a fait—tu peux te concentrer sur ta carrière de Youtubeur, désormais; **Antoine**, le premier co-bureau et grand frère de thèse, je te souhaite bonne chance dans ta nouvelle voie; **Maria**, on se croisera lors des conférences—bonne chance pour tes nouvelles responsabilités et encore félicitations ! **Micka** pour les discussions sur le rap et, surtout, la politique; **Ludovic**, prof de sport que l'on n'a pas assez vu. Merci également aux partenaires stagiaires, de passage, et

à tous les autres.

Sans oublier mes amis, qui y sont également pour beaucoup. **Max C.**, viiite, on se voit en Australie, en Écosse ou à Hawaï ? (...ou les trois ?); **Mâsieur Raymond**, bien à vous cordialement; **Max 'M.' M.**, la bière est moins bonne ici, vivement le V&B ! **Greg**, à ton tour d'enflammer la piste... à Glasgow cette fois ? **Raph**, c'est bon, on a tous les deux de bonnes références maintenant ; **Simon Michel, Muza, Cheklei, Lucas & Laura, Benoît & Agathe, Vincent Chariot & Marlène, Papi, Antoine & Julie**, (courage Antoine, tout va bien se passer); **Alex 'Gunther' S.** et **Chnafi C.**, cf. Raph, c'est également valable pour vous; **Max G.** j'ai moins de mérite que toi. À tous les autres du Lot, de Bordeaux, de Toulouse et de Houba, trop nombreux pour être tous cités ici mais à qui je pense également au moment d'écrire ces lignes.

Pour finir, à ma famille : **mes parents** qui m'ont toujours poussé à faire ce que je voulais sans me fixer de limite, et ce encore aujourd'hui, dans la même optique. Merci pour toutes ces années de soutien. Voici le résultat; c'est aussi un peu votre thèse ! À **ma petite sœur** : merci d'avoir relu cette thèse (je sais que ce n'est pas un cadeau¹), le soutien sans faille et la présence de toujours. À **Danielle, Jordy**, si vous lisez ces lignes un jour (on ne sait jamais), j'espère que cela vous fera sourire en ces moments difficiles.

Enfin, à **Ève** : tu m'as également poussé à entamer un doctorat, tout en sachant d'expérience ce que ça allait impliquer. C'est aussi grâce à toi et tu es toujours là. J'ai une chance insolente.

*Entre Toulouse et Glasgow,
Entre Novembre 2019 et Janvier 2020.*

¹...et merci également pour la mise en page !

List of Figures

1	Introduction: MOTO’s project logo	7
2	Introduction: MOTO’s partners logo	8
1.1	ATC and Remote Towers: The different flight phases and ATC types	18
1.2	ATC and Remote Towers: Different types installations.	20
1.3	ATC and Remote Towers: Example of a Multiple Remote Tower	23
1.4	ATC and Remote Towers: RTS from Sweden	24
1.5	ATC and Remote Towers: Avinor Norwegian RCT	24
1.6	ATC and Remote Towers: Alice Springs RCT	25
1.7	ATC and Remote Towers: German ANSP RCT solution	26
1.8	ATC and Remote Towers: Remote Tower timeline	28
1.9	ATC and Remote Towers: Examples of tag, zoom and infrared vision features	29
2.1	Theoretical backgrounds (human senses): The human ear	41
2.2	Theoretical backgrounds (human senses): Spatial sound perception	45
2.3	Theoretical backgrounds (human senses): Somatosensory system	46
3.1	Theoretical backgrounds (sonification): “Sonification interdisciplinary” [78]	59
3.2	Theoretical backgrounds (sonification): Vintage sonification devices	60
3.3	Theoretical backgrounds (sonification): Common sonification process	64
3.4	Theoretical backgrounds (sonification): “World-of-Windows” concept [20]	67
3.5	Theoretical backgrounds (sonification): OverHear system [158]	67
3.6	Theoretical backgrounds (sonification): Crispian <i>et al.</i> works [41, 40]	68
4.1	Theoretical backgrounds (haptics): “Tactile” device [102]	76
4.2	Theoretical backgrounds (haptics): NASA’s Telerobotic Flight Servicer [173, 5]	77
4.3	Theoretical backgrounds (haptics): DataGlove device [189]	78
4.4	Theoretical backgrounds (haptics): GROPE-III haptic device [28]	79
4.5	Theoretical backgrounds (haptics): Mid 90’s haptic devices examples	80
4.6	Theoretical backgrounds (haptics): Examples of PHANToM-type devices [110].	80
4.7	Theoretical backgrounds (haptics): NormalTouch [17] and Haptic Revolver [178].	83
4.8	Theoretical backgrounds (haptics): Tactile feedback and kinesthetic interaction.	87
5.1	First experimental campaign: Design process	98
5.2	First experimental campaign: Horizontal angle resolution for Audio Focus modality	100

5.3	First experimental campaign: Audio Focus modality design principle.	103
5.4	First experimental campaign: Audio signals used for Audio Focus tactons.	106
6.1	First experimental campaign: Runway circuit, airport description and answering HMI	115
6.2	First experimental campaign: Weather conditions	116
6.3	First experimental campaign: Apparatus	118
6.4	First experimental campaign: Combinatorics of the trials	119
6.5	First experimental campaign: Protocol overview	121
6.6	First experimental campaign: Sandbox	125
6.7	First experimental campaign: Behavioral results	127
6.8	First experimental campaign: Usability results	129
6.9	First experimental campaign: Fatigue results	130
6.10	First experimental campaign: Results from NASA-TLX questionnaire	131
6.11	First experimental campaign: Mental workload results	132
7.1	First experimental campaign: Behavioral averaged values	138
8.1	Second experimental campaign: Design process	148
8.2	Second experimental campaign: Situation of runway incursion	150
8.3	Second experimental campaign: Spatial Sound Alert design	152
8.4	Second experimental campaign: Runway incursion button on ground radar HMI	153
8.5	Second experimental campaign: Waveform for tactons used in CSA modality	154
9.1	Second experimental campaign: Orders HMI and pseudo pilots positions	164
9.2	Second experimental campaign: Scenarios and scripts	167
9.3	Second experimental campaign: Visibility conditions	168
9.4	Second experimental campaign: Apparatus	169
9.5	Second experimental campaign: Protocol overview	170
9.6	Second experimental campaign: Subjective measurements	173
9.7	Second experimental campaign: Results from Reaction time measurements	177
9.8	Second experimental campaign: Neurophysiological workload results	180
9.9	Second experimental campaign: Results from post-experiment questionnaire	181

List of Tables

3.1	Theoretical backgrounds (sonification): Design space for earcons [19, 27, 26]. . .	63
4.1	Theoretical backgrounds (haptics): Design space for tactons [25].	85
4.2	Theoretical backgrounds (haptics): Design space for kinesthetic interaction [55].	88
5.1	First experimental campaign: How KI design space [55] was considered for the design of Audio Focus modality (see reference Table 4.2).	102
5.2	First experimental campaign: How tactons design space [25] was considered for the design of Audio Focus modality (see reference Table 4.1).	105
6.2	First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with <i>Accuracy</i> dependent variable implemented as within factor (Audio Focus modality).	124
6.3	First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with <i>Reaction time</i> dependent variable implemented as within factor (Audio Focus modality).	126
6.4	First experimental campaign: Results from one-way ANOVA with <i>Usability score</i> dependent variable implemented as within factor (Audio Focus modality).	128
6.5	First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with <i>Fatigue score</i> dependent variable implemented as within factor (Audio Focus modality).	128
6.6	First experimental campaign: Results from NASA-TLX with each NASA-TLX sections implemented as within factor (Audio Focus modality).	129
6.7	First experimental campaign: Summary of the results obtained during the first experimental campaign regarding the Audio Focus interaction modality. Right column gives an indication how symbol > should be read.	133
7.1	First experimental campaign: Synthesis of the results regarding the hypothesis formulated in Chapter 6.	141
8.1	Second experimental campaign: How earcons design space [26] was considered for the design of SSA modality (see reference Table 3.1 and Figure 8.3 for an illustration).	155

8.2	Second experimental campaign: How tactions design space [25] was considered for the design of CSA modality (see reference Table 4.1 and Figure 8.5 for an illustration).	156
9.2	Second experimental campaign: Start and stop times for the acquisition of reaction times during the second experimental campaign.	171
9.3	Second experimental campaign: Results from 2×2 ANOVA (CI = .95) with repeated measures (Context [Single, Multiple] \times Augmented [No, Yes]) conducted on Reaction time measurements (SSA modality).	176
9.4	Second experimental campaign: Summary of the results obtained during the second experimental campaign regarding SSA, RWY and CSA modalities.	182
10.2	Second experimental campaign: Synthesis of the results regarding the hypothesis formulated in Chapter 9.	191

Special Terms

affordance refers to the capacity of an object or a concept to suggest its own use (this is the point of view from the field of HCI). 53, 61

audification refers to a sonification technique that is only designating the fact to play any music or pre-recorded sound; see Section 3.3.1, p. 60. xi, 60, 61, 97, 202

auditoryicon refers to the event-based sonification technique called audification; see Section 3.3.2, p. 61. xi, 61

Augmented Reality is the fact to add synthetic elements to Real Reality (RR), such as textual information, images or 3D objects. It's a simple form of Mixed Reality (MR). xi–xiii

earcon refers to the synthetical and more abstract form of auditoryicon; see Section 3.3.3, p. 62. xii, 62, 152

embodied cognition is a term coming from social psychology and cognitive science to consider an integration process taking into account our whole haptics and kinesthetic sensations to construct our body's perception in space. xi, xii, 7, 31, 68, 86

embodiment short term for embodied cognition. 87, 89

haptic refers to the sense of touch. When used within HCI field, it qualifies an in-

terface that use cutaneous sensations to make feedback or interaction. xi, 40

haptic rendering is “*the process of computing and generating forces in response to user interactions with virtual objects*” [148]. 75, 82

haptics see haptic. In the field of HCI, this term is used sometimes to designate both haptic and kinesthetic interaction. xi, 49, 75

haptification refers to the fact of providing data to users in an “*haptic*” way. 75

immersive analytics is defined as “*the use of engaging, embodied cognition analysis tools to support data understanding and decision making*” [50]. 55, 66

Information Visualization refers to the field of Computer Science and HCI aiming at displaying various types of structured information to the users. 89

kinesthetic refers to the sensation of body movement. This sensation is perceived through sensory receptors located in muscles, tendons, and joints. This term is used in HCI field to designate an interaction which consider body's movement and proprioception in its design. xi, xiv, 40, 47, 48, 68, 75, 86, 87

kinesthetics see kinesthetic. 49, 73, 75

Mixed Reality is a term used to designate an advanced form of Augmented Reality (AR) in which the user can interact with synthetic objects. xi, xii, xiv

perception refers to a sensation once it has been treated and interpreted by the brain. 40, 48

Real Reality refers to the real world, real-world experiences, interactions and activities, by contrast with the terms Virtual Reality, Augmented Reality, and Mixed Reality. xi, xii, xiv

sensation designates the information directly felt by sensory organs (sensory receptors) through the body. xi, 40, 48, 75

sense of presence see embodied cognition. 7, 22, 31, 44, 67, 87, 162, 172, 181

serendipity means making an unexpected, unplanned discovery. 87, 88, 101

sonification refers to a technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering) ; see Section 3.1, p. 57, xi, 58, 75, 97, 201

tacton refers to a vibratory pattern designed to provide information to the user using haptic means, it is the haptic counterpart of earcon; see Section 4.4.2, p. 84, 84, 100, 154

Virtual Reality refers to technologies that are used to extract the user from RR, by emerging him or her in an entirely synthetic environment, most of the time through the use of headsets. xii, xiv

XR is an emerging term which regroup all or part of the notions of Virtual Reality (VR), AR, and MR. It is generally used by contrast with the term Real Reality. xiv, 44, 141

Acronyms

- 3DS** Spatial sound only (a modality tested during the first experimental campaign). ix, 112, 113, 117, 119, 120, 122, 124, 126–130, 137
- 3DUI** IEEE Symposium on 3D User Interfaces. 76
- AF** Audio Focus (a modality tested during the first experimental campaign). ix, 97, 104, 111–113, 117, 119, 120, 122, 124, 126–130, 132, 137, 166, 174, 182, 189, 190
- AF+V** Audio Focus coupled with Vibrotactile feedback (a modality tested during the first experimental campaign). ix, 104, 112, 113, 117, 119, 120, 122, 124, 126–130, 132, 137, 139
- AFIS** Aerodrome Flight Information Service. 26
- ANOVA** ANalysis Of VAriance. ix, 120, 122, 124, 126–130
- ANSP** Air Navigation Service Provider. 19, 22
- AR** Augmented Reality. xi, xii, 81, 118, 141
- ATC** Air Traffic Control. 6, 9, 10, 17, 18, 31, 126, 165
- ATCo** Air Traffic Controller. 6, 10, 18, 132
- ATM** Air Traffic Management. 7, 17, 31, 163
- ATS** Air Traffic Service. 6, 18
- CHI** ACM Conference on Human Factors in Computing Systems. 51, 57
- CSA** Call from Secondary Airport modality. viii, x, 153, 154, 156, 162, 165, 166, 168, 170, 171, 181, 189
- DIS** ACM Conference on Designing Interactive Systems. 76
- DoF** degree of freedom. 79
- EBS** Event-based Sonification. 52, 60
- EEG** electroencephalography. 65, 122, 123, 125, 169, 175, 182
- EMBC** IEEE Engineering in Medicine and Biology Society (conference). 123
- FDL** Frequency Difference Limen. 43
- GSR** Galvanic Skin Response. 123, 175
- HCI** Human Computer Interaction (journal). 51, 57, 76
- HCI** Human Computer Interaction. 9, 10, 17, 97, 99
- HMI** Human Machine Interface. xiii, 41, 194
- IAD** Interaural Amplitude Difference. 44
- ICAD** International Community for Auditory Displays. 57
- ICMI** ACM International Conference on Multimodal Interaction. 76
- IDE** Integrated Development Environment. 84, 167
- IFD** Interaural Frequency Difference. 44
- IFR** Instrument Flight Rules. 18–20
- IHM** Interaction Homme-Machine (Fr., see Human Machine Interface (HMI)). 194, 197

IJHCS International Journal of Human Computer Studies. 51, 57, 76

ILD Interaural Level Difference. 44

IMC Instrument Meteorological Conditions. 18–20

ITD Interaural Time Difference. 44

KI Kinesthetic Interaction. ix, 86, 87, 100–102

MAA Minimum Audible Angle. 45, 99, 100, 195

MBS Model-based Sonification. 54, 65

MR Mixed Reality. xi, xii, 21, 81

NASA-TLX NASA Task Load Index. ix, 116, 120, 122, 126, 129, 131, 132, 162, 172, 173, 178, 181, 182, 187, 188

PMS Parameter Mapping Sonification. 54, 64

RCT Remote Control Tower. 17, 21, 31

RR Real Reality. xi, xii, 81

RWY Runway Incursion modality. 151, 153, 154, 156, 162, 165, 166, 168, 170, 171, 182, 189, 190

SA Situational Awareness. 22, 31, 162, 172, 174, 179, 181, 182, 189

SDK Standard Development Kit. 117

SESAR Single European Sky ATM Research. 2, 7, 21, 25, 193, 200

SME Subject-Matter Expert. 172, 173, 181, 182, 187, 188

SPL Sound Pressure Level. 43

SSA Spatial Sound Alert modality. ix, 151, 152, 154, 155, 162, 165, 166, 168, 170, 171, 181, 182, 187, 189

TEI ACM International Conference on Tangible, Embedded and Embodied Interactions. 76

TMTF Temporal Modulation Transfer Function. 43

TOCHI ACM Transactions on Computer-Human Interaction. 51, 57, 76

ToH IEEE Transactions on Haptics. 76

TVCG IEEE Transactions on Visualization and Computer Graphics. 51, 57

UbiComp ACM International Joint Conference on Pervasive and Ubiquitous Computing. 51, 57

UIST ACM User Interface Software and Technology Symposium. 51, 57

VFR Visual Flight Rules. 18, 20, 31, 99

VIS IEEE Visualization Conference. 51, 57

VMC Visual Meteorological Conditions. 18, 20

VR Virtual Reality. xii, 10, 81, 141, 198

XR XR. 81, 89

Table of Contents

Abstract / Résumé	i
Remerciements	v
List of figures	viii
List of tables	x
List of terms	xi
List of acronyms	xiii
Contents	xx
Introduction	1
<i>Résumé (Français)</i>	1
Context & motivations	6
MOTO: the embodied reMOte Tower project	7
Thesis problematic	8
Approach	8
Manuscript walkthrough	9
1 Air Traffic Control and Remote Towers	13
<i>Résumé (Français)</i>	13
1.1 Introduction	17
1.2 Presentation of ATC	17
1.2.1 Brief history	17
1.2.2 Presentation of the field	18
1.2.3 Different types of ATC	19
1.2.4 Flight Rules and Meteorological Conditions	20
1.3 Remote Control paradigms	21
1.3.1 Definitions	21
1.3.2 Hybrid Remote Tower	21
1.3.3 Single Remote Tower	22
1.3.4 Multiple Remote Tower	22
1.3.5 Augmented Remote Tower	22

1.4 Existing RCT environment, prototypes and solutions	23
1.4.1 Sweden	23
1.4.2 Norway	23
1.4.3 Australia	24
1.4.4 Ireland	25
1.4.5 Germany	25
1.4.6 Canada and United Kingdom	26
1.4.7 United States	26
1.4.8 France	27
1.5 Sensorial augmentations in RCT solutions	27
1.5.1 Sight	27
1.5.2 Hearing	29
1.5.3 Touch	30
Wrap-up message	31
I Theoretical backgrounds	33
2 Introduction to non-visual perception	35
<i>Résumé (Français)</i>	35
2.1 The human senses	40
2.2 An overview on the auditory system	41
2.2.1 The auditory system	41
2.2.2 Psychoacoustics and perception of sound	42
2.3 The human somatosensory system and the sense of touch	45
2.3.1 The somatosensory system	46
2.3.2 Three different types of touch perception	47
2.4 Human proprioception	48
Wrap-up message	48
3 Relying on sound: Auditory Displays and Sonification	51
<i>Résumé (Français)</i>	51
3.1 Introduction	57
3.2 Brief history	58
3.3 Basic sonification concepts and techniques	60
3.3.1 Audification	60
3.3.2 Auditory icons	61
3.3.3 Earcons	62
3.4 Other techniques	64
3.4.1 Parameter Mapping Sonification	64
3.4.2 Model-Based Sonification	65
3.5 Interacting using sound	65
3.5.1 Sonic interaction and MBS	65

3.5.2 Other examples of sound interaction	66
Wrap-up message	68
4 Relying on haptics: interacting with body and touch	71
<i>Résumé (Français)</i>	71
4.1 Introduction	75
4.2 History of haptics in computer science and early prototypes	76
4.3 The use of haptics in Augmented, Virtual and Mixed realities	81
4.4 Tactile stimulation to provide information	82
4.4.1 Tactile devices and current technologies	82
4.4.2 Tactile sensations to provide structured information: the concept of tactons	84
4.4.3 Tactile sensations to alert users and to provide spatial information	84
4.5 Kinesthetic Interaction	86
Wrap-up message	89
II Audio Focus modality: involving audification, passive perception and kinesthetic interaction	91
5 Designing Audio Focus interaction modality	93
<i>Résumé (Français)</i>	93
5.1 Introduction	97
5.2 Design approach	97
5.3 Specific observations resulting from the user-centered approach	98
5.4 Informal hypotheses	99
5.5 Improving spatial sound source location in poor visibility conditions	100
5.5.1 Auditory channel and proprioception	101
5.5.2 Touch channel: passive perception through vibrotactile feedback	104
Wrap-up message	104
6 First experimental campaign: Audio Focus evaluation	107
<i>Résumé (Français)</i>	107
6.1 An experiment to evaluate Audio Focus modality	111
6.2 Method	111
6.2.1 Participants	111
6.2.2 Nature of the task and ethics consideration	112
6.2.3 Experimental conditions	112
6.2.4 Hypotheses	113
6.2.5 Airport discrimination using selectable areas	113
6.2.6 Answering interface	114
6.2.7 Metrics	114
6.2.8 Apparatus	116
6.2.9 Combinatorics of the trials	117

6.2.10 Experiment protocol	119
6.3 Data analysis	120
6.3.1 Behavioral data	120
6.3.2 Subjective data	120
6.3.3 Neurophysiological data	122
6.4 Behavioral results	124
6.4.1 Accuracy	124
6.4.2 Reaction time	124
6.5 Subjective results	126
6.5.1 General questions	126
6.5.2 Usability	128
6.5.3 Fatigue	128
6.5.4 NASA Task Load Index	129
6.6 Neurophysiological results	132
Wrap-up message	132
7 First experimental campaign: Discussion	135
<i>Résumé (Français)</i>	135
7.1 Introduction	137
7.2 Behavioral results	137
7.3 Subjective results	138
7.4 Neurophysiological, behavioral and subjective results comparison	139
7.5 Limitations	140
7.6 Synthesis	140
7.7 Conclusion	140
III Multimodal experience in ecological Remote Control Tower environment	143
8 Designing multimodal augmentations for an ecological RCT environment	145
<i>Résumé (Français)</i>	145
8.1 Introduction	148
8.2 Design approach	148
8.3 Identification of relevant ATC situations	149
8.3.1 Location of abnormal events	149
8.3.2 Runway incursion awareness	149
8.3.3 Multiple radio calls in case of multiple RCT	150
8.4 HCI techniques for specific ATC use cases	150
8.4.1 Spatial alert as earcon to provide location of abnormal events	151
8.4.2 Vibrotactile feedback to alert from critical events	151
8.4.3 Other vibrotactile feedback to distinguish between radio calls	153

<i>Wrap-up message</i>	154
9 Second experimental campaign: Multimodal augmentations evaluation	157
<i>Résumé (Français)</i>	157
9.1 Multimodal augmentations in ecological RCT environment	162
9.2 Method	162
9.2.1 Participants	162
9.2.2 Experimental task overview	163
9.2.3 Hypothesis	163
9.2.4 Scenarios	165
9.2.5 Experimental conditions	167
9.2.6 Apparatus	168
9.2.7 Experimental protocol	168
9.3 Measurements and data analysis	171
9.3.1 Behavioral data	171
9.3.2 Subjective data	172
9.3.3 Subject Matter Expert ratings	172
9.3.4 Neurophysiological data	174
9.4 Experimental results	175
9.4.1 Performance analysis	176
9.4.2 Perceived workload analysis	178
9.4.3 Neurophysiological analysis	179
9.4.4 Situational Awareness	179
9.4.5 Post-experiment interview analysis	179
<i>Wrap-up message</i>	181
10 Second experimental campaign: Discussion	183
<i>Résumé (Français)</i>	183
10.1 Introduction	187
10.2 Performance	187
10.3 Perceived workload	188
10.4 Results from neurophysiological measurements and performance analyses	188
10.5 Other subjective results	189
10.6 Limitations	190
10.7 Synthesis	191
10.8 Conclusion	191
Conclusion & Perspectives	193
<i>Résumé (Français)</i>	193
Thesis summary	200
Scientific contributions	202
Conclusion	203
Perspectives	204

List of publications	206
Bibliography _____	207
Appendices	223
First experimental campaign _____	223
Ethics form	223
CERNI form	238
Consent form	240
Welcome text	243
Post-experiment questionnaires	245
Second experimental campaign _____	253
Experimental scripts	253
SASHA questionnaire (SA)	258
Questionnaires (post-experiment guided interview)	260

Introduction

Résumé (Français): Introduction

Contexte et motivations

Pour certains aéroports dotés d'un trafic aérien très faible (environ deux vols par jour), le contrôle aérien, sur le plan logistique, devient de plus en plus compliqué à mettre en place. D'un point de vue économique, ces aéroports s'avèrent être en fait de moins en moins rentables. Pour pallier à ce problème, une solution a été imaginée consistant à déporter le lieu dans lequel se fait le contrôle aérien dans des centres distants. Ainsi, les ressources techniques, logistiques et humaines pourraient être gérées dans un même lieu, ce qui faciliterait l'organisation générale et réduirait les coûts opérationnels. Les autorités de la navigation aérienne et certains laboratoires de recherche ont d'ailleurs d'ores et déjà formaliser des recommandations pour la poursuite d'un développement technologique dans la direction du contrôle aérien déporté [23, 61]. Différentes solutions sont déjà en développement en Europe et dans le reste du monde [120]. Le présent travail de doctorat s'inscrit dans ce sujet de recherche, et a pour but d'améliorer la tâche de travail des contrôleurs aériens évoluant dans un tel contexte, et par conséquent le niveau de sécurité général.

Ces centres seraient composés de plusieurs salles dans lesquelles se feraient de manière effective le contrôle des aéroports distants. La principale différence entre une tour de contrôle *sur site* (ou *physique*) et une tour de contrôle *distante* (ou *déportée*) est que dans ce dernier cas, les contrôleurs aériens utilisent une vue dématérialisée, digitale, sur l'aéroport contrôlé, en parallèle des habituels radars et autres outils plus communs (e.g., moyens radio). Dans une tour de contrôle physique, un certain nombre de stimuli sensoriels leur parviennent, de manière implicite ou explicite. Par exemple, le bruit d'un moteur qui démarre sur l'aire de stationnement indique l'emplacement et la direction de l'avion en question. De manière implicite, cela informe également que le pilote entrera bientôt en communication pour demander des instructions de roulage. Un autre exemple de stimulations sont les oscillations perçues de la tour de contrôle, pouvant indiquer des rafales de vent. Par conséquent, certaines informations implicites qui pouvant être cruciales pour les contrôleurs à un moment donné pourraient être perdues dans un environnement de contrôle à distance. Cependant, et précisément parce que l'environnement de contrôle est dématérialisé, le concept de tour de contrôle déportée offre de nouvelles perspectives en terme d'interaction.

Les contrôleurs aériens font un usage intensif de leur sens visuel. Parfois même trop, ce qui peut saturer ce canal sensoriel. Bien sûr, cette utilisation intensive est justifiée: naturellement, la vision est l'un des canaux sensoriels sur lesquels nous nous appuyons le plus; c'est également

le cas pour les contrôleurs aériens. Elle peut être prédominante dans notre façon de percevoir le monde. La largeur de bande (c.-à-d. la quantité d'informations pouvant être transmises) lui étant attribuée est de loin la plus grande si nous la comparons aux autres canaux sensoriels. Cependant, différentes communautés s'intéressent de plus en plus à ces autres canaux sensoriels dans le but de créer de l'interaction ou de faire parvenir de l'information à l'utilisateur. Par exemple, les personnes ayant une déficience visuelle ont besoin de techniques IHM qui ne reposent pas sur la vision. Les concepteurs souhaitent aussi créer des interactions plus écologiques, en utilisant les sens dont les humains sont naturellement pourvus, tout en recherchant constamment à améliorer l'immersion, la performance et l'utilisabilité. C'est pourquoi l'ouïe, le toucher et les autres sens (odorat, goût, proprioception), associés ou non à la vue, sont de plus en plus considérés comme des moyens possibles d'interaction avec les utilisateurs. Dans ce contexte, l'objet de ces travaux de doctorat est de considérer les sens humains non visuels pour l'interaction dans le contexte du contrôle aérien à distance.

Le projet MOTO

Le financement européen MOTO (“the embodied reMOte TOwer”) est un projet Horizon 2020 dans le cadre de SESAR Action pour la Recherche et l'Innovation (RIA). Son objectif était d'identifier les principaux stimuli multimodaux requis dans un environnement de contrôle aérien à distance, dans le but d'améliorer le *sense de présence* ressenti par les contrôleurs aériens. Le projet MOTO a débuté en juin 2016 et a duré 24 mois, jusqu'en juin 2018. L'état de l'art des neurosciences cognitives a montré que l'information multisensorielle joue un rôle important pour que l'être humain se sente efficacement immergé dans un environnement. La recherche et le développement en matière de performance humaine dans le domaine de la gestion du trafic aérien et des opérations de contrôle à distance a été principalement axée sur les sens visuel et auditif. Par conséquent, deux grandes possibilités de recherche avaient été identifiées :

- Considérer plus de canaux sensoriels dans le cadre du contrôle aérien à distance. En plus d'une approche centrée sur la vue, l'approche dite de la *cognition incarnée* a également été appliquée dans le but d'obtenir une certaine compréhension de l'utilisation des autres canaux sensoriels.
- Explorer des concepts innovants d'interaction homme-système, basés sur d'autres canaux perceptuels comme le toucher, l'audition ou la proprioception. Ce doctorat est précisément inscrit dans ce dernier point.

Le projet MOTO s'est déroulé en s'appuyant sur les compétences de quatre partenaires clés, ayant chacun une expérience en recherche et/ou en développement :

- La société DeepBlue : spécialisée dans la R&D et le transfert de technologies. Son principal domaine d'activité est l'évaluation de systèmes complexes. Elle se concentre sur les interactions, l'intégration et la répartition des fonctions entre les humains, les procédures et les équipements dans des contextes où les exigences en matière de sécurité, de fiabilité et de résilience sont élevées.

- L'Université de la Sapienza à Rome : a été impliquée pour les connaissances des personnes impliquées dans le projet MOTO en neurosciences et non liées au domaine de l'aéronautique.
- L'École Nationale de l'Aviation Civile (ENAC), Université de Toulouse : a fournit les connaissances aéronautiques nécessaires au projet. Elle également fournit son expertise en IHM. C'est sur son campus de Toulouse que les expérimentations les plus "lourdes" du projet ont eu lieu, profitant de ses équipements technologiques adaptés.
- L'Université de Groningen : la deuxième plus ancienne université des Pays-Bas, c'est une institution de recherche de pointe dotée d'installations innovantes pour l'exploration visuelle et l'analyse de grands ensembles de données. MOTO implique l'Institut Johann Bernoulli de mathématiques et d'informatique.

Problématique

Cette thèse s'inscrit donc dans la partie du projet associée à l'ENAC, et concerne principalement les parties liées à la considération des techniques et connaissances issues du domaine de l'interaction Homme-Machine. Cette discipline a donc été considérée dans le but d'améliorer l'expérience utilisateur et la performance au sein des environnements de tour de contrôle déportée.

La problématique générale de cette thèse peut alors être formulée ainsi:

“Au sein d'un environnement immersif de tour de contrôle distante et augmentée, l'ajout d'interactions et de retours d'informations multimodaux basés sur des canaux sensoriels non-visuels peut-il contribuer à l'amélioration de la performance des contrôleurs aériens en situation critique ?”

Approche

Pour répondre à cette problématique, deux grandes phases ont été considérées en dehors du nécessaire travail bibliographique. Ces deux phases ont toutes deux menées à des campagnes expérimentales conçues pour tester les hypothèses formulées. La première expérience s'est déroulée en laboratoire, dans un environnement très contrôlé. Elle avait pour but d'analyser la contributions d'une première modalité interactive que nous avons appelé "Audio Focus". La seconde expérience, quant à elle, s'est déroulée au sein d'un environnement que l'on a voulu plus réaliste, plus écologique. Dans la continuité de la première campagne, notamment au travers de la réutilisation de la modalité Audio Focus, celle-ci a eu pour but d'analyser la contribution d'autres modalités interactives multimodales.

Organisation du manuscrit

Ce manuscrit de thèse est divisé en trois grandes parties. Un chapitre d'introduction, précédent un autre chapitre introduisant les éléments de base de l'aéronautique, ainsi qu'une conclusion générale en fin, s'ajoutent aux 3 parties principales. La première partie correspond à l'état de l'art, qui est ici présenté au travers de trois chapitres distincts mais logiquement enchaînés. Le deux

parties suivantes sont consacrées à la présentation chronologique des deux campagnes expérimentales qui ont été menées dans le cadre du projet MOTO. Elles sont tous deux composées d'un chapitre relatif à la conception de modalités interactives, puis d'un chapitre relatif à leur évaluation expérimentale, et d'un chapitre de discussion. La première d'entre elles présente l'ensemble du processus de conception envisagé pour concevoir et évaluer la modalité interactive Audio Focus, tandis que la seconde présente l'ensemble du processus de conception visant à développer et évaluer trois autres modalités amenées à être introduites, avec Audio Focus, dans un environnement écologique de contrôle déporté.

Chapter 1: *Air Traffic Control and Remote Towers* → Dans ce premier chapitre suivant l'introduction, les différents concepts aéronautiques utiles à la compréhension du présent travail sont introduits, définis et expliqués. Ainsi, après un bref historique de l'aviation, plutôt orienté vers la gestion du ciel, les différents types de contrôles aériens sont introduits, et notamment la notion de contrôle déporté. Nous y définissons les concepts sous-jacents et y présentons quelques-unes des solutions de contrôle déporté déjà existantes. Le lien avec l'informatique et le domaine de l'IHM est fait en fin de chapitre, où nous explorons comment les différents canaux sensoriels peuvent être utilisés dans le cadre du contrôle déporté.

Part I: *Theoretical backgrounds* → Cette première partie correspond à l'état de l'art. Les trois chapitres qui la composent introduisent les notions théoriques utiles au développement de cette thèse.

Chapter 2: *Introduction to non-visual perception* → En premier lieu est introduit le système auditif humain. Nous y présentons comment le son est capté, puis acheminé jusqu'au cerveau pour y être interprété. Certaines caractéristiques de l'audition humaine utiles à la suite de ce travail sont introduites. Nous présentons ensuite le système somatosensoriel humain. Le toucher et la proprioception y sont développés, notamment les différents types de perceptions liées au toucher.

Chapter 3: *Relying on sound: Auditory Displays and Sonification* → Il s'agit d'une revue de la littérature de la prise en compte du son en informatique. Il se focalise donc sur l'audition en IHM. Plus précisément, les notions d'Auditory Displays et de Sonification y sont abordées. Un historique est proposé, puis différents termes, concepts et techniques liés aux domaines de la sonifications sont développés, tels que les icônes audio ou la sonification basée-modèle. Le chapitre se termine par une étude bibliographique de concepts interactifs basés totalement ou partiellement sur le son.

Chapter 4: *Relying on haptics: interacting with body and touch* → Ce chapitre se focalise sur la proprioception et le toucher. Un historique des périphériques haptiques est présenté. Puis, la considération de l'haptique en réalité virtuelle y est abordée, ainsi que les différents concepts, techniques et périphériques liés de près ou de loin à la perception haptique, comme par exemple les tactons. Le chapitre se termine sur une présentation des concepts et techniques interactives basées sur le mouvement.

Part II: *Audio Focus modality: involving audification, passive perception and kinesthetic interaction* → Dans cette partie nous présentons en détails la première campagne expérimentale qui fut envisagée lors de ce travail.

Chapter 5: *Designing Audio Focus interaction modality* → Ce chapitre relate le processus de conception de la modalité Audio Focus. Celle-ci, basée à la fois sur du son spatialisé et interactif, des stimuli vibratoires et l'interaction kinesthésique, a été développée en collaboration avec des professionnels du contrôle aérien. Elle répond à une situation ayant été isolée avec leur aide et correspondant à un cas d'utilisation précis.

Chapter 6: *First experimental campaign: Audio Focus evaluation* → Ce chapitre relate de manière exhaustive tous les processus expérimental ayant été envisagé dans le but de quantifier la contribution de la modalité interactive Audio Focus. 22 contrôleurs aériens ont participé volontairement à une expérience de laboratoire. Les résultats de cette expériences sont présentés en détail en fin de chapitre.

Chapter 7: *First experimental campaign: Discussion* → Dans ce chapitre, les résultats précédents sont alors discutés sous l'angle du contrôle aérien et de l'IHM. Nous y fournissons nos conclusions par rapport à la modalité Audio Focus et proposons des perspectives à court et moyen termes.

Part III: *Multimodal experience in ecological Remote Control Tower environment* → De la même manière que la partie précédente, cette partie correspond à la présentation complète de la seconde campagne expérimentale.

Chapter 8: *Designing multimodal augmentations for an ecological RCT environment* → Dans ce chapitre nous retranscrivons la méthode de conception que nous avons utilisée pour concevoir les trois autres modalités testées dans ce manuscrit (à savoir l'Alerte Spatiale Sonore, l'Incursion de Piste et l'Appel provenant du Second Aéroport). Ces trois modalités correspondent elles-aussi à des situations identifiées avec des professionnels du domaine.

Chapter 9: *Second experimental campaign: Multimodal augmentations evaluation* → Tout comme pour le second chapitre de la partie précédente, dans celui-ci nous présentons l'ensemble du processus expérimental envisagé pour tester la contribution de ces trois modalités. Une seconde fois, des contrôleurs ont pris part à une expérience. Celles-ci s'est déroulée dans des conditions techniques les plus réelles possibles afin d'être en mesure de tester les modalités dans un environnement écologique. Les résultats engendras sont présentés en fin de chapitre.

Chapter 10: *Second experimental campaign: Discussion* → Ce chapitre correspond à la discussion des résultats précédemment obtenus. La contribution de chacune des modalités y est discutées ainsi que les perspectives.

Enfin, le document se termine par une conclusion générale sur le travail mené durant ces 3 années. La première section de chaque chapitre en est un résumé en langue française.

Context and motivations

For some airports with very low traffic density², providing constant Air Traffic Service (ATS) is increasingly becoming a complex logistic task. So that from an economic point of view, such airports are becoming less profitable. One solution could be to centralize Air Traffic Control (ATC) into centers in which air traffic would be remotely controlled and human resources would be brought together. Incidentally, Air Navigation Authorities and laboratories have already contributed information and recommendations for further development in this direction [23, 61], and different solutions are already in development across Europe and worldwide [120]. The present Ph.D. work falls in this research field with the aim of enhancing Air Traffic Controllers (ATCos) performance, and consequently, safety.

These centers would be composed of several rooms called Remote Control Towers (or simply Remote Towers; this will be further defined in next Chapter I, Section 1.3, p. 21). The main difference between an *on-site* (or *physical*) control tower and a *remote* one is that in a latter, ATCos are working using a dematerialized view of the controlled airport, among radars and other common facilities such as radio means. In an on-site 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.

ATCos make extensive use of their visual sense. Sometimes even too much, which can make this sensory channel saturated. Of course, this intensive use is justified: vision is one of the sensory channels on which we naturally rely the most. It may be predominant in our way to perceive the world. The bandwidth (i.e., the amount of information that can be conveyed through) assigned to this sensory channel is by far the larger one compared to other ones. However, different communities, especially from the field of Human-Computer Interaction, are increasingly investigating other sensory channels to make interaction or to provide information to the users. For example, people with visual impairments need HCI techniques that do not rely on vision. Researchers and practitioners also want to build interactions that are more ecological, using the senses that humans are naturally equipped with, constantly wishing to enhance immersion, performance, and affordance. This is why hearing, touch (i.e., haptics), and other senses (smell, taste, proprioception), coupled or not with sight, are more and more considered as possible ways to interact with users. In this context, the subject of this Ph.D. works is to consider non-visual human senses for interaction in remote control context. It was conducted within the framework the MOTO project.

²Approximately 2 flights per day.

MOTO: the embodied reMOte TOWer prject

European funding MOTO is a Horizon 2020 project in the framework of the SESAR Research and Innovation Action (RIA). Its aim was to identify the key multimodal stimuli required on remote air traffic control platforms to enhance the sense of presence experienced by ATCos. This embodiment improvement was ultimately expected to benefit ATCos' attention levels, their decision-making process and facilitate real-world knowledge and skills transfer to a simulated work-site. MOTO project has initiated in June 2016, lasted approximately 24 months, until June 2018.

The state of the art of cognitive neuroscience has shown that multi-sensory information plays an important role for humans to effectively feel immersed in an environment. Human performance R&D in Air Traffic Management (ATM) and remote control operations has been mainly focused in visual and auditory senses. Consequently, two major research opportunities were identified:

- To *consider the role of all the human senses in tower operations*. Embodied cognition approach was applied to achieve a full understanding on the use of all senses for ATCos, beside the visual one.
- To *explore innovative human-system interaction concepts* on the understanding of embodied aspects of interaction, i.e., considering other perceptual channels like haptics, hearing or proprioception, and of ATM human performance. This Ph.D. is precisely inscribed within this last point.

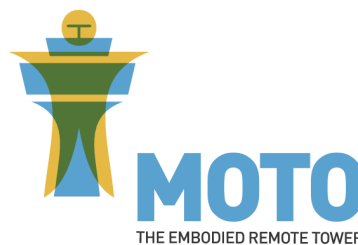


Figure 1 — MOTO project's logo.

MOTO project was built on the complementarity of four core partners that cover the whole value chain from research to application and development of prototypes. ENAC and Deep Blue provided the primary venues for exploitation, having existing contacts with key industrial partners, while both Sapienza University research teams have corresponding sister spin-off companies that could ensure support for mature exploitation. The University of Groningen had supported the team in designing new visualization techniques. Here is a more detailed presentation of the different partners:

- DeepBlue: this is a company specialized in R&D and technology transfer. Its main area of activity is the evaluation of complex systems. Focusing on the interactions, integration and allocation of functions between humans, procedures and equipment in contexts with high security, dependability and resilience requirements, like ATM.

- Sapienza University of Roma: has brought state-of-the-art knowledge from domains other than aviation. MOTO involves both the Industrial Neuroscience and the Social and Cognitive Neuroscience laboratories.
- French Civil Aviation School (ENAC), University of Toulouse: has provided continuing education for the main players of the French civil aviation, ENAC also offers a favorable environment for research activities: it has its own training facilities, access to ATM experts as well as engineers and pilots, and a long-term expertise in innovative interaction technology, simulation and modelling.
- University of Groningen: the second-oldest university in the Netherlands, it is a leading research institution featuring state-of-the-art facilities for the visual exploration and analysis of huge datasets, together with know-how in large-scale data visualization. MOTO involves the Johann Bernoulli Institute for Mathematics and Computing Science.



Figure 2 — The logos of the different MOTO’s partners: ENAC (top left), Sapienza University (top right), DeepBlue (bottom left), university of Groningen (bottom right).

Thesis problematic

This thesis is therefore inscribed into the part of MOTO project which is associated to ENAC, and mainly concerns the topics related to the consideration of techniques and knowledge coming from the field of Human-Computer Interaction. This discipline was therefore considered with the aim of improving the user experience and performance within Remote Control Tower environments.

The general problematic of this thesis can then be formulated as follows:

“In the immersive environment of Remote and Augmented Control Towers, can multimodal interaction and feedback based on non-visual sensory channels improve the performance of controllers during critical Air Traffic Control events?”

Approach

In order to answer this problem, two main phases have been considered, apart from the necessary bibliographical work. These two phases both led to experimental campaigns designed to test the hypotheses we formulated. The first experiment took place in laboratory, highly controlled

environment. It aimed at analyzing the contributions of a first interactive modality that we called “Audio Focus”. The second experiment took place in an environment that we wanted to be more realistic, more ecological. In the continuity of the first campaign, notably through the reuse of the Audio Focus modality, this one aimed at analyzing the contribution of other multimodal interactive modalities.

Manuscript walkthrough

This Ph.D. thesis manuscript is divided into three main parts, identified by an introductory chapter, preceding another chapter introducing the basic elements of aeronautics, and a general conclusion. The first part corresponds to the state of the art, which is presented here in three distinct but logically linked chapters. The next two parts are devoted to the chronological presentation of the two experimental campaigns that were carried out as part of the MOTO project, presented above, for which ENAC was in charge and which constitute the core of this doctoral work. They are both composed of a chapter presenting the design of interactive modalities, then a chapter presenting their experimental evaluation, and a discussion chapter. The first presents the entire development process related to Audio Focus interaction modality (*Second Pilot Exercise*), and the second presents the entire development process aimed at developing three other modalities to be introduced, along Audio Focus, in an ecological remote control environment (*Second Validation*).

Chapter 1: *Air Traffic Control and Remote Towers* → In this first chapter after the introduction, the different aeronautical concepts useful for understanding this doctoral work are introduced, defined and explained. Hence, after a brief history of aviation, more oriented towards sky management, the different types of ATC are presented and the notion of remote control is introduced. We define the underlying concepts and present some of the existing remote control solutions, in an experimental or commercial form. The link with computer science and the field of Human Computer Interaction (HCI) is made at the end of the chapter, where we explore how the different human sensory channels can be used in the context of remote control.

Part 1: *Theoretical backgrounds* → This first part corresponds to the state of the art. Its three chapters introduce the theoretical notions useful for the development of this thesis.

Chapter 2: *Introduction to non-visual perception* → First of all, the human hearing system is introduced. We present how sound is captured and then sent to the brain for interpretation. Some characteristics of human hearing that are useful for this study are introduced. We then present the human somatosensory system. Touch and proprioception are developed, including the different types of perceptions related to touch.

Chapter 3: *Relying on sound: Auditory Displays and Sonification* → This is a review of the literature on the consideration of sound in Computer Science. Eventually, it focuses on hearing in HMI. More precisely, the notions of Auditory Displays and Sonification are introduced. A brief history is proposed, then different terms, concepts and techniques related to the fields of sound engineering are developed, such as auditory icons or model-based

sonification. The chapter ends with a bibliographical study of interactive concepts based on sound.

Chapter 4: *Relying on haptics: interacting with body and touch* → This chapter focuses on proprioception and touch. A brief history of haptic technologies is presented. Then, the consideration of haptics in VR is introduced, as well as the different concepts, techniques and devices closely or far related to haptic perception, such as tactons. The chapter ends with a presentation of interactive concepts and techniques based on body movement, including kinesthetic interaction.

Part II: *Audio Focus modality: involving audification, passive perception and kinesthetic interaction* → In this section we present in detail the first experimental campaign that was envisaged during this work.

Chapter 5: *Designing Audio Focus interaction modality* → This chapter describes the entire design process of the Audio Focus modality. Based on both spatial and interactive sound, vibrotactile feedback and kinesthetic interaction, it has been developed in collaboration with ATCos. It responds to a situation that has been isolated with their help which is then materialized using a specific use case.

Chapter 6: *First experimental campaign: Audio Focus evaluation* → This chapter exhaustively describes all the experimental processes that have been considered to quantify the contribution of the Audio Focus interaction modality. 22 ATCos voluntarily participated in a laboratory experiment. The results are introduced in detail at the end of the chapter.

Chapter 7: *First experimental campaign: Discussion* → In this chapter, the results previously presented are discussed from the perspective of ATC and HCI. We provide our conclusions in relation to the modality tested and propose short and medium-term perspectives.

Part III: *Multimodal experience in ecological Remote Control Tower environment* → In the same way as in the previous part, this one corresponds to the complete presentation of the second experimental campaign.

Chapter 8: *Designing multimodal augmentations for an ecological RCT environment* → In this chapter we re-translate the design method we used to design the other three modalities tested in this experiment (i.e., Spatial Sound Alert, Runway Incursion and the Call from the Second Airport modalities). These three modalities also correspond to situations identified with professionals in the field.

Chapter 9: *Second experimental campaign: Multimodal augmentations evaluation* → As in the second chapter of the previous section, this one presents the entire experimental protocol envisaged to test the contribution of these three modalities. Another time, ATCos took part in an experiment with the aim of providing them with the most real conditions possible and thus testing the modalities in an ecological environment. The results obtained are presented at the end of the chapter.

Chapter 10: *Second experimental campaign: Discussion* → This chapter corresponds to the discussion of the results previously obtained. The contribution of each of the three modalities is discussed as well as the perspectives.

The first section of each of these chapters is a summary in French. The introductory and concluding chapters are introduced by a French translation of their content. In the next chapter, therefore, we introduce the basic aeronautical concepts needed for this study.

Chapter 1

Presentation of the field of Air Traffic Control and the concept of Remote Control Tower

Résumé (Français)

Présentation du domaine du Contrôle Aérien et du concepts des Tours de Contrôle Déportées

Le contrôle du trafic aérien, ou ATC

La *Gestion du Trafic Aérien* est une discipline de l'aéronautique qui est apparue dans les régions du monde dotées d'un trafic aérien dense dans le but de résoudre des problèmes de congestion de l'espace aérien. De nos jours, elle désigne l'ensemble des activités menées dans le monde entier pour assurer la sécurité ainsi que la fluidité du trafic aérien. Le contrôle de trafic aérien (ATC en anglais pour *Air Traffic Control*) en est une composante. L'ATC peut être divisé en 3 types distincts :

- le contrôle *en route*, qui concerne la phase de croisière des vols commerciaux,
- le contrôle d'*approche*, en charge des phase de descente, de montée, de décollage et d'atterrissage,
- le contrôle *sol* ou d'*aéroport*, qui concerne le trafic au sol.

La Figure [1.1](#) ci-après dans le corps du chapitre présente une vue de ces différents types de contrôle aérien. Ce travail de doctorat porte exclusivement sur le contrôle de types *approche* et *aéroport*.

L'ATC est un service rendu par les contrôleurs aériens. Ce service consiste à fournir le contrôle aérien d'un espace aérien, ce qui implique des problématiques liées à la logistique aérienne ainsi qu'au maintien d'un niveau de sûreté nécessaire. En fonction du type du contrôle rendu, la tâche de travail des contrôleurs aérien peut se dérouler dans une tour de contrôle avec vue sur l'espace aérien contrôlé (c'est le cas des contrôles d'approche et d'aéroport), ou bien dans une salle sans vue vers l'extérieur (car inutile), dans le cas du contrôle en route. Le contrôle aérien peut être fournit

par des instances gouvernementales, des compagnies privées ou bien les corps armés, dépendant à la fois de la nature du contrôle concerné ainsi que du pays.

Le concept de tour de contrôle déportée

Le concept de contrôle aérien déporté, comme son nom l'indique, est le fait de fournir ce service d'ATC à distance, c'est-à-dire depuis un point géographique distant de l'espace aérien contrôlé. Comme vu précédemment dans le chapitre introductif de ce document, les contrôleurs aérien sont alors situés dans des centres de contrôle à distance, dans lesquels ils font leur travail depuis une installation nommée Tour de Contrôle Déportée (ou RCT en anglais pour *Remote Control Tower*). Le niveau de sûreté du service ATC rendu doit donc être au moins équivalent à celui de la tour de contrôle physique correspondante située sur le terrain de l'aéroport ou à l'aérodrome distant concerné. Ce service est donc réalisé de manière à l'aide de caméras situées sur l'aéroport, dont le flux vidéo est acheminé pour être diffusé dans l'environnement de la tour déportée, d'où se fait effectivement l'activité de contrôle aérien. Le son capté sur l'aéroport peut également être acheminé jusqu'à la tour déportée. Le concept de tour de contrôle déportée peut être décliné en 4 sous-concepts évoqués dans l'introduction.

De nombreux projets existent d'ores et déjà dans plusieurs pays. Ces concepts, à l'état de prototype ou de réelles solutions commerciales clefs en main, sont majoritairement développés par la Suède (Figure 1.4), la Norvège (Figure 1.5), l'Australie (Figure 1.6), l'Irlande, l'Allemagne (Figure 1.7), le Canada, le Royaume Uni, les États Unis, la France ou encore l'Estonie. Ces environnements nouveaux, de part leur nature digitalisée engendrée par la reproduction dans la tour déportée de la vue sur l'aéroport distant, offrent un support supplémentaire à l'interaction. Ainsi, les solutions existantes et les concepts actuels de tour de contrôle déportée proposent des fonctionnalités telles que la vision infrarouge, le zoom, ou encore la vue directionnelle. Des fonctionnalités en réalité augmentée sont également disponibles, par exemple la surimpression d'informations concernant les avions dans l'environnement de l'aéroport (Figure 1.9).

L'utilisation des différents canaux sensoriels en environnement de contrôle déporté

Pour finir cette synthèse, sachant que nous travaillons sur des techniques d'interaction reposant sur des aspects perceptuels, il est important de mentionner les différents canaux sensoriels impliqués dans un environnement de travail ATC commun et de décrire leur reproduction dans les tours déportées. Cela nous permettra d'avoir une vision plus détaillée de l'environnement dans lequel ce travail s'inscrit.

La vision

Les aspects visuels sont essentiels dans une tour de contrôle physique. En plus des différents radars et autres équipements, les contrôleurs aériens utilisent souvent leur vision pour localiser et discriminer les aéronefs en approche, au décollage et au sol. Dans un environnement de tour de contrôle déportée, la vue sur l'environnement contrôlé doit par conséquent être reproduite fidèle-

ment : c'est la fonction des caméras installées sur place. Comme expliqué précédemment, un tel environnement devient propice à toutes sortes d'améliorations visuelles grâce à ce processus de digitalisation. Cependant, les publications existantes relatent pour la plupart des augmentations partiellement ou totalement liées à ce canal sensoriel, car il est préférentiellement utilisé par rapport aux autres sens. Les autres sens dominants, notamment l'ouïe et encore plus le toucher, ne sont en effet souvent pas considérés pour les environnements de contrôle déportés. C'est par exemple le cas dans [36], où Cordeil *et al.* considèrent les contrôle déportés comme un environnement entièrement immersif permettant l'analyse de données [50] en utilisant un affichage tête-porté pour analyser la trajectoire des avions. Par ailleurs, une liste détaillée des repères visuels et des caractéristiques prévues a été identifiée dans [169].

L'audition

L'environnement audio des contrôleurs aériens est souvent pauvre. Il leur est cependant suffisant pour disposer d'une représentation auditive mentale de l'espace aérien contrôlé. En plus des communications radio bien sûr basées sur l'audition et la parole, l'ouïe est souvent stimulée dans une tour de contrôle physique par des événements relativement discrets, captées par nos sens de manière consciente ou non. Par exemple, le vent peut être ressenti sur les murs de la tour (via ses oscillations), ou certains moteurs peuvent être entendus s'ils sont situés relativement proche de celle-ci. Dans une tour de contrôle, les contrôleurs aériens sont en fait assez isolés de la zone qu'ils contrôlent et leur environnement audio est en fin de compte assez pauvre, mais contribue cependant au sentiment d'immersion. Ce point doit donc être considéré pour les tours de contrôle déportés afin de recréer cet environnement auditif, aussi faible soit-il.

Certaines technologies interactives basées sur le son existent pour les tours de contrôle déportés, mais elles sont assez peu nombreuses. Par exemple, une méthode innovante de spatialisation du son utilisant la stéréo binaurale afin de discriminer les communications des contrôleurs (en route) est présentée dans [69]. Les contrôleurs en route travaillent en binôme. Ils ont tous deux des tâches distinctes, qui se complètent l'une l'autre pour accomplir la tâche de contrôle. Ce système permet de diffuser simultanément deux signaux audio, chacun destiné à l'un des deux contrôleurs, sans affecter l'environnement audio de l'autre.

Le toucher

Comme mentionné plus haut, les oscillations dues au vent peuvent être perçues dans une tour de contrôle physique. Cela représente une partie des stimuli haptiques que les contrôleurs peuvent ressentir. Les moteurs au démarrage ou à l'accélération peuvent également être ressentis de la même manière. Cependant, tout comme l'audition, le toucher et, plus généralement, les sens haptiques, ne sont pas très sollicités dans un contexte aussi isolé.

À la connaissance des auteurs, peu d'études portant sur le sens du toucher dans les tours de contrôle ont été publiées. L'un des exemples les plus concrets de travaux de recherche portant sur la perception tactile dans l'ATC est [101, 85], où les auteurs présentent Strip'TIC, un système visant à remplacer les bandes de papier physiques utilisées par contrôleurs pour suivre

l'état des aéronefs contrôlés.^[1] Cette recherche a été basée sur le fait que l'utilisation de papier à bandes ("strips") ne permet pas de mettre à jour facilement le système avec des informations temps réel relatives aux l'aéronef. Cette observation peut amener à des situations problématiques, alors que l'environnement de travail des contrôleurs aériens est de plus en plus digitalisé. Ces derniers restent attachés à l'aspect tangible de leur tâche de travail et sont généralement réticents à l'abandonner. C'est pourquoi Hurter *et al.* proposent un système combinant un suivi basé sur la vision, une projection augmentée, une visualisation radar et des bandes de papier électroniques utilisables, manipulables à l'aide d'un stylet. Ce type de système est maintenant couramment utilisé, principalement par les contrôleurs en route.

¹Le concept de Strip'TIC est un concept haptique dans le sens où il vise à remplacer le sens du toucher, c'est-à-dire le fait de saisir un objet physique (par exemple des bandes de papier), par un concept IHM.

1.1 Introduction

In this chapter we present the field of ATC. We start with a general presentation of the field in Section 1.2, including some interesting historical aspects (Section 1.2.1), a presentation of the discipline (Section 1.2.2), and a reminder of the different sort of ATC (Section 1.2.3), namely ground control, approach and en-route ATC. We then present the most common flight rules in Section 1.2.4. The main section for our study is the Section 1.3, in which the concept of Remote Control Tower (RCT) is introduced. We first introduce the main definitions in Section 1.3.1. We then define the more precise concepts of Single RCT in Section 1.3.3, Multiple RCT in Section 1.3.4, Hybrid RCT in Section 1.3.2, and Augmented RCT in Section 1.3.5. After this presentation of the concepts related to Remote Towers, we present some of the main existing RCT solutions and market products in Section 1.4. The link between RCT concepts and HCI is done in Section 1.5, where we present some HMIs for ATC and RCT based on sight (Section 1.5.1), hearing (Section 1.5.2), and touch (Section 1.5.3). This chapter ends with a general summary in Section 1.5.3.

1.2 Presentation of ATC

1.2.1 Brief history

The first motorized airship flight took place fifty years before the first Wright brothers' flight on September 24, 1852. This machine was designed by the French engineer Jules Henri Giffard and did not look like the aircraft model that we know today (i.e., a fuselage equipped with two wings, a rear tail and control surfaces). The flight was 27 km long, from Paris to Élancourt. Almost forty years later, in 1889, the first international aeronautical congress was held in Paris to discuss of the aerostation general rules. This resulted in 1891 with the publication of the first aviation law treaty. The first launches of an high speed airship, built by the count von Zeppelin, occurred during the year 1900. Soon after, the American Wright brothers did their first flight with their *Flyer* aircraft, on the 17 of December, 1903. Still, ATC was not yet created. In 1907, the Frenchman Louis Blériot crosses the Channel on-board an “heavier-than-air” aircraft. However, things started to take shape and accelerate years after in 1919, when 18 European countries, on behalf of France, meet at a conference in Paris to establish the legal status of airspace. After this meeting, all civil aircraft started to be equally treated in airspace.² The same year, the *Farman F.60 Goliath* aircraft, from the French enterprise *Avions Farman*, was the first airliner to link Paris to London and Paris to Brussels. Three years later, the Air Navigation International Commission (ANIC) was created. The “Réglementation de la Navigation Aérienne” convention was written the same year, in 1922. Finally, the first version of International Civil Aviation Organization (ICAO), the “Provisional International Civil Aviation Organization” (PICA0), was built twenty two years later in 1944 during the Chicago conference. Its universal and international character was recognized in 1947. It now depends on the United Nations, with the role of participate in the development of policies and standards that allow international ATM standardization. Flights within the same

²Which is referred nowadays as the *right of non offensive transit*.

country are not concerned by ICAO and its head office is located in Montreal.

1.2.2 Presentation of the field

The term *Air Traffic Management* first appeared in areas of the world concerned with dense aircraft traffic. The goal was to solve issues related to airspace traffic congestion, with a rational point of view. ATM refers to all activities carried out to ensure worldwide the safety, but also the fluidity of air traffic. Its organization into 3 main sub-functions was first propose at the Chicago convention in 1944, and includes:

- ATS, which has the purpose to enhance the safety of air traffic. It should prevent collisions between aircraft in the air and on ground, but also between aircraft and obstacles, coordinate and speed-up the traffic. It should pass on data that are relevant to process efficiently and safely the air traffic. Finally, it alerts relevant ATM organs when aircraft are in need of help from emergency services;
- Air Traffic Flow Management (ATFM) is designed to prevent saturation of airports, aerodromes and controlled air sectors. To do this, take-off and landing schedules are known in advance and regulated by a dedicated organization;
- Air Space Management (ASM) is the entity that coordinates the various sub-features of ATM and ensures their inter-compatibility.

ATS is the one that concerns the works related in this manuscript. ATC encompass these functions, and can be divided into 3 distinct types: *En-route*, *Approach*, and *Ground* or *Airport* control (see Figure 1.1). We consider Approach and Ground controls for the present work. Within these different types of ATC, pilots and ATCos always follow specific rules defined by OACI [123] and presented in Section 1.2.4: Visual Flight Rules (VFR), Instrument Flight Rules (IFR), Visual Meteorological Conditions (VMC) and Instrument Meteorological Conditions (IMC).

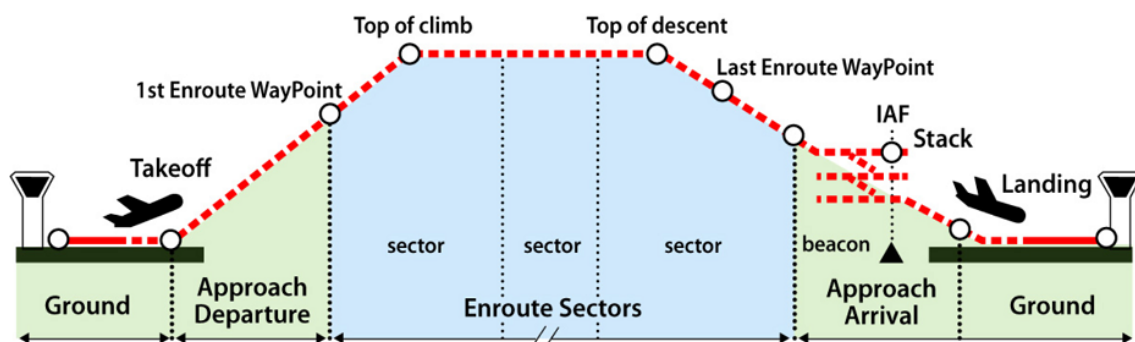


Figure 1.1 — Flight phases and ATC types. The present study is concerned by ATC types in green (Approach and Ground). Circles are the main fix points of a regular commercial flight (credits: Letondal *et al.* [101]).

1.2.3 Different types of ATC

ATC is a service provided by Air Traffic Controllers, whose function is to provide control, safety and management of the air traffic. Depending on the nature of the controlled area, the profession is conducted in the look-out of a control tower (approach and ground controls), or in a control room (en-route control), which is most of the time located at an airport, or in an activity area. According to the concerned country, region or aerodrome, ATC service is provided by private companies, public entities designated with the term Air Navigation Service Provider (ANSP), or the armed forces.

The common point between these 3 ATC services is that when flying, as well as when moving on the ground, aircraft shall be controlled in accordance with Visual Flight Rules and Visual Meteorological Conditions. Instrument Flight Rules and Instrument Meteorological conditions (resp. IFR and IMC) are also defined to supplement visual rules. These are the four main axes that shape the activity of the ATCos.

1.2.3.1 En-route control

En-route centers are responsible for providing ATC services to aircraft in “cruise” flight phase, i.e., when aircraft are flying outside the vicinity of an aerodrome³. En-route ATCos retrieve aircraft managed by approach ATCos at the time of transition from the climb phase to the cruise phase (first en-route waypoints, see Figure 1.1). Then, they manage the trajectories of aircraft in cruise phase, before handing over control to approach ATCos when aircraft leave their cruise phase to transit to approach phase (last en-route waypoint). This type of ATC service do not concern the present manuscript. Actually, approach ATCos’ working task is quite different from the one of en-route ones.

1.2.3.2 Approach control

Approach centers are responsible for providing ATC services to aircraft in the vicinity of an airport, or, more generally, in a controlled area. ATCos are located either in the lookout of a control tower or in a dedicated radar room. Approach ATCos manage aircraft on take-off, initial climb to the top of climb before passing the control to en-route ATCos. They also manage aircraft leaving their cruise phase to enter the initial descent phase, descent, until landing. They also control the trajectory and logistic of aircraft on ground, on taxiways and on the parking area.

1.2.3.3 Ground control

Ground control is responsible for providing ATC services in aerodrome traffic, i.e., in a restricted area around an aerodrome. The difference between ground control and approach control is that the latter concerns airports, while ground control generally concerns smaller areas, known as aerodromes. At some airports, ATCos’ tasks can be to control aircraft on ground, and also aircraft

³In France, there are 5 such centers, which are called *Centres en Route de la Navigation Aérienne (CRNA)*, located in Athis-Mons (near Paris), Reims, Brest, Bordeaux and Aix-en-Provence. The French Civil Aviation School (ENAC) in Toulouse also has a supported en-route center in case one of the other five is out of order.

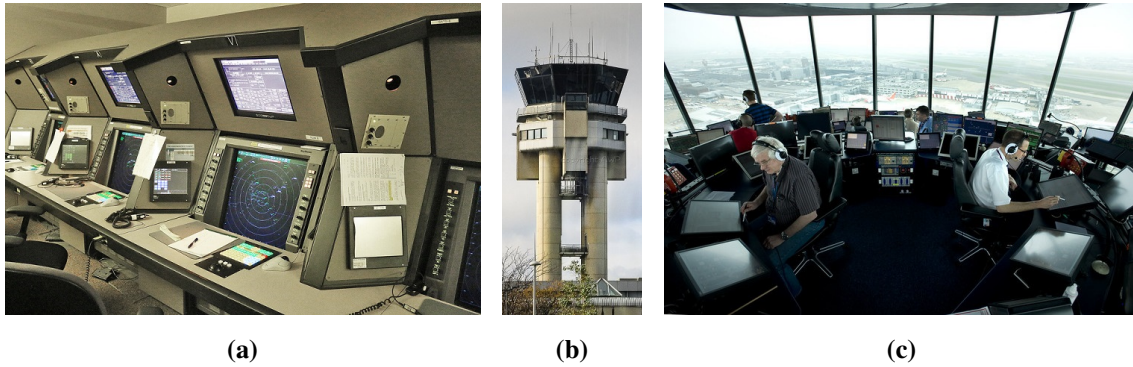


Figure 1.2 — (a) Part of en-route control room of Atlanta airport (b) Toulouse Blagnac approach and ground control tower (c) Inside of London Heathrow NATS control tower.

during take-off, climbing, descent and landing. This is also referred to as ground control and approach control, respectively. The service need to be provided from the lookout of a control tower to have a view on the controlled airport or aerodrome vicinity (“Out-of-the-Window” view).

1.2.4 Flight Rules and Meteorological Conditions

Visual Flight Rules (VFR) represents the simplest and freest flight regime, where collision avoidance is essentially based on the principle of *seeing and avoiding*. They are also the first set of rules to have been used at the beginning of aeronautics and the ones that requires the least number of instruments. VFR flight is only permitted when there is Visual Meteorological Conditions (see hereafter), in which the pilot must usually have a minimum visibility of 1500 to 8000 meters (it varies from one aerodrome to another) and remain clear of cloud (at a minimum distance of 1500 meters horizontally and 300 meters vertically). Special conditions apply to visual flight above cloud (VFR on top), poor visibility (special VFR) or night flight (VFR at night). VFR are opposed to IFR. Instrument Flight Rules (IFR) defines a flight regime where collision avoidance is based on ATC and the respect for specific flight paths. Using indications from aircraft instruments and instructions received from ATCos, the pilot must maintain the aircraft in a flight-specific configuration (altitude, speed), as well as follow a trajectory imposed by ATM authorities to ensure separation from elements that they cannot see (terrain, obstacles and other aircraft), and comply with published regulations and procedures. IFR allow flight in Instrument Meteorological Conditions. Instrument Meteorological Conditions (IMC) refer to conditions in which Visual Meteorological Conditions are not met. In contrast, Visual Meteorological Conditions (VMC) are the minimum conditions for VFR flight. In other words, VMC are the conditions for which the pilots have sufficient visibility not to be required to comply with IFR.

To sum up, the acronym VFR is used to qualify visual flight, using only the sight and vital instruments; while, in opposite way, the acronym IFR is used to qualify instrument flight, where pilots are guided by ATM, predefined flightpath and on-board instruments. The term VMC qualifies the minimum visual conditions necessary for VFR, while IMC refers to non VMC conditions, where only IFR is possible.

1.3 Remote Control paradigms

The definition of the remote ATC concept was adopted by the European Commission under SESAR H2020 consortium. The different related names and acronyms are introduced before detailing the different paradigms of remote controls.

1.3.1 Definitions

Remote Tower Center is basically providing Air Traffic Services remotely. This allows ATCos to provide ATC service from a different location than the airport which is actually controlled. In such a case, ATCos' working task is operated from a Remote Tower Center that can be everywhere but usually at a reasonable distance from all the airports to be controlled (e.g., in order to reduce any latency issue and to increase technical reliability). More specifically, ATCos are working in a room called RCT⁴ [59]. The ATC service is therefore remotely rendered, and its safety level should be equivalent to the related physical control tower located at the concerned airport or aerodrome.

RTS is executed by streaming a real-time view from high definition digital cameras which have been installed at the distant controlled area. The view on the airport and its vicinity is replicated onto LCD screens in the control room to make an Out-of-the-Window image. The fixed cameras give an opportunity to have close-up view on particular details within the airport environment. Also, movable cameras can be directed as required, mimicking the way binoculars may be used in conventional tower. The ATCos' visual attention is supplemented by a range of environmental sensors and microphones capturing sounds, meteorological data, and any other sort of operational data that may serve the purpose of the RTC.

As required in early specifications, remote tower services may improve night and low visibility operations using additional sensing equipment, such as infrared or night-vision lenses and MR overlays. When an airport is expanding quickly, and consequently the view from the original tower may be obstructed, this could be an interesting alternative to the construction of a new physical control tower. Furthermore, using remote tower services can enhance contingency arrangements⁵

The first beneficiaries of remote tower services are regional airports. However, in near future there is almost no doubt that the technology may be used in busier areas or as a redundancy solution for outages in coverage. Using the technology in less dense traffic situations is an interesting opportunity to demonstrate the in-service reliability of the concept, allowing any modification to be made before looking at what other applications might be possible. Hereafter, we define the four distinct paradigms of remote control.

1.3.2 Hybrid Remote Tower

The term of *Hybrid Remote Tower* defines a contingency facility to be used when a physical control tower is unserviceable for a short period of time (e.g., due to renovation, fire, technical failure, etc). While considering this paradigm, a limited number of instruments is envisaged, including at least the instruments which are vital to ATC such as radio, air and ground radars, and goniometers.

⁴Also designated by Remote and Virtual Tower (RVT) within SESAR projects.

⁵Additional features are seen as ways to increase regular standby systems capacity.

Through such a hybrid case, remote control operations will then assure at least a basic level of service. It can be seen as a simpler version of the Single Remote Tower paradigm.

1.3.3 Single Remote Tower

A *Single Remote Tower* is a RCT in which one ATCo is responsible for continuously controlling one and only one airport. Controlling several airports can still be possible in such a case, but only one at a time. This is the first paradigm of RCT to have emerged and the most widespread. It provides a solution for controlling airports located in sparsely populated areas, with low airport traffic but being of vital necessity for the region in which they are located. From a broader perspective, the Single RCT paradigm also makes it possible to control larger airports, such as London Heathrow for example, in an area that is easier to access and where human resources can be more easily managed. It could also represent an alternative solution which could be useful when a failure occurs in the physical tower (e.g., any system failure, power failure, fire, etc).

1.3.4 Multiple Remote Tower

Unlike Single RCT paradigm, *Multiple Remote Tower* allows one ATCo to control several airports *simultaneously*. This requires multiple ratings for each ATCo and careful staffing schedules. This paradigm is still relatively new compared to Hybrid and Single ones, and still requires to be studied because of its relatively huge impact on the ATCo's working task.⁶ In fact, Multiple RCT context has been little studied in scientific literature due to the fact that it is still a rather recent topic. A first study was published in 2010 to demonstrate the feasibility of Multiple Remote Tower, at least to control a minimum of 2 small airports at the same time [127, 116].

Apart from scientific studies, some more concrete projects have recently emerged, such as the SESAR PJ05 project's *Remote Tower For Multiple Airports*⁷, jointly managed by Lithuanian ANSP Oro Navigacija (ON), the national aeronautics and space research center of the Federal Republic of Germany DLR and Frequentis systems supplier. Through different text scenarios, one ATCo provided ATS to 3 Lithuanian airports simultaneously. Also, 6 Lithuanian ATCos managed extensive traffic in a mixed VFR/IFR environment. However, and to the best of the author's knowledge, no deliverable nor scientific publication have been published so far.

1.3.5 Augmented Remote Tower

The *Augmented Remote Tower* paradigm designates one of the abovementioned remote control paradigms when they are enhanced, or augmented, with advanced techniques of interaction and feedback. In this context, new HCI modalities are used to enhance ATCo's working experience, SA, level of performance, immersion and sense of presence. This is done with the global aim to enhance the general level of safety. The present study is inscribed in this topic and considers this specific concept, which will be discussed in more detail in the next chapters.

⁶Of course, controlling at least 2 airports at the same time can significantly change the way ATCos work. In fact, much of their training would need to be reconsidered, and new interaction and feedback techniques should be designed to "absorb" the workload generated by such a paradigm.

⁷See https://www.remote-tower.eu/wp/?page_id=9 for more details.



Figure 1.3 — An example of Multiple Remote Tower validation test held under SESAR 2020's PJ05 project *Remote Tower for Multiple Airports* in April 2018. With this exercise, 3 airports are simultaneously experimentally controlled by one ATCo.

1.4 Existing RCT environment, prototypes and solutions

In this section we provide an overview of the first RCT solutions currently in operation, offered for sale, or under conception. For a few years now, RCT concepts have been democratized all over the world (Europe, Australia, United States, India). We provide in this section an introduction to the first and well-known RCT and RTC implementations. Nowadays, the concept is growing increasingly more faster; to get a glimpse of this trend, we count nowadays 12 distinct projects related to remote ATC just in France.

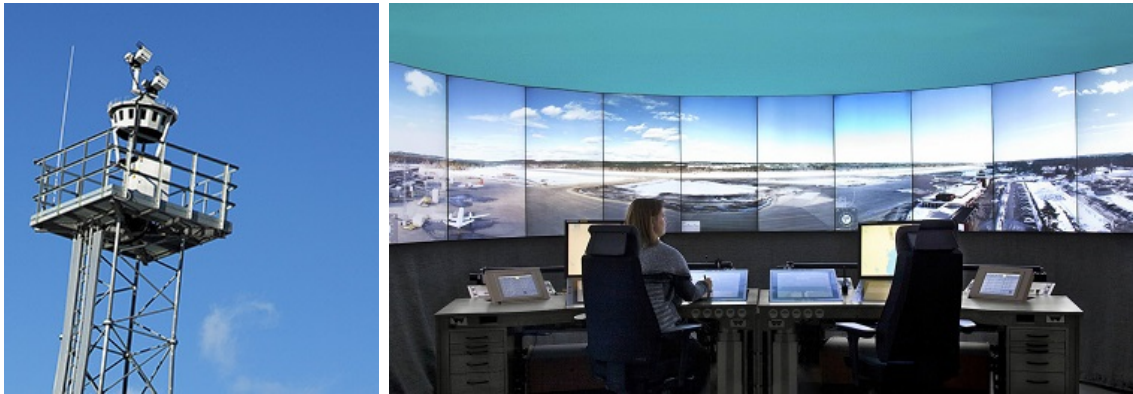
1.4.1 Sweden

The world's first Remote Control solution was built at the Örnköldsvik airport in Sweden (see Figure 1.4). The provided RTS, called *Saab Digital Air Traffic Solutions*, was the result from a joint venture between the Swedish Saab manufacturer and Swedish ANSP (Luftfartsverket). The remote airport was enabled using a set of cameras and sensors installed on site, providing real-time data to the RTC whose facilities are located approximately 100 km away in Sundsvall.⁸ Currently, Sundsvall RTC is remotely managing the operations at Sundsvall for Örnköldsvik and Härnös airports. The Sundsvall Remote Tower Center targets RCT implementation in small airports having physical tower unmanned or only momentarily staffed during daytime.

1.4.2 Norway

Avinor, the Norwegian ANSP in charge of RCT implementations in the country, tested a RTC implementation at Bodø to remotely manage operations at an heliport in Værøy airport (see Figure 1.5). This heliport represents an area with low-density traffic, as it is serving only 4 scheduled

⁸The solutions proposed on the market can be located even further away from the controlled airport.



(a) RCT equipment installed at Örnköldsvik airport. (b) RCT installation at Sundsvall to remotely control Örnköldsvik and Härnös airports.

Figure 1.4 — Real RTS solutions provided by Saab manufacturer and Swedish ANSP in Sweden.

flights a day. Bodø RTC has also the mandate to remotely manage Roest airport. The ATCos have control over all the sensors and airfield lighting systems, such as the infrared cameras. Since this project, fifteen RCT solutions have been implemented in Norwegian airports under Avinor's control, in one unique RTC. The majority of these airports are implemented in remote areas with severe meteorological conditions.



Figure 1.5 — The RCT facility provided by Avinor in Norway and implemented in Bodø to remotely control Vaerøy airport.

1.4.3 Australia

Air services Australia (Australian ANSP) is as well interested in the benefit of the RCT concept as it concerns the sixth largest country in the world in terms of surface (approximately 7 million of km²). Therefore, a RCT has been installed at Alice Spring, which is located in central Australia.

The corresponding RTC facilities (Figure 1.6) have been deployed in Adelaide, located in south Australia. One of the main drivers for this RCT implementation is the remote location of airports with particular harsh meteorological conditions in the center of the country, which have made it understandably difficult to motivate ATCos to work in such remote areas.



Figure 1.6 — The RCT facility to remotely controlled Alice Springs airport located in south Australia.

1.4.4 Ireland

Irish ANSP and Civil Aviation Authority use Saab systems, the same one which was tested and deployed in Sweden (see Section 1.4.1), to remotely control Cork and Shannon airports since 2016. These airports are controlled from Dublin's airport, in which an RTC has been built. This project was framed in the scope of the SESAR project called *Large Scale Demonstration*. The contract included the use of Electronic Flight Strips environment in the RCT implementation. The system comprises a series of high definition cameras with pan-tilt-zoom feature, weather sensors, microphones, as well as signal light guns that can be remotely operated from Dublin RTC.

1.4.5 Germany

Deutsche Flugsicherung (DFS, German ANSP) has implemented a RCT solution for Saarbrücken airport in 2017. Frequentis, an Australian company, was appointed to provide the necessary equipment, enabling Saarbrücken airport to be remotely controlled from a RTC at Leipzig (see Figure 1.7). DFS's road map included further RCT implementations in both Erfurt and Dresden airports to be consequently integrated into Saarbrücken RTC. The main goal of these implementations pursues more efficient shifts planning, as in this case DFS expected ATCos to be simultaneously rated for the different airports.



Figure 1.7 — The RCT solution used by German ANSP DFS to remotely controlled Saarbrücken airport from Leipzig.

1.4.6 Canada and United Kingdom

Canada covers a very wide airspace with many aerodromes having a very low traffic density, and for some of them, very difficult accessibility. In case of emergency, vehicles and aircraft can be remotely controlled by Remote Aerodrome Advisory Services. Nav Canada with the support of Avinor (respectively Canadian and Norwegian ANSP) implemented an RTC in London international airport. Nav Canada RCT implementation follows a different approach to the European ones called “Networked Tower Systems”. This approach targets aerodromes with few or no night operations, however fully-staffed during peak hours. Within this approach, daylight operations are still carried as classical aerodrome control, and night operations can be consolidated in a sort of RTC, formally named Flight Service Station, which only provides Aerodrome Flight Information Service (AFIS) from a remote location.

1.4.7 United States

US implementations differs from the European interpretation of the concept. One of the biggest interest for the country is to eliminate the loss of vision to the movement area when building several airport towers in medium and large American airports. However in 2015, a 3-month trial of a RTC was performed in Leesburg, Virginia. The implemented RTC used a similar system than the one deployed in Sweden. It was set up at the airport itself during the test phase. In addition, multiple high definition cameras, 2 ATCo working positions and a mobile ATC tower was used to guarantee safety. More recently, the Federal Aviation Administration has nominated the Fort Collins-Loveland Municipal airport as the official test facility for the new RCT technology. This engaged ground-based video and aircraft ground detection capabilities to relay the information to ATCos in a remote location. One of the main factor for the location decision was its close proximity to the FAA approach and Air Route Traffic Control Centers, located in Denver and Longmont, Colorado. This testing equipment have been installed in 2016, with preliminary testing, analysis and validation subsequently programmed.

1.4.8 France

The Direction des Services de la Navigation Aérienne (DSNA, French ANSP) is interested in studying the applicability and future implementations of RCT in some of the airports in the overseas, in particular those qualified as Extended-range Twin-engine Operation Performance Standards airports, as they are required to provide ATS service on a 24/7 basis. Remote control has been so far already experimented from Saint Pierre et Miquelon⁹ airport tower with 8 high definition cameras.

Figure 1.8 shows a timeline on which the main advances related to remote control concept have been positioned in time since 2009.

1.5 Sensorial augmentations in RCT solutions

Knowing that we are working on interaction techniques relying on perceptual aspects, it is important to mention the different sensory channels involved in a common ATC working environment and to describe their reproduction into RCT contexts. This will permit us to have a more detailed view on the environment in which this work is inscribed (namely the Augmented RCT).

1.5.1 Sight

Visual aspects are vital in a physical control tower. In addition to the different radars and other equipment they can use, ATCos often use their sense of vision to locate and discriminate approaching, taking-off and on ground aircraft. In a RCT environment, the view on the airport vicinity must therefore be accurately reproduced. This is made possible by the installation of cameras on the controlled area, coupled with a data link to transmit this view to the RTC. Obviously, this visual reproduction capabilities should ensure that the classical tower working environment can be strictly reproduced remotely to its further extent, at least.

The visual transmission can be described as the ability of offering an out-of-the-window real-time image of the controlled area in comparison to the one ATCos would obtain in case the tower is physically located onto the airfield. Therefore, such an environment becomes conducive to all kinds of visual enhancement via this digitization process. New features, which cannot normally been found in a physical control tower, can then be added. Hence, some remote control solutions offer zoom functionality to provide a more detailed view of a selected portion of the controlled area, or thermal infrared video capabilities for night and/or low-visibility operations. We can also mention tagging systems to add additional textual information to aircraft so that ATCos can quickly identified them (see Figure 1.9).

Existing studies and publications in HCI for RCT context bring for the most part increases partly or fully related to this sensory channel, as it is mainly used in regards to the other senses. This is for example the case in [36], where the Cordeil *et al.* consider RCT environment as a fully immersive environment which is enabling immersive analytics [50] with the use of immersive

⁹Saint Pierre et Miquelon is a French archipelago in North America located in the North Atlantic ocean, 25 km from the south Terre-Neuve Canadian island.

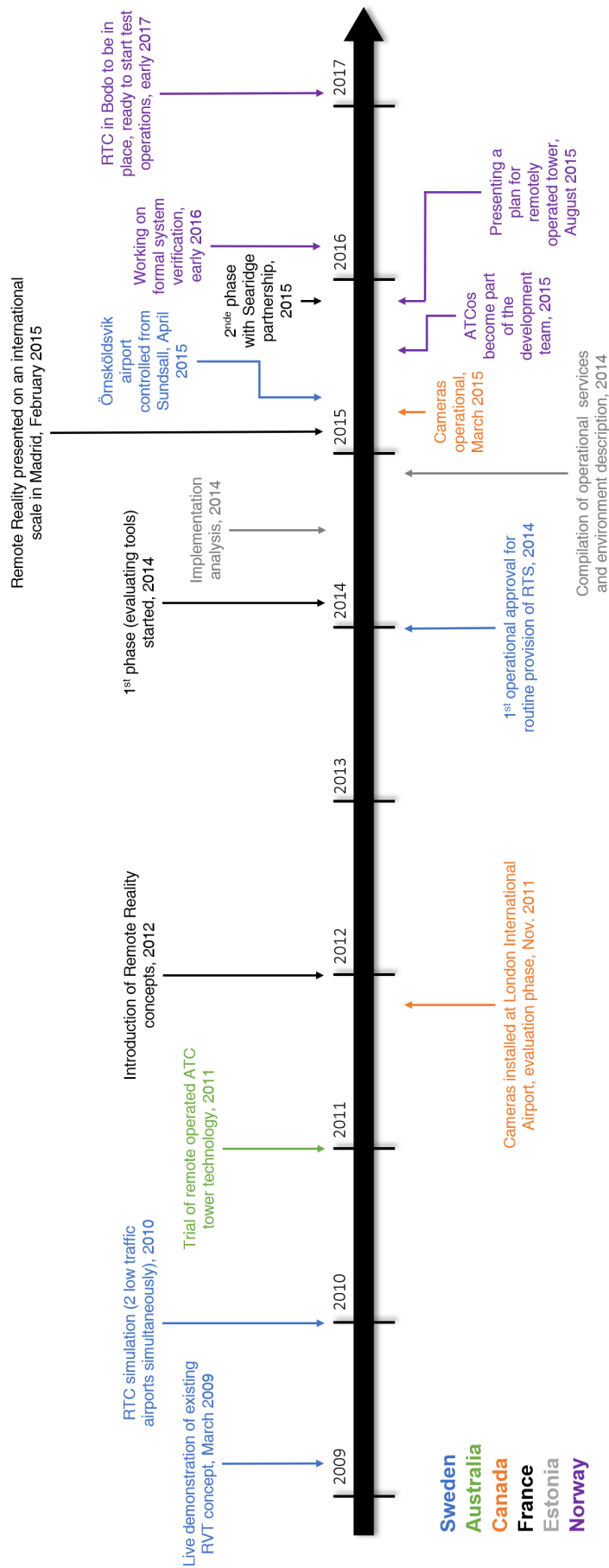


Figure 1.8 — Timeline of some of the main project related to Remote Tower concept since 2009.



(a) Infrared visual augmentation feature offered by Rheinmetall German manufacturer.



(b) Tag and zoom visual augmentation features offered by NATS RCT solution.

Figure 1.9 — (a) infrared vision, aircraft detection and tag features in RCT solution and (b) some examples of AR tag and zoom enhancements.

head-mounted display to analyze aircraft trajectories. Also, a detailed list of anticipated visual cues and features have been identified in [169].

1.5.2 Hearing

The audio environment of ATCos is quite poor. It is however sufficient for them to have a mental auditory representation of the controlled airspace. In addition to radio communications based on hearing and speech, the auditory senses can be stimulated in a physical control tower by relatively discrete events. For example, the wind may be felt on the walls of the tower, or some engines may

be heard if they are located relatively close to the tower. In a control tower, ATCos are actually quite isolated from the area they control and their audio environment is ultimately quite poor, but still contributes to the sense of immersion. This point must therefore be considered in a RCT context in order to recreate this auditory environment, how weak it may be.

Sound-based interactive solutions already exist in the context of RCT, but are quite few. To give an example, an innovative method of spatial sound using binaural stereo in order to discriminate the communications of en-route ATCos is reported in [69]. ACC ATCos work in pairs, usually sitting on separate chairs in front of the same desk. They both have separate tasks, which complement each other to complete the en-route control task. This system allows different audio channels, each intended for one of the two ATCos, to be broadcast simultaneously and without affecting the audio environment of the other ATCo. 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.

1.5.3 Touch

As mentioned above, wind-driven oscillations can be perceived in a physical control tower. This represents a portion of the haptic stimuli that controllers can feel. Engines at start or acceleration can also be felt in a similar way. However, haptic senses are not much solicited in such an isolated context.

Very little work exists to study the sense of touch in control towers. One of the most concrete example of research work considering touch sensory channel in ATC is [101, 85], where the authors introduce Strip'TIC, a system that aims at replacing actual tangible paper strips used by ATCos to track the state of controlled aircraft. The concept of Strip'TIC concerns haptics in the sense that it is aiming at replacing the sense of touch, i.e., the fact to grab a physical object (e.g., paper strips) using an HCI concept. The research is based on the observation that using strip papers the system cannot be easily updated with real-time aircraft-related information. And this could be problematic while the working environment is more and more digitalized. ATCos are attached to this tangible aspect of their working task and are usually reluctant to abandon it. This is why Hurter *et al.* proposes a system combining vision-based tracking, augmented rear and front projection, radar visualization, and electronic paper strips which is made usable, graspable with a digital pen. They wanted to propose a system in which the tangible/touch aspect could be conserved, at least in part. This type of system is now commonly used, mainly by ACC ATCos.

Wrap-up message

In this chapter we presented the fields of Air Traffic Management (ATM) and Air Traffic Control (ATC). After a brief historical overview in which we saw how these sectors were born, we have presented in more detail ATC and its structure. We saw that ATC is divided into 3 main types, namely *en-route* control, *approach* control, and *ground* control. These different types of ATC are all concerned with the same rules. This thesis work is part of the approach and ground control types, and as we will see later, in the specific situation of almost no VFR conditions (i.e., degraded or no visibility).

We introduced the concept of remote control, which is basically the fact of controlling a distant airspace, most often small aerodromes located in sparsely populated areas and/or subject to harsh weather conditions, from a remote control position. However, this concept is increasingly being applied to larger airports due to budgetary constraints and ease of human resource management. To qualify in more details remote control, a set of terms have been developed. All equipment used to capture not only the visual, but also the sound or any other information on the controlled area is called Remote Control Tower (RCT). This data is then sent to a Remote Tower Center, and more precisely in a room designated by Remote and Virtual Tower. The present works take place in this type of installation.

There are several types of RCT. Hybrid RCT is the simplest one and consists in the minimum set of equipment necessary for remote control. Single RCT is when one ATCo is controlling a single airport at a time. Multiple RCT is when one ATCo is controlling several airports simultaneously. Augmented RCT represents one of the 3 previous concepts to which HCI augmentations have been added in the form of interactions or feedback. Nowadays, the concept of remote control is more and more applied to real operational situations. The first countries to have studied it are Sweden, Norway, Australia, Ireland, Germany, Canada, England, the United States, Estonia and France, and the concept is now democratizing at high speed. France, for example, has currently a dozen projects underway related to the concept of remote control.

We saw that the sensory channels of ATCos are not used very much. Vision is predominant in their working task; naturally, it allows them to search for the aircraft with which they are in communication in the airport vicinity, or for example to read radar screens. However, hearing and touch are not really considered. In a digitized environment such as that of a RCT, this represents a study track with the goal to unload the visual channel. Situational Awareness (SA) could also be enhanced in RCT environment, compared to a physical control tower, where it is higher. Consideration of other sensory channels would improve ATCos' sense of presence. Hence, taking advantage of sensory channels in which there is still bandwidth left may promote natural perception through embodied cognition and increase ATCos' performance. We will discuss these points later in this manuscript. The next 3 chapters, making the first part of this manuscript, focus on a review of the HCI literature related to hearing and touch. A presentation of the mechanisms of human perception is available in the next chapter.

PART I

THEORETICAL BACKGROUNDS

Chapter 2

Introduction to non-visual perception

Résumé (Français)

Introduction à la perception non visuelle (État de l'art, partie 1)

Introduction

L'auteur Kantz a un jour proposé une métaphore des sens humains convenant relativement bien à l'informatique; en voici une traduction possible : “[...] *les sens peuvent être vus comme des transducteurs capturant l'information du monde physique, réel, pour en créer une traduction adaptée aux codes de notre monde intérieur, créant ainsi notre perception du monde qui nous entoure.*” Le terme technologique *transducteurs* est ici employé dans le cadre d'une définition biologique, ce qui correspond bien à la composante *augmentation* du présent travail. Dans ce chapitre, nous présentons la façon dont notre corps capte l'information du monde réel sous la forme de *sensations* pour ensuite créer nos propres *perceptions*. Nous nous attardons principalement sur les systèmes auditifs et somatosensoriels, responsables respectivement de notre audition, ainsi que de notre toucher et notre proprioception. Dans la suite de ce manuscrit, nous nous servirons des caractéristiques intrinsèques de ces systèmes sensoriels pour concevoir des modalités d'interaction non-visuelles adaptées à certaines situations d'intérêt pour le contrôle aérien d'approche et d'aéroport.

Une *sensation* représente l'information qui est directement capturée par les récepteurs sensoriels. La *perception* représente la même information après traitement par le cerveau. Les sensations, et donc ce que nous appelons les sens humains, sont créées par différentes structures de notre corps appelées systèmes sensoriels. Il n'existe pas de définition uniforme et de classification stricte de ces sens. Ils peuvent cependant être divisés en trois catégories [93] :

- l'*extéroception* [109], qui est la perception de stimuli provenant de l'extérieur du corps; elle correspond aux sens communs (vue, audition, toucher, odeur et goût) ;
- l'*intéroception* [38, 93], correspondant désormais à une classification bien admise englobant les sens dont la fonction est d'acquérir l'information provenant des organes internes;
- la *proprioception* [167], qui correspond à la perception de notre propre corps dans l'espace ainsi que de ses mouvements.

Le système auditif et la perception du son

Anatomiquement, le système sensoriel lié à l'audition humaine, également appelé *système auditif* est composé de deux parties : l'oreille, qui est responsable de la transformation des grandeurs physiques en signaux biologiquement, et le système nerveux, qui est responsable de l'acheminement jusqu'au cerveau et de l'interprétation de ces signaux physiques. L'oreille elle-même est composée de trois groupes anatomiques : l'*oreille externe*, composée de la partie visible de l'oreille et du tube externe, l'*oreille moyenne*, essentiellement composée du tympan et des os transmettant les vibrations à l'oreille interne, et, donc, l'*oreille interne*, faisant également partie du *système vestibulaire* responsable de l'audition mais aussi de l'équilibre, et contenant les labyrinthes osseux et membraneux (voir Figure 2.1 ci-après). À l'instar d'autres systèmes, le système auditif a évolué pour nous doter d'"armes" utiles à la survie de l'espèce. Notre capacité à filtrer, sélectionner les sons est issue de ce processus. Ses caractéristiques doivent être prises en compte dans la conception d'interfaces. De manière générale, nous sommes en mesure de percevoir des signaux sonores au travers de cinq paramètres :

- le *pitch*, qui d'après [49] représente la fréquence sonore *perçue*. Fondamentalement, la fréquence d'un son est le nombre de périodes en une seconde (la *hauteur* d'un son est la perception de cette vibration, qui est transmise au cerveau lorsque les osselets de l'oreille moyenne sont en mouvement) ;
- le *timbre* correspond à la différence existant entre deux sons perçus différemment mais ayant cependant la même hauteur ainsi que la même intensité sonore ;
- le *niveau sonore*, mesuré en décibels (dB), relatif à la *puissance sonore* ;
- les dimensions *spatiales*, faisant référence à toutes les caractéristiques du son qui nous permettent de percevoir sa position absolue ou relative dans l'espace ;
- et la dimension *temporelle*, i.e., le temps, intrinsèquement lié au concept de son.

Dans sa gamme de fréquences, grossièrement comprise entre 20 Hz à 20 kHz, l'oreille humaine est capable de détecter une grande variété d'intensités sonores. Dans la plupart des environnements rencontrés, cette détection est liée au nombre de fréquences différentes perçues. Le plus petit son audible est fonction de l'intensité globale. Cette sensation peut être décrite à l'aide de la loi de Weber-Fechner [44], qui est une fonction logarithmique décrivant la relation entre une sensation mentale et la magnitude physique d'un stimulus [44] : $20 \log_{10}(P/P_{ref})$, exprimé en dB, où P est la pression mesurée et P_{ref} est la pression de référence. Les sons à fort niveau sonore, peu communs ou non pertinents dans le contexte, sont détectés presque instantanément car ils peuvent représenter un danger ou, plus généralement, être liés à une information urgente. Ces caractéristiques concernent également l'aspect temporel du son : les cellules nerveuses de la cochlée (oreille interne) ont en effet leur propre fréquence vibratoire maximale définissant leur propre bande passante. Des variations temporelles soudaines à l'intérieur de cette bande passante peuvent être très bien détectées. Deux aspects temporels contribuent à cette détection : l'enveloppe sonore, qui varie lentement, ainsi que la structure du son lui-même, qui varie rapidement. Notre sensibilité à

la détection de changements de cette fréquence semble avoir son pic autour de 55 Hz, mais elle diminue rapidement jusqu'à 1000 Hz où aucune variation n'est détectée. De plus, le système auditif est capable de détecter au moins 2 à 3 ms de variations dans la structure sonore [131]. L'oreille humaine peut également percevoir une différence de 1 Hz à 100 Hz, mais ne peut pas entendre une différence de fréquence inférieure à 2 ou 3 Hz à 1 kHz. En outre, nous avons des difficultés à reconnaître une différence de 1 octave entre deux sons au-delà de 5 kHz. Et par ailleurs, l'oreille humaine peut créer une fréquence fondamentale perçue supplémentaire lorsque les sons sont joués simultanément et sont harmonieusement bien séparés.

Pour la présente étude, une caractéristique intéressante de l'oreille humaine est l'*effet cocktail* (plus connu sous sa traduction anglaise *cocktail party effect*). Celui-ci peut être défini comme la faculté de se concentrer sur un orateur particulier parmi plusieurs autres, constituant un bruit de fond. Cet effet a d'abord été identifié par S. Handel [74, 7]. Il a depuis lors largement été étudié. L'oreille humaine est capable de faire ce processus de sélection, alors qu'il est presque impossible de le faire en analysant visuellement le spectrogramme correspondant. Par ailleurs, notre capacité à percevoir les sons dans l'espace nous est essentielle. Cette faculté contribue à construire une représentation 3D précise de notre environnement. Contrairement à la vision, l'audition nous permet d'intégrer les 360 degrés d'éléments environnants à notre représentation interne de l'espace, sans champs morts. Comme les auteurs de [98] l'ont remarqué, contrairement à la vision, l'audition n'a pas besoin d'être orientée dans une direction particulière. Cette faculté contribue également à notre perception des distances, à attirer notre attention sur une zone d'intérêt, y compris un danger, ou à intégrer des informations qui ne sont pas perçues par d'autres canaux sensoriels (par exemple des pas à travers des murs). Ces caractéristiques contribuent à affiner la perception générale de notre environnement.

D'un point de vue IHM, la représentation sonore de l'espace a récemment été identifiée comme un point clé contribuant au *sens de présence* et plus généralement à l'*immersion*. Cette faculté permet également de localiser les alertes ou tout événement méritant l'attention de l'utilisateur, l'IHM (sonore) concernée étant spatialement structurée. Cette perception auditive de l'espace est permise grâce à la géométrie de l'oreille externe et à l'espacement entre les deux oreilles, qui sont de plus isolées acoustiquement l'une de l'autre par la tête. Lorsqu'un son est émis par une source, la distance entre la source et les deux oreilles est perçue par l'intermédiaire des distances entre la sources et chacune des deux oreilles, qui ne sont pas égales (si la source sonore en questions n'est pas située pile en face de l'écouteur). Ce principe affecte également la hauteur ou les fréquences perçues. De même, l'amplitude perçue ne sera pas la même pour les deux oreilles. Par ailleurs, la différence de phase entre les deux oreilles est également incluse dans le calcul effectué dans le cortex auditif pour interpréter l'emplacement d'un son. Toutes ces données donnent au système auditif des repères binauraux qui sont nécessaires pour déterminer la distance et la direction d'une source sonore [114]. Cependant, deux erreurs sont possibles dans la perception de l'emplacement des sources sonores: les erreurs de type *cône de confusion*, qui se produisent lorsque la source sonore est localisée dans un cône allant de l'oreille jusqu'à l'emplacement de la source sonore. Un second type d'erreur se produit lorsque la source sonore est localisée à proximité de l'emplacement réel de la source sonore avec une précision de $\pm 6 \sim 7$ degrés [31].

Toucher et système somatosensoriel

Le système somatosensoriel est un sous-système du système nerveux et le principal système sensoriel de notre corps. Il est composé des organes et des processus qui contribuent à nos sensations (pression, chaleur, douleur, équilibre). Ces sensations sont créées par les récepteurs sensoriels à partir d'informations provenant de diverses parties du corps (peau, os, tendons, muscles, organes, oreille interne). Les récepteurs sensoriels sont des structures biologiques qui réagissent aux changements de l'environnement.^[1] Il en existe trois types : les *mécanorécepteurs* permettant de ressentir des forces mécaniques (pressions légères ou intenses, mouvement des cheveux, toucher, vibration de basse ou haute fréquence, étirement, étirement des muscles, étirement des tendons), les *thermorécepteurs* permettant de ressentir les différences de températures, et *nocicepteurs*, permettant de percevoir la douleur. Ces récepteurs sensoriels se retrouvent dans tout le corps. Puisque notre peau est le plus grand organe de notre corps (environ 2 m² en moyenne [117]), elle représente une grande surface d'interaction potentielle. Les caractéristiques du système somatosensoriel sont particulières dans la mesure où elles permettent de percevoir des vibrations jusqu'à 10 kHz, ce qui, comparé à la vision par exemple qui a un taux de rafraîchissement d'environ 25 Hz, est assez significatif. Cependant, nous ne sommes en mesure de percevoir des variations de pression que dans une bande passante de 30 à 300 Hz. Nous percevons des mouvements inattendus à une fréquence d'environ 1 Hz, et nos réflexes sont perçus jusqu'à 10 Hz. La plus petite variation de température ressentie est d'environ 0,01 °C.

Il existe trois types de perception liées au toucher :

- La perception *cutanée* ou *passive* correspond à une stimulation cutanée appliquée à la surface de la peau alors que la partie concernée du corps est immobile (par exemple, la perception d'un insecte se posant sur notre main ou du vent sur notre visage). Ce type de perception a été pris en compte dans la conception des modalités d'interaction présentées dans la partie II (voir Chapitre 5, p. 93).
- La perception *tactilo-kinesthésique* ou *active*, par opposition à la précédente, correspond à une perception passive impliquée dans un processus d'exploration volontaire, avec la main ou toute autre partie du corps, dans le but de sentir une surface, une texture. Le fait de toucher un objet pour en appréhender la forme, comme le ferait un ébéniste, correspond à la perception active. La perception active seule n'existe pas vraiment en soi. Elle est dans la plupart des cas couplée à d'autres sensations.
- La perception *haptique* est engendrée lorsque la perception active précédente est couplée à la perception de la gravité et à la proprioception. Ceci résulte d'un processus complexe d'intégration de sensations permettant d'appréhender inconsciemment le poids, la forme, la densité d'un objet. La réflexion sur ce type de perception haptique a été affinée en 1966 par James J. Gibson [65] et confirmée par Carl E. Sherrick et Ronald Rogers la même année [155]. Gibson a défini la perception haptique d'un individu comme la perception du monde l'environnant par l'utilisation de son propre corps. Il a en outre été établi que

¹Ils peuvent être apparentés aux transducteurs de la métaphore de Krantz introduite en début de chapitre.

ce processus est automatique et inconscient, et qu'il prend en compte non seulement les forces mécaniques volontairement appliquées au corps et la gravité, mais aussi les forces mécaniques internes involontaires telles que la contraction des muscles.

La proprioception

La proprioception fait référence à la perception des mouvements de notre corps, associée à la perception de sa position, ou de la position de ses parties, dans l'espace. La proprioception *est* une perception, c'est-à-dire le résultat de l'intégration et de l'interprétation de sensations provenant de récepteurs sensoriels multiples répartis dans notre corps. La plupart de ces informations proviennent des récepteurs sensoriels des muscles et des tendons, de l'oreille interne, et des systèmes nerveux impliqués. Ce processus nous permet de percevoir, inconsciemment, nos propres mouvements et l'orientation spatiale de notre corps. Cette perception nous permet également de percevoir en détail la position de notre corps et de ses parties par rapport aux éléments de notre espace environnant. On parle aussi parfois de *sensibilité profonde*. Il n'y a pas de différence évidente avec le terme *kinesthésie*. Cependant, on parle de proprioception lorsque l'on souhaite désigner cette faculté de notre corps, et plutôt de kinesthésie lorsqu'il s'agit de conception d'interfaces. Nous développons plus en détails ce concept IHM dans le Chapitre 4 (p. 71).

2.1 The human senses

Krantz gave a metaphor for human senses that suit well for Computer Science [99]:

“[...] senses are transducers from the physical world to the realm of the mind where we interpret the information, creating our perception of the world around us.”

In this Ph.D. work, we rely on human perception (especially hearing, touch and proprioception) to build novel interaction modalities. There is a fundamental difference between *perception* and *sensation*: *sensation* correspond to the consciousness of variations in the internal environment and the environment, as our body physiologically retrieve these variations, while the *perception* is the conscious interpretation of these stimuli. A *sensation* represents the information which is directly captured by the sensory receptors, and the *perception* represents the same information after treatment by the brain. Sensations, and therefore what we called *human senses*, are provided by different structures in our body called *sensory systems*. There is no uniform definition and strict classification for the human senses. However, we can distinguish them into 3 main categories [93]:

- *Exteroception* [109], which is the perception of external stimuli originating outside of the body, e.g., common senses (sight, hearing, touch, smell and taste);
- *Interoception* [38], which is now a well-admitted classification containing the senses whose function is to retrieve information from internal organs. According to [93], it's *“the perception of the body's internal state through the processing of signals arising from within the body, e.g., blood pressure, heart beats, etc. Interoceptive features may reflect the emotional valence, arousal and other somatic states”*;
- *Proprioception* [167], which is the perception of body's own position, motion and outside state.

These three human senses categories does not correspond to the sensory systems, which could be part of more than one of these categories, such as the somatosensory system, for example. Traditional exteroceptive senses are those corresponding to the sensory systems relying on sight (visual system), hearing (auditory system), touch, taste, and smell (somatosensory system). Exteroception also includes other human sensations, such as *nociception*, which is the faculty to feel the pain (somatosensory system), *thermoception*, which consists in the perception of differences in temperature (also somatosensory system), and the possible *magnetoception*, which may be the perception of the direction (this one persists to be a strong assumption). Other important human senses are related to proprioception, which is the awareness of the body's position and its different parts in space. Proprioception is also about *equilibrioception*, which is the sense of balance.

In this Ph.D. dissertation, we concentrate on exteroceptive senses of hearing, via sonification methods and interactive sound, and the exteroceptive and proprioceptive haptic senses, via tactile cutaneous sensations and kinesthetic perception. In this perspective, the next sections propose a overview of the human auditory and somatosensory systems.

2.2 An overview on the auditory system

In this section we describe the human auditory system from an anatomical (Section 2.2.1) and psychoacoustic point of view (Section 2.2.2). In this last section, we address important themes for the next chapters, namely the perception of sound through several physical dimensions (Section 2.2.2.1), the characteristics and limitations of the human ear (Section 2.2.2.2), the cocktail party effect (Section 2.2.2.3), and spatial hearing (Section 2.2.2.4).

2.2.1 The auditory system

The sensory channel related to human audition is composed of two main parts. The first part is the one responsible for the transformation of physical quantities (acoustic energy) into signals that are biologically understandable. This work is done by the ear. The second part is the one responsible for the interpretation of this physical signal, namely the nervous system. These two parts are making the auditory system.

The ear is composed of three main anatomic groups: *outer*, *middle* and *inner ears* (Figure 2.1). The outer ear is composed of the visible part of the ear and the external tube. The middle ear makes the junction between outer and inner ears, and is essentially composed with the eardrum and the bones attached to it, which transmit the vibrations to the inner ear. At the end, the inner ear is part of the vestibular system, which is responsible for hearing and balance. It contains the sensory organs related to the vestibular system: the bony and membranous labyrinths. This system (i.e., the ear and the associated nervous system sub-parts) is no exception and has evolved in such a way as to basically give us weapons to survive. For example, the capacity to filter sounds of interest is one of them. Human ear has endowed intrinsic characteristics throughout the ages, which naturally have to be considered when designing HMI.

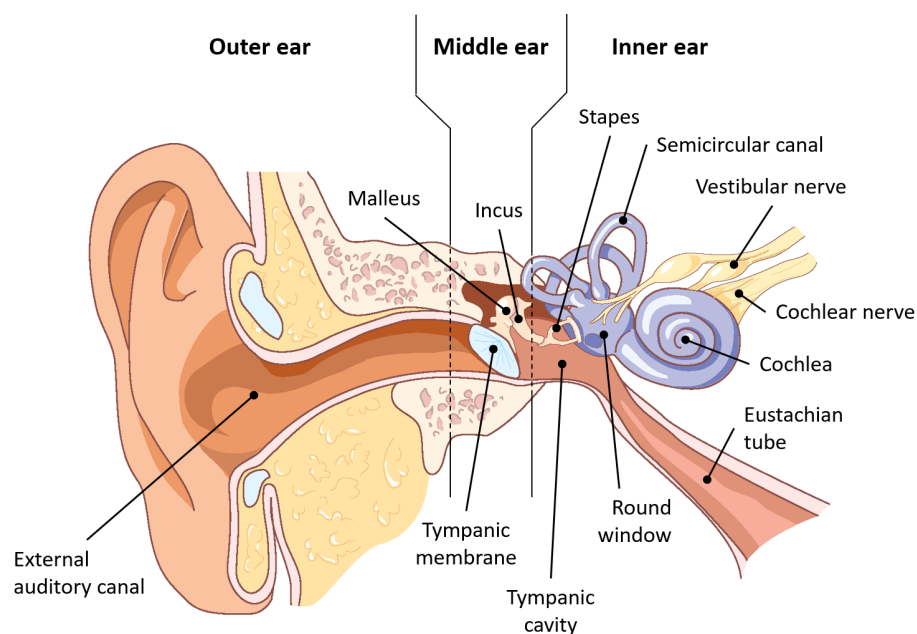


Figure 2.1 — The human ear, its 3 main groups of structures and their different sub-parts (this scheme was taken from [15] and then modified).

These physical characteristics are imbued by these 3 different auditory stages into the acoustic signal before making a biological signal. Then, it is sent via the vestibulocochlear nerve (or auditory vestibular nerve, one of the cranial nerves) to the brain, where it will be interpreted. However, the first steps of this interpretation process begins in the outer ear: its characteristics, that are impregnated by the ear into the signal, can be seen somehow as a post-treatment. Actually, the sound transmission process begins in the aperture of the pinna of the outer ear where sounds are collected and then conveyed to the middle ear through the external auditory canal. Here in the pinna, the sound is first filtered [153] by a complex process of signal reflections thanks to the asymmetrical structure of the tissues. This filtering is partly responsible for our ability to localize sounds [31]. As explained in [78], the pinna is also responsible for the perception of sounds “outside the head”, which is the sensation that the sounds we hear are actually not localized in our ear, but really *outside*. Once in the middle ear, the signal will first make the eardrum vibrate, which in turn will transmit these vibrations to the chain of ossicles. These ossicles are 3 in number: the hammer (or *malleus*), the anvil (or *incus*) and the stirrup (or *stapes*). These are the smallest bones of the human body. Their function is to convert the sound vibrations into mechanical movements. These movements are then transmitted to the inner ear. In this last stage, the stapes transmit the movements to the semi-circular canal and cochlea. They contain numbers of little hairs, each of them being connected to a distinct nerve cell. Here, the auditory vibrations, previously converted into a mechanical movement, are finally converted into nerve impulses and carried out to the brain (auditory cortex). The inner ear is not bigger than a pea, but it realizes a complex signal processing, by realizing high definition frequency and temporal filtering.

2.2.2 Psychoacoustics and perception of sound

2.2.2.1 Perceived sound dimensions

Human ear perceives sound signals through 5 main dimensions: the *pitch* dimension, the *timbre* dimension (e.g., timbre, spectral power, amplitude and frequency of harmonics, roughness, brightness, etc), the *loudness* dimension (e.g., the general intensity of a sound signal), the *spatial* dimension (sound sources disposition in space, stereo panning, interaural differences, etc), and the *temporal* dimension (tempo, duration, sequential position, decay, etc). These dimensions have been classified in a systematic review of the literature for the sonification of physical quantities in [49]. Sound frequency, or *pitch*, which is commonly associated to the perception of the frequency when we talk about audio (and not in other fields), is related to the vibratory dimension of sound. Basically the frequency of a sound is the number of periods in one second. Talking about auditory perception, the pitch is the perception of this vibration, which is transmitted to the brain when the ossicles are moving. The *timbre* is commonly defined as the characteristics that makes two sounds having the same pitch and loudness different from our perception. The *loudness* can be defined as the perceived sound volume. *Spatial* dimension refers to every characteristics of sound that let us perceiving any absolute or relative position in a space. Finally, speaking of sound, *time* dimension is kind of trivial because sound is intrinsically related to this physical dimension: it refers to all temporal aspects of sound.

2.2.2.2 Human ear characteristics and limitations

In the next steps of this manuscript, we will explain how we build interactive modalities based on spatial hearing. It is therefore necessary to provide some of the main characteristics of sound perception.

Within its frequency range of about 20 Hz to 20 kHz, the human ear is capable of detecting a wide variety of sound intensities. For most contexts and environments encountered, this detection is linked to the context complexity in terms of number of different frequencies. That means that the smallest audible sound is a function of the overall intensity. This sensation can be described using by the Weber-Fechner's law, which is a logarithm function [44] describing the relationship between a mental sensation and the physical magnitude of a stimulus:

$$20 \log_{10} \left(\frac{P}{P_{ref}} \right),$$

where P is the measured pressure and P_{ref} the reference pressure. This number is called Sound Pressure Level (SPL) and is expressed in decibels (dB). The minimum perceptible pressure is about 20 μ P and the auditory system is most sensitive to frequencies around 4 kHz.

From an evolutionary point of view, human ear has over time been endowed with particular characteristics that are useful for surviving. For example, sounds with high loudness, that are uncommon or not relevant to the current context are almost instantly detected because they can carry a danger or more generally be related to urgent information. These evolution-related characteristics also relate to the temporal aspect of sound. Nerve cells of the inner ear in the cochlea have their own maximum vibratory frequency which defines their perceptual bandwidth. However, sudden temporal variations within this bandwidth are very well detected. Two temporal aspects contribute to this detection: the sound envelope, which varies slowly, and the structure of the sound itself, which varies rapidly. The perception of the modulation of this envelope varies according to the rate of its temporal change. This effect is called Temporal Modulation Transfer Function (TMTF). The sensitivity to detect changes in this frequency seems to have its pic at around 55 Hz, but this rapidly decreases until 1000 Hz where no variation are detected. Also, the auditory system is able to detect at least 2 to 3 ms variation gaps in the sound structure [131].

Specific perceptive properties can be attributed to human ear, including the Frequency Difference Limen (FDL) [118], which is the fact that few differences in frequency are more perceptible for low frequencies than for high frequencies. Actually, human ear can perceive a 1 Hz difference at 100 Hz, but cannot ear a frequency difference of less than 2 to 3 Hz at 1 kHz. Besides, we have difficulties to recognize a 1 octave difference between two sounds that are beyond 5 kHz. Moreover, human ear can “creates” an added perceived fundamental frequency when sounds are played simultaneously and harmonically well separated. This last effect is called *pitch residue*.

2.2.2.3 The cocktail party effect

For the present study, one interesting characteristics of the human ear is the cocktail party effect. It can be defined as the faculty to focus on one particular speaker among several others constituting a background of noise. This effect was first identified by S. Handel [74, 7] and since vastly

studied by psychologists and audio perception related specialists. It refers to *selective attention*. The human ear is capable of doing this selection process, while it is almost impossible by visually analysing the related spectrogram. In cases where this could be relevant, this particular characteristics should be considered in sonification to make sounds more salient.

2.2.2.4 Spatial hearing

Still from an evolutionary point of view: the ability to perceive sounds in space is essential to human perception. This faculty contributes to construct a precise 3D representation of our environment. Unlike vision, hearing allows us to integrate 360 degrees of surrounding elements to our internal representation of space, with no dead fields, as this is the case for vision. As the authors of [98] have noticed, contrary to the vision, hearing doesn't need to be oriented in a particular direction. In addition to this perception of the scene in which we evolve, this faculty helps us to evaluate distances, to attract our attention to an area of interest, including a danger, or to integrate information that is not perceived by other sensory channels like vision (e.g., footsteps through walls, distant sounds). Thus, these characteristics contribute to refine the general perception of our environment by complementing it.

This encoding of space resulting from the auditory system is particular in that it's radically different from the one coming from the visual system. This latter is of *spatiotopic* type, because the images from the environment world are directly encoded in the retina in the form of 2D images: it is typically a visual projection. Conversely, the space encoding coming from the auditory system is of *tonotopic* type, because it is built according to perceived frequencies and cannot be directly represented in a 2D or 3D space: this is a model of the environment. Our abilities to localized sound sources in space rely on spatial (i.e., binaural) and frequency cues [7].

From an HMI point of view, sound representation of space has recently been identified as a key point contributing to the *sense of presence*, which improves the impression of being part of the scene and thus the immersion into XR environments (see Section 2.3). It also makes it possible to locate alerts or any events deserving user's attention, since the concerned HMI is spatially structured. This representation is generally based on 3 axes, namely *direction* and *distance* of the source, and the *spaciousness* of the environment. Spaciousness is perceived thanks to the outer ear geometry and the spacing between the two ears, which are acoustically isolated from each other by the head. When a sound is emitted by a source, the distance between the source and the two ears will be heard through the delays in the time of arrival which will not be the same for both ears. This difference is called Interaural Time Difference (ITD); this principle also impacts the pitch or perceived frequencies (Interaural Frequency Difference, IFD). In the same way, the perceived amplitude will not be the same for the two ears, which is the Interaural Amplitude Difference (IAD), or Interaural Level Difference (ILD, see Figure 2.2). All these parameters are considered in the elaboration of numerical models of the ear called Head-Related Transfer Functions (HRTF), which provide an ecological spatial sound. Besides, the phase difference between the two ears is also included in the computation made in the auditory cortex to interpret the sound location. All these data give the auditory system binaural cues which are requisite for determining the distance and the direction of a sound source [114]. Two possible errors in the location of sound sources

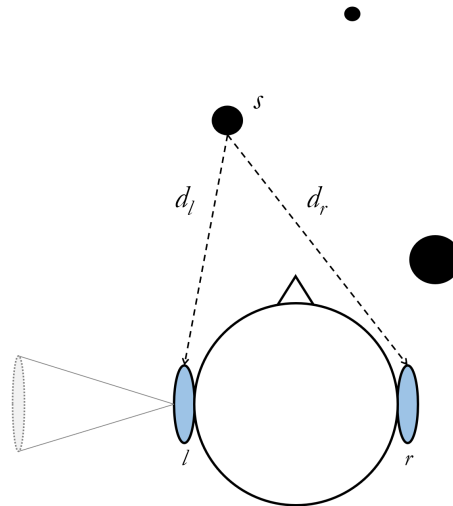


Figure 2.2 — A representation of how spatial sound is perceived by the human ear. ITD, and consequently IFD, is represented via d_i distances: here $d_l < d_r$ which means that the left ear will hear the sound sooner than the right ear (in the order of milliseconds). ILD (or IAD) is represented via the size (or diameter) s of a sound source: here, 3 sources are represented; the closer (resp. farthest) a source is from an ear, the higher its intensity is (resp. lower) (this should be applied to the two ears). Finally, a *cone of confusion* effect is represented in grey on the left ear: the dotted ellipse represent the area of location possibilities of a sound source.

have been identified. Errors of type *cone of confusion* are made when the sound source is localized in a cone starting from the ear and ending on the location of the sound source, but not exactly on the real sound location. The second type of location error is made when the sound source is localized in the vicinity of the real sound source location with a precision of $\pm 6 \sim 7$ degrees [31].

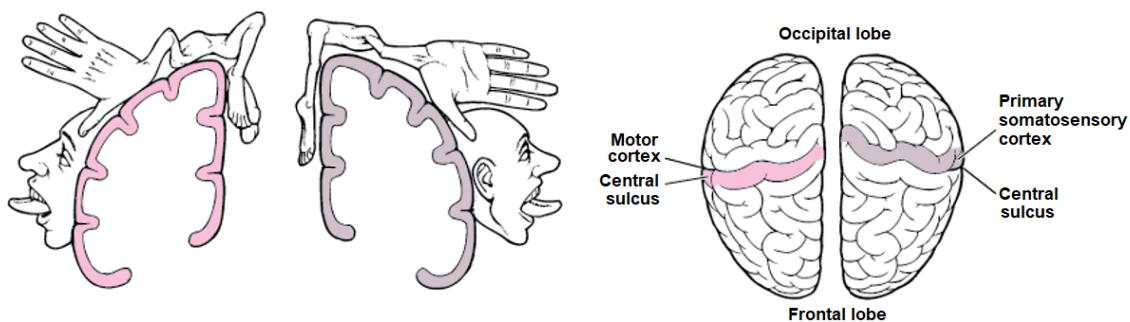
The perception of sound source direction can also be viewed as a discrimination of the space. Sound can be located along the horizontal axis (azimuth), with a particular height and a specific distance. Human ear is pretty good at locating sounds along the horizontal axis, with a Minimum Audible Angle (MAA) of about $1 - 2$ degrees [115]. With this measure, we have an approximation of the minimum localization precision of the human ear. The distance, meanwhile, is apprehended according to the quantification of 4 different dimensions [186]: intensity of the sound, air attenuation (air density), reverberation ratio, and the nature of the enclosed space (if it is the case).

2.3 The human somatosensory system and the sense of touch

In this section we propose an overview of the human sensory system related to the sense of touch and kinesthetic perception. Firstly, the somatosensory system is presented. Then, we discuss the different types of perceptions associated with the sense of touch. Distinct types of perception are involved in HMI. More particularly the so-called passive perception of the skin, which is used since tactile feedback is involved in the communication with the system.

2.3.1 The somatosensory system

Somatosensory system is a sub-system of the human sensory nervous system, and the main sensory system of our body. This specific sensory system is composed of the organs and processes that contribute to our sensations (pressure, heat, pain, balance). These sensations are created from information generated from various parts of the body (i.e., skin, bones, tendons, muscles, organs, inner ear) by sensory receptors. There are 3 types of sensory receptors: *mechanoreceptors*, *thermoreceptors* and *nociceptors*. Sensory receptors are structures that respond to changes in the environment, that is, stimuli. Mechanoreceptors allow to feel mechanical forces (light or intense pressures, movement of the hair, discriminating touch, vibration of low or high frequency, stretching, stretching of the muscles, stretching of the tendons). Thermoreceptors allow to feel the temperature (heat or cold). Nociceptors serve to perceive the pain. These sensory receptors are found throughout the body: in most tissues (free nerve endings), in the hair follicles (plexus of the hair root), in the glabrous skin (Meissner corpuscles), in the connective tissues of the mucous membranes (Krausse corpuscles), in the epidermis (Merkel's disks), the hypoderm and the articular capsule (corpuscles of Raffini and Pacini), in the subcutaneous tissues (Pacini corpuscles), as well as in the muscles (neuromuscular spindles) and tendons (neurotendinous spindles). All these receptors are involved in the elaboration of tactile, thermal and nociceptive perception. Since our skin is the largest organ of our body (about 2 m² on average [117]), it represents a large surface of potential interaction using touch sensations.



(a) The “sensorial *homonculus*”, or literally the “sensorial little man”, is a schematic view showing the relation between the different parts of the body and the related areas in the cortex that are involved in the somatosensory perception.

(b) The somatosensory system is everywhere in our body through sensory receptors. In the brain, it is located in the central sulcus in the left hemisphere of the brain, and integrates feedback information from the motor cortex, located in central sulcus of the right hemisphere.

Figure 2.3 — The sensorial “*homonculus*” and the corresponding areas in the brain. The surface of the cortex is unevenly distributed according to the sensitivity of the body parts. Thus, the fingers have a surface, so potentially a number of neurons, much larger than the feet, for example.

The somatosensory system characteristics are particular in that they allow us to perceive vibrations up to 10 kHz, which, compared to the vision for example which has a refresh rate of about 25 Hz, is quite huge. However, we are only able to experience pressure variations in a bandwidth of 30 to 300 Hz. We perceive unexpected movements at a frequency of about 1 Hz, and our reflexes

are perceived up to 10 Hz. The smallest temperature variation that can be felt is approximately 0.01 °C. Therefore, the sense of touch is a subset, or a specific “feature”, of the somatosensory system. Regarding the abovementioned description, what is interesting and usable for haptics HMI within the somatosensory system is what is perceived through the mechanoreceptors.

2.3.2 Three different types of touch perception

In this section we present the 3 different types of perception that are involved in HCI devices using tactile or vibrotactile feedback.

2.3.2.1 Cutaneous perception

We talk about *skin perception*, or *passive perception*, when a cutaneous stimulation is applied to the surface of the skin while the concerned part of the body is immobile. This is the case when a part of the body is in contact with a moving object, without having moved this part of the body. For example, an insect landing on our hand, the perception of the wind on the face or even the cutaneous perception of the rain are concrete cases of passive perception. This type of perception has been considered in the design of the interaction modalities presented in Part II (see Chapter 5, p. 93).

2.3.2.2 Tactile-kinesthetic perception

When a process of voluntary exploration is set up, with the hand or any other part of the body, with the goal to feel a surface, a texture, then we talk about *tactile-kinesthetic perception*, or *active perception*. The fact to touch an object to apprehend its shape (as a cabinetmaker would do for example) is active perception. It involves consciously and with a specific goal the sense of touch with the help of voluntary movements, resulting in the application of mechanical pressure on the skin, the muscles and more generally the concerned tissues. Active perception alone does not really exist in itself and is coupled with sensations to form haptic perception.

2.3.2.3 Haptic perception

We speak of *haptic perception* when active perception is associated with the perception of gravity and proprioception (i.e., the consciousness of our surrounding environment and our own movements). This results from an integration process of complex sensations that allows us to understand the weight, shape, density of an object that we raise without seeing it (among others and for example). This reflection on the *haptic sense* was refined in 1966 by James J. Gibson [65] and confirmed by Carl E. Sherrick and Ronald Rogers the same year [155]. Gibson defined the haptic system as “[the] sensibility of the individual to the world adjacent to his [her] body by use of his [her] body”, emphasizing what Dessoir tried to express more than sixty years earlier, namely the *active* aspect of haptic perception and its close relation with the movements of the body. It was further established that this process is automatic and unconscious, and that it take into account not only the mechanical forces voluntarily applied to the body and gravity, but also the involuntary internal mechanical forces such as the contraction of the muscles.

2.4 Human proprioception

As said before, the term *proprioception* refers to the perception of the movements of our body, associated with the perception of its position, or the position of its parts, in space. Proprioception is a *perception*, that is, the result of integration and the interpretation of multiple sensations from different sensory organs housed throughout the body (see Section 2.3.1). Most of this information comes from the sensory receptors of the muscles and tendons, the inner ear (see Section 2.2.1), and the nervous systems involved. This process allows us to perceive, unconsciously, our own movements and the spatial orientation of our body. This perception also allows us to perceive in detail the position of our body and its parts in relation to the external elements of our surrounding space. We also sometimes talk about *deep sensitivity*. There is no obvious difference to the terms kinesthetic or kinesthetic term. However, we speak of proprioception when we want to designate this faculty of our body, and rather of kinesthesia when we do interface design. But no rule really exists about it. We will take a closer look at this sense from an HCI point of view in Chapter 4 (Section 4.5, p. 86).

Wrap-up message

In this section we have presented some of the human perceptual systems, especially those we consider later in this manuscript. Firstly, we have categorized human perception into 3 main classes, namely *exteroception*, which encompasses sensations created by external stimuli (sight, hearing, touch, smell and taste), *interoception*, which includes sensations from our internal organs, and *proprioception*, which refers to the consciousness of our movements and our body in space. Then, we provided more details on perceptual systems related to hearing, touch and proprioception in order to understand them more finely and thus better exploit them for the design of auditory interfaces.

The auditory system proves to be a very good candidate among the remaining non-visual senses to unload the often overloaded sense of vision. From an anatomical point of view, the auditory system is mainly composed of the outer ear, then the middle ear and finally the inner ear. These 3 anatomical stages participate in the transform process of a physical audio signal (energy) in a biological signal (sensation). This process, coupled with the interpretation that is made by the brain, gives some characteristics to our sound perception. In particular, we perceive sounds according to 5 main dimensions: *pitch* (frequency perception), *timbre*, *loudness*, *time* and *spatialization*. A remarkable characteristic is the *cocktail party effect*, which is the ability to focus on a particular speakers in a noisy sound environment. Hearing strongly contributes to the mental representation we build from our surrounding environment, and allows us to “visualize” areas that are normally inaccessible to vision (such as dead fields or occulted areas), and to react almost instantly to an unfamiliar or out of context sound. Auditive information, coupled with visual information, allows us to perfect our evaluation of distances. The human ear is quite effective when it comes to locate sound sources placed horizontally along the transverse plane, but we are lacking of precision when it comes to locate sound sources which are placed on the sagittal plane (axis vertical). Moreover, when sound sources are aligned on a same azimuth, as it could be the

case in ATC, our localization abilities are bad (*cone of confusion*).

Next, we presented the somatosensory system, as well as the fact that our sensations are the result of the treatment of sensory receptors with external stimuli, located everywhere in our body. In particular, touch sensation is mainly derived from cutaneous sensory receptors. Also, there are 3 types of perceptions related to touch (or haptics): *cutaneous*, or *passive perception*, which is indirect and involuntary, *tactile-kinesthetic perception*, which is an active cutaneous perception in which a conscious tactile exploration is put in place, and *perception haptic*, which is a tactile-kinesthetic perception integrating other sensations, in particular those coming from our proprioception.

Afterwards, we briefly presented the sensory system related to the perception of our body, our movements and their place in our surrounding space, namely our *proprioception*. This is the result of a process of complex sensory integration and can also be referred to as kinesthetics. In the rest of this dissertation, we provide a state of the art on the existing interactive systems and concepts which are based on our auditory, tactile and proprioceptive perceptions. These are the topics of the following two chapters, namely *auditory displays* and *sonification* (Chapter 3), and, respectively, *haptics* and *kinesthetics* (Chapter 4).

All of this knowledge coming from the HCI literature will then be used to design new ways of interaction (Parts II and III).

Chapter 3

Relying on sound: Auditory Displays and Sonification

Résumé (Français)

Sonification et son interactif (État de l'art, partie 2)

Introduction

La prise en compte de notre perception auditive dans la conception d'interfaces Homme-machine correspondant à une discipline appelée *Auditory Displays*. De la même manière que les écrans représentent un médium pour fournir une information visuelle, les Auditory Displays représentent un médium pour fournir une information sonore. Le terme *sonification* a durant quelques années fait l'objet de réflexions, jusqu'à aboutir en 1997 à la définition suivante : “[...] l'utilisation d'un son non vocal pour transmettre de l'information. Plus précisément, la sonification est la transformation de relations au sein de données, en relations perçues dans un signal acoustique dans le but d'en faciliter la communication ou l'interprétation.” [97, 98].

Depuis lors, le sujet est de plus en plus étudié, et de nombreuses études ont été publiées dans les principales conférences d'IHM et de Visualisation de Données (CHI, UbiComp, UIST, VIS, Audio Mostly) et principaux journaux (TOCHI, TVCG, HCI, IJHCS). Néanmoins, des différences de points de vue subsistent encore au sein de la communauté. Comme expliqué dans [49], la frontière est encore mince entre la sonification en tant que méthode scientifique pour la transformation d'information en sons, et la sonification en tant que discipline abstraite, englobant la musique et certaines autres formes artistiques liées au son. Les limites de ce domaine de recherche ne sont ainsi pas encore totalement définies. C'est la raison pour laquelle Thomas Hermann a proposé en 2008 une définition plus robuste :

“Une technique qui utilise des données en entrée et génère des signaux sonores (éventuellement en réponse à une excitation supplémentaire facultative) peut être appelée sonification si et seulement si [77]

(C₁) le son en résultant reflète les propriétés objectives ou les relations qui existent dans les données d'entrée;

- (C₂) *la transformation est systématique, ce qui signifie qu’il existe une définition précise de la façon dont les données (ou les interactions facultatives) provoquent les variations du son en sortie;*
- (C₃) *la méthode de sonification employée est reproductible (i.e., partant des mêmes données et interactions, le son résultant doit être structurellement identique);*
- (C₄) *le système peut être utilisé intentionnellement avec des données différentes en entrée, ainsi que de manière répétée”.*

Comme le montre le *The Sonification Handbook* [1], la sonification est un vaste domaine de recherche dans lequel plusieurs disciplines sont impliquées (voir Figure 3.1 ci-après). Les plus courantes sont l’informatique, l’ingénierie audio, la psychologie et les sciences cognitives, mais on peut aussi ajouter à cette liste (non exhaustive) l’IHM, les technologies d’assistance, les facteurs humains ou bien encore l’ergonomie. Dans cet ouvrage, les auteurs ont considéré 4 axes pour leur présentation, à savoir : *Les principes fondamentaux de la sonification* (théorie, psychoacoustique, perception, conception des interactions, évaluation et conception), *les technologies de sonification* (sonification statistique, synthèse sonore, méthodes expérimentales et sonification interactive), *les techniques de sonification* (audition, icône auditive, earcon, sonification paramétrée et sonification par modèle) et pour finir des *Applications*. Bien que cette structuration soit intuitive et entièrement adaptée à une présentation complète du domaine, son exhaustivité présupposée nécessiterait un travail trop fastidieux. Par conséquent, les sujets les plus importants et les plus couramment étudiés liés à la sonification ont été conservés et synthétisés dans ce manuscrit, ainsi que ceux servant de base théorique aux chapitres suivants.

Techniques de sonification et concepts les plus courants

De manière générale, la sonification fait référence au processus qui transforme les données en son.. Cette opération peut prendre plusieurs formes, pouvant être basées soit sur des événements, soit sur des modèles spécifiques, soit les deux. Trois techniques correspondent à la sonification “événementielle” (ou EBS) :

- l’*audification*, qui correspond à “*la lecture directe de flux de données sous forme d’ondes sonores, permettant seulement quelques traitements mineurs pour que le signal devienne audible*” [49] ;
- les *icônes auditives* (*auditory icons*), basés sur une approche écologique de la perception auditive, associant de courts sons environnementaux à des événements discrets dans les données d’entrées, afin de créer des relations métaphoriques (par exemple l’obturation de l’objectif d’un smartphone est souvent associé à un “click” audio) ;
- les *earcons*, similaires aux icônes auditives dans la façon dont les données sont considérées et leur dimension temporelle brève, mais ne devant être composés que de sons entièrement synthétiques, sans valeur métaphorique préalable (par exemple, une mélodie indiquant un niveau d’autonomie).

Conceptuellement, l'*audification* correspondant à la seule action de jouer de la musique ou un son préenregistré. Si les données d'entrée peuvent être modélisées avec une fonction continue du temps, alors le simple fait d'associer des valeurs numériques successives à des valeurs de pressions successives (par exemple l'énergie sonore), puis de les lire dans le même ordre par l'intermédiaire d'un auditory display dans le but de les rendre audibles, est de l'audification. C'est pourquoi Florian Dombois et Gerhard Eckel ont ensuite donné la définition suivante : "*l'audification est une technique qui consiste à donner un sens aux données en interprétant tout type de signal unidimensionnel (ou un ensemble de données bidimensionnel, comme un signal) en amplitudes dans le temps, pour ensuite le reproduire via un haut-parleur afin de pouvoir l'écouter*" [47]. L'audification est la plus simple des techniques de sonification et semble être la plus couramment utilisée dans les phases de prototypage des processus de conception. Les exemples contemporains d'audification sont nombreux et concernent surtout les sciences : physique, médecine, économie, statistique. Elle a été fréquemment appliquée à la sismologie planétaire [46], par exemple. La première personne qui s'est intéressée à la contribution de l'audio en géologie, avec l'aide de Gregory Kramer, a été Chris Hayward lorsqu'il a présenté en 1992 ses travaux sur l'audification de données sismiques. Par ailleurs, une découverte scientifique a été faite lorsque Pereverzev *et al.* ont découvert en 1997 les oscillations quantiques dans les atomes de ^3He [129]. Ne rencontrant rien de remarquable avec leur équipement habituel, ils ont décidé de connecter leur signal à des hauts parleurs. Ils ont ainsi remarqué une dérive de fréquence importante, qu'ils ont ensuite été capables d'interpréter. La technique que nous avons utilisée dans la conception de la modalité Audio Focus, présentée ci-après dans ce manuscrit (voir Chapitre 5, p. 93), peut être qualifiée d'*audification interactive* dans le sens où les sons de moteur considérés sont joués à mesure que le mouvement de la tête de l'utilisateur module leur volume.

Alors que l'audification correspond au simple fait de rendre des données d'entrée audibles, les *icônes auditives* confèrent de surcroît à l'information fournie à l'utilisateur des dimensions structurante et événementielle. Le concept a été présenté pour la première fois par Bill Gaver en 1989 [63]. Il correspond au fait de lier explicitement des sons de la vie courante à des actions ou à un événements. Par exemple, dans la métaphore du bureau d'Apple, des actions ou événements particuliers ont leurs propres sons (supprimer un document, effectuer une action non autorisée ou simplement être notifié par une application). Ces sons sont reconnaissables par les utilisateurs, qui étaient auparavant formés naturellement à les reconnaître et à reconnaître leur sémantique parce que ce sont des sons environnementaux pourvus d'une grande affordance. Cependant, les sons utilisés pour concevoir les icônes auditives ne peuvent pas être parlés. En fait, la façon dont un son est utilisé et le contexte dans lequel il est utilisé détermineront sa nature d'icône auditive ou non. Cependant, une icône auditive doit être familière à l'utilisateur, en ce sens qu'elle ne doit conduire à aucune dissonance entre sa propre nature et l'événement ou l'action qu'elle suggère.

Les *earcons* sont des icônes auditives synthétiques, c'est-à-dire non écologiques, non environnementales. La première définition a été donnée par Blattner *et al.* en 1989 : "*[les earcons] sont des messages audio non verbaux utilisés dans les interfaces homme-machine pour fournir à l'utilisateur des informations sur un objet, un fonctionnement ou une interaction informatique*" [19]. Plus tard en 1994, Stephen Brewster a affiné cette définition dans sa thèse de doctorat : "*[sons] abstraits et synthétiques pouvant être utilisés dans des combinaisons structurées pour*

créer des messages auditifs” [27]. La différence fondamentale entre les icônes auditives et les earcons réside dans le fait que les earcons n’exigent pas de l’utilisateur qu’il soit préalablement habitué à les entendre. La relation entre les earcons et l’événement auquel ils sont associés doit être apprise explicitement par l’utilisateur. Un espace de conception comprenant un ensemble de paramètres à considérer pour leur conception est donné dans le tableau 3.1 ci-après.

Autres techniques de sonification

D’autres techniques de sonification correspondent à une formalisation des données sous forme de correspondances ou de modèles. Dans ce contexte plus technique, le concept de procédé de sonification ou d’algorithme est plus présent car le modèle est plus précis, contrairement à l’audification, aux icônes auditives et aux earcons qui font généralement référence à des sons pré-enregistrés. Ces techniques doivent satisfaire aux conditions précédemment évoquées, à savoir être *objectives* (C_1), *systématiques* (C_2), *reproductibles* (C_3), et *répétables* (C_4).

La technique de sonification dite “paramétrée” (en anglais *Parameter mapping* ou PMS) associe directement les données d’entrées aux paramètres sonores introduits dans le chapitre précédent, afin que tout ou partie des données d’entrée puisse être entendue(s) par l’utilisateur. Puisqu’il s’agit d’une représentation des données dans une dimension différente, les changements quantitatifs dans le temps sont conservés. Les informations sélectionnées sont ainsi communiquées en permanence à l’utilisateur. C’est actuellement la technique de sonification la plus populaire [49]. Elle offre une plus grande flexibilité dans la représentation des données par rapport à l’audification, aux icônes auditives ou aux earcons, cependant la correspondance (*mapping*) doit être soigneusement définie afin de ne pas générer de fausses interprétations des données d’entrée.

Puisque les données d’entrée sont *mappées* au travers de dimensions sonores en sortie via des techniques de sonification spécifiques, les données, littéralement, “jouent”, comme l’on pourrait jouer d’un instrument [77]. La sonification par modèle (MBS, pour *Model-Based Sonification*) a été introduite par Thomas Hermann en 1999 [79]. La principale caractéristique de cette technique, notamment par rapport à la PMS, est que le modèle de sonification est conçu de telle manière qu’aucun son ne soit audible sans intervention externe au modèle. Pour continuer l’analogie proposée dans [77], avec la MBS, l’instrument *est* la donnée d’entrée, ou plus précisément, l’instrument est un modèle de celle(s)-ci. Les interactions sont ici confiées aux utilisateurs, qui interagissent avec ledit “instrument” dans le but de recevoir des réponses sonores en guise de résultat. La MBS est plus sophistiquée que la PMS en ce sens qu’elle n’est pas une transposition des données telle qu’elles en sons, mais une agrégation composée des données (modélisées) et des interactions, tout en préservant l’intégrité temporelle des données d’entrées. L’objectif de la MBS est de donner aux utilisateurs la possibilité d’explorer les caractéristiques intrinsèques des données via l’audition.

Son interactif

De nombreuses techniques de sonification n’intègrent pas de moyens d’interaction. L’interaction dans la sonification apparaît lorsque les utilisateurs interagissent avec des systèmes qui fournissent de l’information sous forme sonore. Plus précisément, il s’agit d’interpréter les données qui sont

sonifiées, mais cela se produit lorsque l'utilisateur est focalisé sur ces données. La MBS a permis de construire un cadre conceptuel pour la conception d'interactions sonores. Cependant, l'une des premières applications utilisant la sonification interactive a été proposée en 1992 par M. Grohn avec Sonic Probe [68]. Le concept a ensuite été enrichi par Barrass *et al.* qui l'ont appliqué à l'exploration pétrolière et gazière en combinant la sonification des données géologiques des puits et un compteur Geiger virtuel. Cette dernière métaphore permettait aux utilisateurs d'entendre les variations de sons lors de l'exploration des fichiers de log, en fonction de la nature des données rencontrées au cours de leur exploration manuelle (pétrole, gaz ou roche). Barrass, inspiré par les concepts des Design Patterns [62], a également proposé une méthodologie pour la conception des interactions sonores qu'il a appelée Sonification Design Patterns [13]. De plus, un modèle de sonification interactive des données sismiques et médicales a été proposé par S. Saue dans [149]. Ce travail a permis de définir comment sonifier de grands ensembles de données spatiales et comment interagir sur le plan sonore avec celles-ci en introduisant une métaphore dans laquelle les utilisateurs ont été invités à parcourir les données en les écoutant.

Cependant, en dépit de toute considération pour un cadre conceptuel pour la conception de méthodes de sonification interactives, un système interactif agissant sur le son et présentant des similitudes avec celui introduit dans le présent manuscrit (voir Chapitre 5, p. 93) a été publié en 1981, 11 ans avant les travaux du Grohn [20]. Richard A. Bold, inspiré par les travaux de Peter G.W. Keen et Michael S. Scott Morton [94], a ainsi proposé sa contribution aux interfaces Homme-machine en anticipant ce qu'allaient être nos interactions quotidiennes avec les systèmes dans le futur à court et moyen terme. Il a ainsi imaginé que nous serions rapidement exposés à différentes vidéos et sources sonores de manière simultanée, et a proposé des techniques interactives dans le but d'adresser un tel environnement. Plus précisément, l'objectif était de focaliser l'attention de l'utilisateur sur les sons tout en faisant face à un mur d'écrans diffusant différentes images simultanément. Les sons émanant de ces différents téléviseurs étaient amplifiés en fonction du regard des utilisateurs ("eyes-as-output").

Cette anticipation rappelle bien sûr le concept d'interfaces fenêtrées ayant été introduit quelques années plus tard par Apple en 1984 [180]. Il représente une solution interactive au problème de l'attention dans un environnement qui offre plusieurs sources d'information simultanées. La volonté d'amplifier le son pour renforcer l'attention de l'utilisateur a été identifiée à plusieurs reprises dans la littérature récente. Nous pouvons également citer le système OverHear [160] (voir Figure 3.5) qui propose une méthode d'amplification des sources sonores à distance basée sur la direction du regard de l'utilisateur en utilisant des microphones directionnels. Appliqué au domaine de la visioconférence, cet outil permet aux utilisateurs, à distance, de se focaliser sur la parole d'un interlocuteur en particulier, dans un environnement pouvant facilement devenir bruyant.

Un autre domaine dans lequel l'interaction sonore peut contribuer est le domaine de l'*analyse immersive*, ou *immersive analytics* [50]). Ce domaine, à la frontière de la visualisation des données, des environnements immersifs et de la réalité virtuelle, de l'infographie, de l'interaction et de l'analyse de données, vise à définir un cadre de recherche pour le développement de méthodes permettant l'analyse naturelle et collaborative des données et la prise de décision. Dans ce contexte, le son, associé ou non à l'interaction, a un rôle à jouer. Dans un environnement immersif, par exemple, l'interaction spatiale des sons pourrait accroître le sentiment de présence

et fournir des moyens plus naturels et/ou efficaces d'interagir. En fait, plusieurs travaux liés à l'interaction sonore 3D ont été réalisés. Un concept présenté dans [40] ressemble à certains égards à la recherche qui sera présentée dans les prochains chapitres de ce manuscrit (Partie II, p. 93). Dans cette étude, les utilisateurs est entouré par des sources sonores organisées selon une topologie en anneau" (voir Figure 3.6). Ils peuvent sélectionner des sources sonores spécifiques via la désignation 3D, leurs gestes et un système de reconnaissance vocale. L'objectif est de fournir un moyen d'explorer la scène auditive via l'utilisation du concept de manipulation directe [87] et cette métaphore en anneau qui représente un environnement structuré.

3.1 Introduction

The consideration of our auditory perception in HCI design is inscribed in a field called *Auditory Displays*. As auditory displays are for auditory sensory channels what visual displays are for visual sensory channels, we can extend the comparison by saying that sonification is for hearing what visualization is for vision. This analogy with data visualization is a good entry point to understand what Auditory Display is as a research topic. As stated by Reuter *et al.* in 1990 [136], “*the use of sound in data representation is the auditory counterpart of data visualization*”. Even before, in 1989, Bill Bruxton seemed to be the first one to create—or at least to use publicly—the term *sonification* at the ACM Conference on Human Factors in Computing Systems (CHI), where he suggested that “[...] *this activity [will be] called sonification*” [136]. Since this event, the term has matured into people’s minds and has become a well-established research area. The International Community for Auditory Displays (ICAD) was created in 1992 under the initiative of Gregory Kramer, and the auditory displays and sonification community started to grow. Nowadays, sonification seems to be a sub-part of the Auditory Display as a scientific field. In 1997, a more modern and precise definition was formulated by Kramer:

“Sonification is defined as the use of non-speech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretations.” [97, 98]

Since then, the topic became more and more studied, numerous studies were published in top HCI and Visualization conferences (CHI, UbiComp¹, UIST², VIS³) and journals (TOCHI⁴, TVCG⁵, HCI⁶, IJHCS⁷). Another specialized conference was also born in 2001, named *Audio Mostly*⁸. Nevertheless, there are still viewpoint differences within the community to develop a strict definition of the term *sonification*. As explained in [49], the frontier is still thin between sonification as a scientific way to map information into sounds and sonification as a more abstract field, also encompassing music and sound-related art forms⁹. Hence, the boundaries of this research area are still not completely defined. That is the reason why a more robust definition was proposed to contributors of the field by Thomas Hermann in 2008, which includes the following 4 distinct conditions and on which we rely on in this manuscript:

¹ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)

²ACM User Interface Software and Technology Symposium (UIST)

³IEEE Visualization Conference (VIS)

⁴ACM Transactions on Computer-Human Interaction (TOCHI)

⁵IEEE Transactions on Visualization and Computer Graphics (TVCG)

⁶Human-Computer Interaction journal (HCI)

⁷International Journal of Human-Computer Studies (IJHCS)

⁸See <https://audiomostly.com/>

⁹To give one example, in music, we can think of the concept of *ambient* and shortly after, *generative* music introduced in the early 1970s by Brian Eno and others. Without a strict definition of sonification as a scientific field, the latter technique could be seen as a form of model-based sonification (see Section 3.4.2).

Sonification: “A technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering) may be called sonification if and only if

- (C₁) The sound reflects objective properties or relations in the input data;
- (C₂) The transformation is systematic. This means that there is a precise definition provided of how the data (or optional interactions) cause the sound to change;
- (C₃) The sonification is reproducible: given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical;
- (C₄) The system can intentionally be used with different data, and also be used in repetition with the same data.” [77]

As depicted in *The Sonification Handbook* [1], sonification remains a vast domain of research in which several disciplines are involved (Figure 3.1). The most common ones are Computer Science, Audio Engineering, Psychology and Cognitive Sciences, but we can also add HCI, Assistive technologies, Human Factors and Ergonomics. In this handbook, the authors considered 4 axes for their presentation, namely *Fundamentals of sonification* (theory, psychoacoustics, perception, interaction design, evaluation and design), *Sonification technologies* (statistical sonification, sound synthesis, experimental methods and interactive sonification), *Sonification techniques* (audification, auditory icons, earcons, parameter mapping and model-based sonification), and *Applications*.

Although this structure is intuitive and entirely adapted to a complete overview of the field, obviously it has already been proposed and its presupposed completeness would require a too tedious work—indeed, other topics had to be addressed in this dissertation (e.g., ATC, human sensory systems, haptics and kinesthetics). Therefore, the most important and commonly studied topics related to sonification have been retained and synthesized, as well as ones serving the purpose of subsequent chapters of this manuscript. After a brief historic presentation of the research field (Section 3.2), the rest of the presentation will be organized along the 3 following topics: i) *Basic sonification concepts and techniques* (Section 3.3), ii) *Other techniques* (Section 3.4), before finishing with the most interesting part for the present research, namely iii) *Interacting using sound* (Section 3.5). In the last section, a chapter overview will be given as well as the main information to consider for the rest of this dissertation (Section 3.5.2).

3.2 Brief history

Depending on the point of view, Auditory Displays and Sonification can be qualified sometimes as novel, and sometimes as well-established research area. Actually, they are nearly 30 years old now (in 2019), which can be seen as a quite mature age for a Computer Science branch. Nevertheless, as a research field, this remains fundamentally novel. However, beyond the research field or even the concepts, the wish to convey information through sound is almost as old as the first human civilizations themselves. Actually, it is generally accepted (while there is still no clear evidence) that one of the first examples of sonification was conceived by the people of the first Mesopotamian states, which are dated from the end of the fourth millennium B.C. (to the beginning of the third

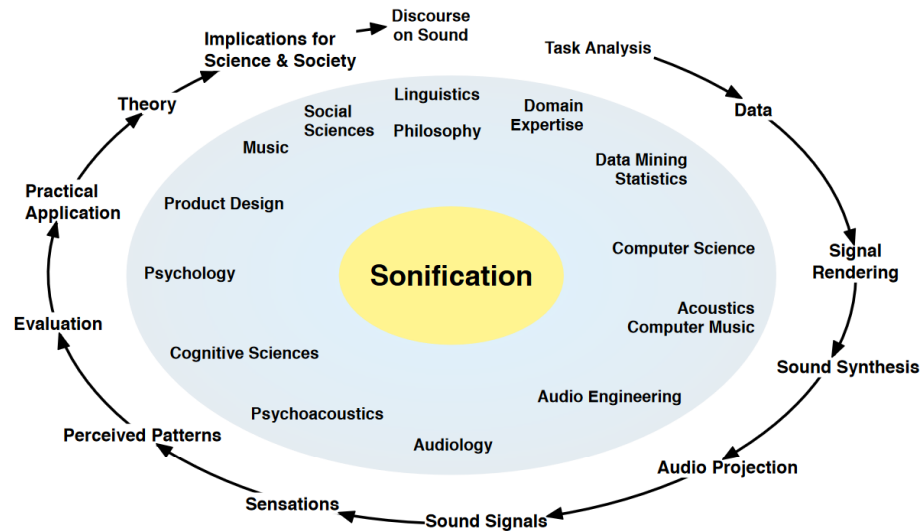


Figure 3.1 — The different disciplines involved in sonification. “The outer perimeter depicts the transformations of information during the use cycle, the inner circle lists associated [...] disciplines [...] This diagram is surely incomplete” [78].

century). Around 3500 BC approximately, specialized people called “auditors” reported potential anomalies in Pharaoh merchandise accounts by listening to the sound made by the grains moving in and out the warehouse, before and after a transaction¹⁰ [185].

Well after, ancient Greeks designed an alarm that was made of a clepsydra and a water organ [177]. This technique was also used during medieval China to have a quantification of elapsed time. During Pythagoras period, between approximately 575 and 495 BC, another technique similar to sonification was used way before its formalization to express the quantity (or ratio of energy) that exists within mathematical relationships. The concept, called *The Harmony of the Spheres* [185], associated musical tones to the size of the objects that can be seen from the Earth. Well after, Kepler was inspired by this principle in his *Harmonices Mundi* (1619) in which he applied the same principle to the orbit of the planets around the Sun, by modulating the pitch according to their angular displacement as seen from the Sun¹¹ [172].

More recently, several devices based on sonification technique have been commonly used. Perhaps the most famous of them is the Geiger counter, invented by Hans Geiger in 1913 and configured by Geiger and Walther Müller in 1928. Even today (however in more technologically up-to-date versions), his tool is used to measure ionizing radiation by sonifying radioactive particles using an old telephone (or at least the same principle) every time one of them collide with the radiation chamber (the “tube”). Another example is the optophone, made by Dr Edmund Fournier d’Albe in 1913 to help people with visual impairment to read by associating letters with time-varying chords of tones. A last great example of sonification device is the well-known stethoscope, which was invented in 1816 in France by René Laënnec (see Figure 3.2).

¹⁰This example may seem a bit excessive at first sight, however, the fact of undertaking a mechanical action in order to create a sound to extract information is, by definition, sonification.

¹¹Since then, this principle have been repeated several times [12, 133], even by the NASA to sonify the Sun (see <https://www.nasa.gov/feature/goddard/2018/sounds-of-the-sun>).



(a) Early version of a Geiger counter (Geiger-Müller tube, 1932). (b) A blind women reading a book on an optophone in 1921 (Credit: Blind Veterans UK). (c) An antique model of stethoscope (no known date).

Figure 3.2 — Examples of vintage sonification devices that have been made during the 18th and 19th centuries.

3.3 Basic sonification concepts and techniques

There are few HCI concepts and metaphors associated to sonification, which are presented in this section from the most general to the most specific. Generally speaking, sonification refers to the process that transforms data into sound, i.e., to represent the information with audible elements. This is done by playing a sound with the goal to associate it to a particular information. This operation can take few forms, which can be based on events or specific models. The “basic” techniques presented in this section may eventually refer to Event-based Sonification (EBS), and are 3 in number: *audification*, *auditory icons* and *earcons*. As define in [49]: “audification is the direct playback of data streams as sound waves, allowing only some minor processing for the signal to become audible”, “auditory icons are based on ecological approach to auditory perception, associating short environmental sounds with discrete events in the data in order to create metaphorical perceptual relationships, e.g., the mechanical “click” sound in digital cameras”, and “earcons are similar to auditory icons regarding how data are considered and with respect to brevity, but using entirely synthetic sounds with no prior metaphorical value, e.g., a melody indicating the battery level in mobile phones”. Strictly speaking, these three techniques cannot be assimilated to sonification because they do not refer to a mapping of the data, but only to sonic events that are temporally constraint (and generally quite short ones) and associated to a meaning. Anyway, there is still a confusion in the literature concerning this point. In this section, we will consider these three techniques as “basic forms of sonification”.

3.3.1 Audification

The only action to play music or any pre-recorded sound is audification. Gregory Kramer gave another definition in [97]: [...] “the direct translation of a data waveform into sound”. To go further, if the input data can be modelled with a continuous function of time, then the simple fact to associate successive numerical values with successive air-pressure values (e.g., sound energy), and then to play them in the same order through an auditory display to make them audible, is *audification*. This is why Florian Dombois and Gerhard Eckel then stated the following definition:

[audification] “is a technique of making sense of data by interpreting any kind of one-dimensional signal (or of a two-dimensional signal-like data set) as amplitude over time and playing it back on a loudspeaker for the purpose of listening” [47]. Audification is then the most “simpler” of the existing sonification techniques, including auditory icons, earcons, parameter-mapping and model-based sonifications (see next sections), and appears to be the most commonly used when approaching a new topic where sonification could be used. These latter techniques are then preferred but audification can be useful in particular context, for example when raw data can already be sonified without modification.

Contemporary examples of audification are numerous, and are most concerned with sciences: physics, medicine, economy, statistics. This sonification technique was frequently applied to planetary seismology [46]. The first person who investigated the contribution of audification to the field, supported by Gregory Kramer, was Chris Hayward when he presented in 1992 his work on the audification of seismic data coming from earthquake. Still in physics, a scientific discovery was made when Pereverzev *et al.* found in 1997 the quantum oscillations in ^3He atoms [129]. While meeting nothing remarkable with their usual equipment, they decided to connect their signal to audio headphones. Then, they noticed a significant frequency drift they were able to interpret. We could also cite the previously mentioned sonification of solar wind [4]. Another example is a sonification method of market stock data that have been proposed in [60, 121]. The authors showed that using sonification techniques, subjects significantly better predicted (70%) prices movements (or using sonification and visualization, also 70%) than using only the visualization (61%).

Besides, the technique we used in the design of Audio Focus interaction modality, presented in this manuscript, can be qualified as *interactive spatial audification*, since engine sounds are played while user’s head movement modulate their loudness. This specific point will be addressed in more detail in the Chapter 5 (p. 93).

3.3.2 Auditory icons

While audification is the simple fact to make input data audible, *Auditory icons* provide a structuring and event-driven dimension to the information transmitted to the users. The concept was first introduced by Bill Gaver in 1989 [63]. They represent everyday life sounds explicitly linked to an action or event. For example in the Apple Mac OS desktop metaphor, particular actions or events have their own sounds (deleting a document, doing a unauthorized action or simply be notified by an application). These sounds are recognizable by the users, who were previously naturally trained to recognize them and their semantic because they are environmental sounds—we then speak about *auditory affordance*. They will quite-directly understand what is going on by hearing these auditory icons. However, the sounds used to design auditory icons could not be made of, nor integrate, speech or part of speech. Besides, a piece of music can be an auditory icons. Actually, the way a sound is used and the context in which it is used will determine its nature of auditory icons or not. However, an auditory icon must be familiar to the user, in that they must not lead to any “dissonance” between their own nature (i.e., what they are associated with) and the event or action they suggest. For example, the association between a sound of applause and a prohibited action may obviously lead to a misunderstanding, and ineluctably it will destabilize the users.

Several examples of auditory icons are given in [24]. As mentioned above, SonicFinder [63] is the most know example of everyday use of auditory icons: the authors associated ecological short sounds to particular actions and events in the Mac OS environment. The concept was later on enlarged by the same authors in the design of the SoundShark application [64]. An interesting study was led by Fitch and Kramer in which they applied auditory icons concept to medicine field. The findings were that the subjects (who were students short trained in anaesthesiology) performed quicker their actions using auditory display incorporating auditory icons than using only common visualization [54].

3.3.3 Earcons

Earcons are synthetic auditory icons but not ecological nor environmental sounds. The first definition was given by Blattner *et al.* in 1989: “[*earcons are*] *non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction*” [19]. Later on in 1994, Stephen Brewster refined this definition in his Ph.D. thesis: “[...] *abstract, synthetic tones that can be used in structured combinations to create auditory messages*” [27]. Because of auditory icons are made of environmental sounds, and earcons are made of synthetic sounds, the fundamental difference that exists between auditory icons and earcons is that earcons do not require the user to be previously accustomed to the sounds. In fact, the relationship between earcons and the event that they are linked with need to be learned by the user. They are made of musical tones essentially. A common example of earcon is the opening sound of Skype software, or the alert in airplane to signify that the seatbelt must be fasten. We also used earcons in our everyday lives while using our computers, smartphones, televisions, etc. A design space with a set or parameters to consider is given in Table 3.1.

Several types of earcons have been formalized, namely *one-element earcons*, which are basically made of simple synthetic sound like a simple musical note for example (however, they can have rhythmic properties). *Compound earcons* are composed with multiple one-element earcons. Since earcons, unlike auditory icons, need to be learned (as weak as this learning is), some more sophisticated techniques can be considered while designing them. In particular, several researches have led to the design of harmonically consistent earcons. *Transformational earcons* are designed along a grammar, which can be a musical range. Finally, *hierarchical earcons* are similar to transformational earcons in that they are constructed along rules, but they also incorporates hierarchical relationships. They are designed with tree structures, in which each node is an earcon that inherits the properties of its ancestors.

We can also mention the concept of *spearcons*, which have been introduced by Walker *et al.* in 2006 [171]. They are made of short speech messages conveying speech-only information. They are created by speeding-up pre-recorded (or generated) sentences, and are employed in the same way as auditory icons or earcons. Results suggest that they are more efficient (in terms of accuracy and reaction times) than auditory icons and earcons to give navigation cues to users in menu-based interfaces. We will not address them more in this manuscript because by definition, sonification seems to preferentially concern non-speech sounds.

Table 3.1 — Guidelines with a set of parameters for the design of earcons by Brewster *et al.* [19, 27, 26].

Parameter	Designation	Definition
<i>Timber</i>	E_1	When possible, the use of sounds with multiple harmonics is preferable. Sounds that are easy to understand are also recommended.
<i>Register</i>	E_2	The register of sounds used in the sonification cannot be meaningful if users are required to make absolute judgment.
<i>Pitch</i>	E_3	The combination of complex frequency structures (in a range of 150 Hz to 5 kHz) to build earcons is efficient when used along rhythm parameters to differentiate between each other's.
<i>Rhythm, duration, tempo</i>	E_4	Temporal dimensions should be different so that earcons are as differentiable as possible (but earcons should be kept short).
<i>Intensity</i>	E_5	Sound intensity is a very important parameter because it is the most common source of annoyance. It must remain high enough for the earcons to be perceived and understood, and low enough to fully integrate into the interface's sound environment.
<i>Spatial location</i>	E_6	This parameter may be considered using stereo or spatial sound. It is a good parameter to play earcons simultaneously (from different spatial locations).

3.4 Other techniques

Several techniques have been designed that make stronger links between data and sounds. These techniques refer to a formalization of the data in the form of mappings or models, which are designed to generate sounds as outputs (Figure 3.3). In this more technical context, the concept of sonification algorithm is more present because the model is more accurate, unlike for the audification, auditory icons and earcons which generally refer to pre-recorded sounds.

These techniques must satisfy that resulting sounds should be (see Section 3.1): *objective* (C_1) which refers to the fact that the meaning of the sounds must be ecologically appropriate; *systematic* (C_2) which means that every changes in the input data must be reflected in the resulting sonification; *reproducible* (C_3) which means that, using an identical sonification technique, a data set must lead to the same output sonification; and *repeatable* (C_4) which means that using different data, a same sonification technique is always applicable [77]. In this section we expose the main types of “advanced” sonification mapping and model-based techniques.

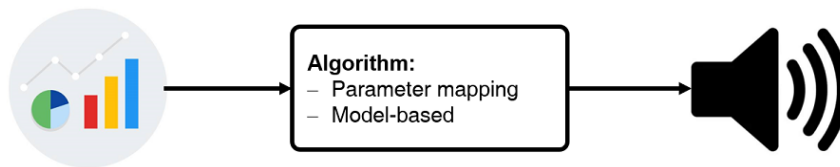


Figure 3.3 — The general process for sonification techniques: the data feed a model or an algorithm that produce sounds following a well-defined formalism.

3.4.1 Parameter Mapping Sonification

Sound can easily be modelled using several dimensions. Parameter-based Sonification (PMS) technique directly associate information to sound dimensions (see Section 2.2.2.1). This involves associating a particular dimension of the input data (depending on its nature) with an output sonic dimension such as volume/loudness, duration, pitch, time stamp, envelope characteristics, brightness and so ever, so that the input data, or at least one of the subpart, can be heard. Since this is a representation of the data in a different dimension, the quantitative changes over time are kept. The selected information is thus communicated continuously to the user. This is currently the most popular sonification technique [49]. It gives greater flexibility in the representation of data compared to audification, auditory icons or earcons, however the mapping must be carefully chosen so as not to generate false interpretations of input data (at least perceptual ones).

Several study have been led ton sonify data using PMS. Actually, in the same way that there is a strong analogy between visualization (set of techniques used to visualize data in a structured way) and sonification (set of methods to sonify data in a structured way), there is also a strong analogy between scatter plots visualizations and PMS. In this regard, Madhyastha and Reed qualify PMS as *multi-dimensional “sonic scatter plot”* [106].

Likewise, in a will to design new interactive technique to help people with visual impairments to apprehend pie chart diagrams, or simply to have a sonic method to enrich visual representations, Franklin and Roberts studied in 2003 a method to sonify pie chart diagrams [56]. Several sound

dimensions (or parameters) have been investigated, including pitch, loudness, timbre, location, rhythm, duration and tempo, to produce audible pie charts representations. After implementing and evaluating five different designs, they found that different mappings are efficient (including one which also seemed to be efficient to sonify bar charts). Another example could be the parametric “orchestral” sonification of electroencephalography (EEG) signal proposed in [81]. Wanting to rely on the ability of the human ear to distinguish several simultaneous signals, the authors designed a parametric sonification of EEG signal, by not only linking the “simple” pitch dimension to the amplitude values (one parameter), but by assigning 6 different EEG frequency bands to MIDI instruments (*real-time multi-parametric data binding*).

3.4.2 Model-Based Sonification

Since input data are mapped into output sonic dimensions using advanced sonification techniques, the data literally “plays” on an instrument [77]. Model-based Sonification (MBS) was introduced by Thomas Hermann in 1999 [79]. The primary point is that the sonification model is designed in such a way that none is audible in output without external interaction. To continue the analogy proposed in [77], with MBS the instrument *is* the data, or more precisely, the instrument is a model of the data. Here, interactions are left to the users, who excite the so-called “instrument” to receive sonic response(s). MBS is more sophisticated than PMS in that it is not a basic transposition of data into sounds, but an aggregation composed of the (modelled) data and interactions, while also preserving data continuity into a temporal dimension. The aim of MBS is to give users a way to explore the intrinsic characteristics of the data through sound.

Examples of MBS are numerous. Bovermann, Hermann and Ritter introduced in 2006 a MBS technique called *Tangible Data Scanning sonification model* that immerse the users into a sonic 3D environment, in which they can interact with using a physical object in their hand [21]. In particular, the users can discover geometric surfaces or other excitable objects using tangible interaction to create a personal representation of the data [21, 142]. As another application example, Sturm collaborated with oceanographer in 2002 to sonify ocean waves using MBS [164].

3.5 Interacting using sound

Many sonification techniques does not integrate any interaction mean. It is the case for the most part of alerting, monitoring or ambient sonifications (such as auditory icons). Interaction in sonification appears when humans interact with systems that are providing information through sounds. More specifically, the purpose is still to interpret the data which are sonified, but now it occurs when the user is focused on this data. In this section we relates some of the previous studies involving interaction in sonification.

3.5.1 Sonic interaction and MBS

As we saw in the previous section, MBS was first introduced in 1999 [79]. Since then, MBS allowed to build a conceptual setting for sonic interaction design. However, one of the first application using interactive sonification was proposed in 1992 by M. Grohn with Sonic Probe [68].

The concept was later enriched by Barrass *et al.* [58, 14], who applied it to oil and gas exploration by combining the sonification of well-logs geological data and a virtual Geiger counter. This latter metaphor allowed the users to hear variations in sounds while exploring the log files, depending on the nature of the data encountered during the exploration (oil, gas or rock). This 3D manipulation tool was presented to many domain experts and was well accepted. Sonic interaction was also applied to medicine by proposing real-time acoustic tool to support surgeons [90]. Barrass, inspired by Design Patterns [62], also proposed a methodology for sonic interaction design which he called Sonification Design Patterns [13]. Additionally, a model of interactive sonification of seismic and medical data was proposed by S. Saue in [149]. This work defined how to sonify large spatial data sets and how to interact sonically with it by introducing a walking metaphor in which users were invented to simply walk through the data by listening to sonic cues.

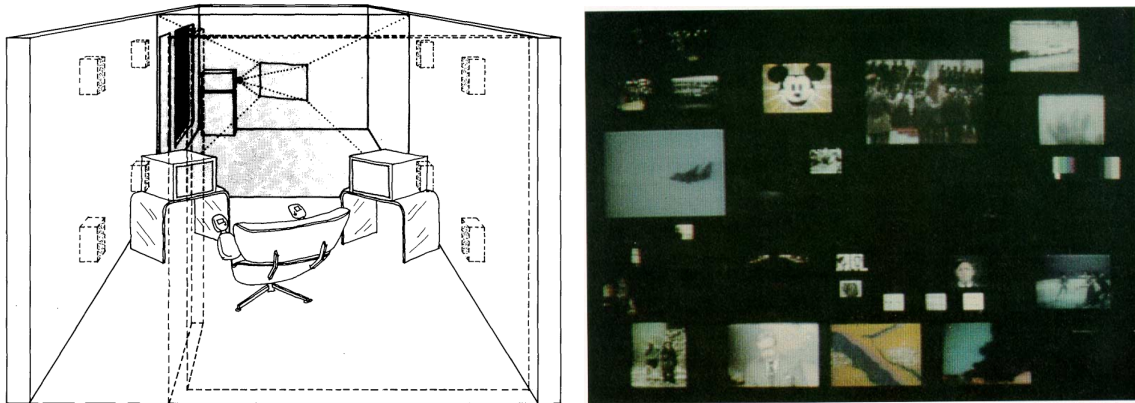
3.5.2 Other examples of sound interaction

However, despite all consideration for a conceptual design framework for interactive sonification, an interactive system acting on sound that has similarities with the one presented in the present manuscript (see Chapter 5, p. 93) was published in 1981, eleven years before the work of Grohn [20]. Richard A. Bold, inspired by the work of Peter G.W. Keen and Michael S. Scott Morton [94], proposed his contribution to “man-machine interfaces¹²” with an anticipation of what he expected for the everyday future interactions. At this time, he imagined that we will be quickly exposed to different simultaneous videos and sound sources. He gave a name to this particular environment, “World-of-Windows”, and proposed specific techniques to deal with. More precisely, the goal was to focus user’s attention on sounds while facing a wall of screens operating simultaneously and broadcasting different images (e.g., approximately 20 simultaneous images and sounds, see Figure 3.4). The interaction was in that the sounds emanating from these several televisions were amplified according to the users’ gaze (“eyes-as-output”). More pragmatically, when the users looked at a televisions, the sounds associated with this television was amplified so as to focus their attention.

This metaphor of the “world-of-windows” is of course reminiscent of the concept of windowed interfaces that was introduced few years later by Apple in 1984 [180]. It represents an interactive solution to the problem of attention in an environment that offers several simultaneous sources of information. The wish to amplify the sound to reinforce user’s attention has been identified several times in recent literature. These works, like the above mentioned Bolt’s one, can all be seen as attempts to computationally reproduce the cocktail party effect (see Section 2.2.2.3) through interactions with sound. In particular, we can cite the OverHear system [160] (see Figure 3.5) which provides a method for remote sound sources amplification based on the user gaze direction using directional microphones. Applied to the field of video conferencing, this tool allows the users’ to be remotely focused on a particular interlocutor speech, in an environment that can be noisy. The system uses an integrated eye-tracker and a webcam.

Currently, another area in which sound interaction can contribute may be the field of immer-

¹²At this time we rather talk of “man-machine interface” than HCI; the field seemed to not yet being clearly identified



(a) The experimental room, called “Media room”, was composed of two TV, a large screen at the end of it on which back-projected images were displayed, a chair on which the users were seated and two pressure-sensitive joysticks.

(b) A representative frame from the early Bolt’s “World-of-Windows” metaphor with several television displays displaying simultaneously full-color videos, non-regularly arranged. The size of each window may reflect the associated density of information.

Figure 3.4 — Bolt’s installation for his 1981 works exploring new kind of interaction techniques acting on displays and sounds, in a representative environment of near future anticipated interactions, using eye-tracking technology [20].

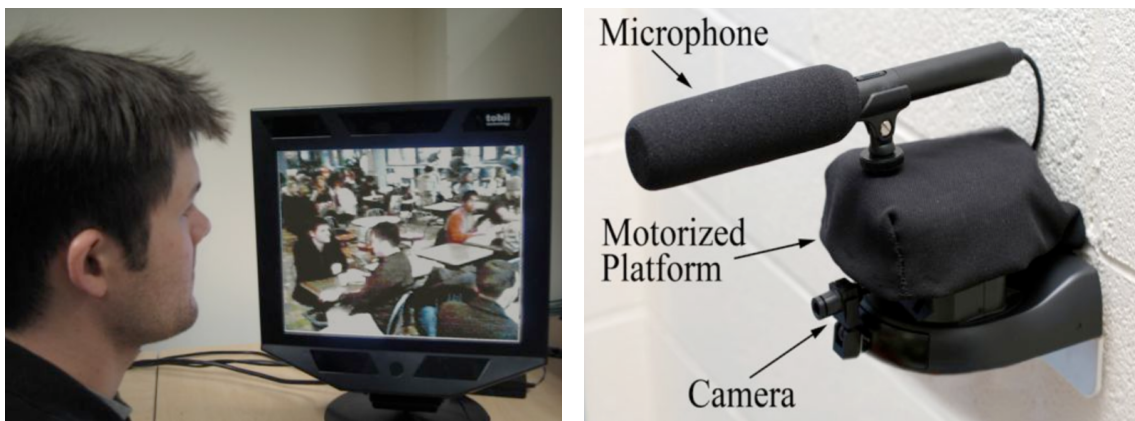
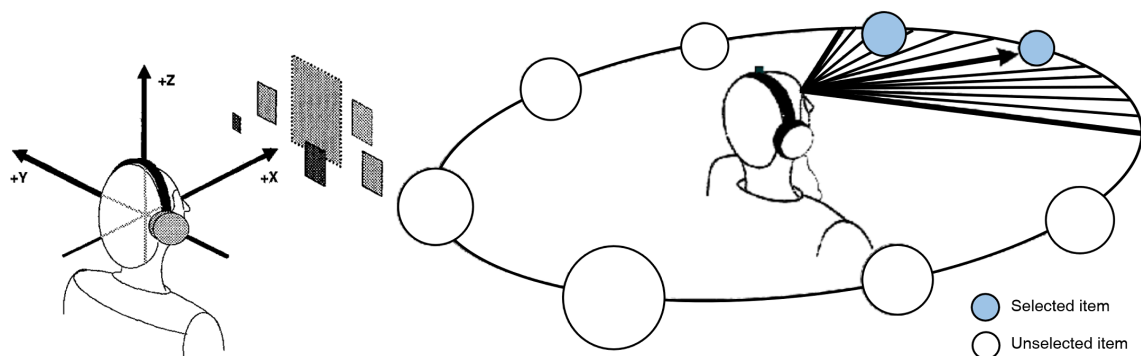


Figure 3.5 — The OverHear system [160]: OverHear remote display with full screen video stream and integrated eye tracker (left). OverHear robotic directional microphone with webcam (right).

sive analytics [50]). This field, at the frontier of data visualization, immersive environments and VR, computer graphics, HCI and data analytics, aims at defining a research framework for developing methods for understanding natural and collaborative data analysis and decision making. In this context, sound, whether or not associated with interaction, has a role to play. In an immersive environment, for example, spatial sound interaction could increase the sense of presence and provide more natural and/or efficient ways to interact. Actually, several works related to 3D-sound interaction have been performed. A concept presented in [41] and [40] resembles in some aspects to the research that will be presented in Chapter 5 of this manuscript (p. 93). In this study, the user is surrounded by interactive sound sources organized in a “ring” topology (see Figure 3.6). 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 functions [29, 34, 176]—with the use of direct manipulation [87] via this ring metaphor mapping of a structured environment.

As we saw in Chapter 1 (p. 13), Remote and Virtual Tower concept represents an immersive environment. The aim of the MOTO project, in which this thesis is inscribed, is to bring to this environment new elements of interaction that are not based on a the sense of vision, while promoting the immersion of ATCos by considering the principle of the embodied cognition. Sound is considered in this approach, as well as touch and, as we will see later in this dissertation, proprioception. A structured interactive sound environment such as the one proposed in [40], which involves hearing and kinesthetic interaction, can solve some issues related to particular situations that we have identified in this context (see Chapter 5, p. 93).



(a) Conceptual illustration of a spatial auditory display. (b) Conceptual illustration of the dynamically changeable auditory focus area, derived by processing head movements of the user.

Figure 3.6 — Spatial audio interaction in a ring topology [41, 40].

Wrap-up message

In this chapter we have presented the research field of Auditory Displays, and more particularly Sonification. A brief historic of the field allowed us to realize that we artificially use sound to transmit relevant information from the ages of the first civilizations.

Spatial sound plunges us more deeply into our task by improving our immersion. Therefore, the contribution of spatial sound would be beneficial to immersion in an augmented RVT environment. However, some limitations exist (see Section 2.4, p. 48). In the context of the design of sound interfaces intended to unload the visual sense, we could however rely on this skill in the perception of horizontal locations.

We have presented the contribution of the community of researchers interested in the Auditory Display. Hence, different sonification techniques already exist. Auditory icons and earcons are basic concepts for building structured sound interfaces. Audification and MBS allow interaction, the former being more archaic than the second, which is one of the most advanced techniques for sonifying data.

We have also introduced some existing concepts to interact with spatial sound. This theme is of particular interest to us for our study because in a context of degraded visibility or more generally

when the operators are deprived of their visual sense, sound is a good alternative to represent the environment thanks to our auditory spatial perception abilities. In RVT context, this was a solution that we considered and that we will detail in following chapters of this manuscript.

Chapter 4

Relying on touch and movements: the use of haptic perception and kinesthetics in HCI

Résumé (Français)

Perception haptique et sens sens kinesthésique pour l'interaction (État de l'art, partie 3)

Introduction

Le terme *haptique* est un mot-valise souvent employé de manière imprécise. En physique par exemple, il désigne une branche de la mécanique classique dont le but est d'étudier les relations existant entre le mouvement et ses causes. Ce mot a été introduit en 1892 par le psychologue allemand Max Dessoir afin de disposer d'un terme lié au sens du toucher, de la même manière que nous en disposions déjà pour la vue ("optique") et pour l'audition ("acoustique"). Ses racines lexicales proviennent du grecque *haptós* signifiant "palpable". De nos jours, ce terme est souvent employé pour désigner des technologies et concepts exploitant les sens du toucher et de la proprioception.

En informatique, le mot est plutôt utilisé pour désigner des technologies dites haptiques, ce qui correspond à une branche de l'IHM dont le but est d'étudier la contribution de la perception résultant de nos organes sensoriels liés au toucher et à notre proprioception dans le cadre d'une interaction avec les systèmes. Alors que le terme anglais *haptics* est lié au sens du toucher, l'autre terme anglais *kinesthetics* est lié à la proprioception. Il fait référence aux forces appliquées au corps, à ses mouvements et à sa perception dans l'espace. On peut aussi parler de *communication kinesthésique*, et on parle de technologie *kinétique* lorsque le dispositif concerné est exclusivement basé sur les mouvements du corps (comme le périphérique Kinect de Microsoft par exemple, voir Figure 4.8). Par conséquent, la formule peut parfois faire référence de manière abusive à toutes les technologies, concepts et dispositifs qui utilisent notre perception tactile, sous forme de vibrations ou de forces, ainsi que nos mouvements et notre sens de l'équilibre, afin d'interagir avec des systèmes.

Stimulation tactile pour fournir de l'information

Technologies existantes

La plupart des techniques de stimulation cutanée peuvent être purement mécaniques (vibrateurs de smartphones, interacteurs pour la réalité virtuelle, voir Figure 4.8). Ainsi, les vibrations sont généralement obtenues à partir d'un poids mis en mouvement à l'aide d'aimants à commande électrique. Une rotation crée un déséquilibre rapide et donc des vibrations perceptibles. Des stimulations mécaniques peuvent également être obtenues à l'aide de transducteurs sonores. Ces appareils sont en quelque sorte des haut-parleurs infrabasses, qui doivent être fixés sur le support que l'on veut faire vibrer. Ces appareils sont souvent utilisés dans les cinémas ou autres installations audiovisuelles. C'est ce type de dispositif que nous avons utilisé dans les expériences décrites ci-après dans ce manuscrit (Parties III, p. 93 et III, p. 145). Cependant, des technologies sont récemment apparues pour créer des sensations tactiles dans l'air grâce aux ultrasons [84]. Ce type de technologie est désignée en anglais par la formule *mid-air ultrasonic haptic feedback* [183]. Le plus connu de ces appareils est Ultrahaptics [32]. Il s'agit d'une matrice de 12×12 ou plus émetteurs à ultrasons. Le principe physique en jeu est basé sur le fait que lorsque des signaux ultrason de haute intensité se croisent, une sensation de picotement cutané peut alors être ressentie à l'endroit de ces points d'intersection. A l'aide de ce type d'"affichage" haptique, des objets 3D peuvent ensuite être *haptifiés* dans les airs. Cependant, la technologie n'est pas encore assez mature pour faire ressentir les vibrations avec une intensité suffisante [184, 37].

Fournir de l'information structurée au moyen du concept de tacton

Différents concepts IHM basés sur des sensations cutanées ont été élaborés pour fournir des informations structurées aux utilisateurs. Nous pouvons mentionner les icônes haptiques (ou *haptic* ou *tactile icons* en anglais) et la famille de concepts sous-jacents. Le concept d'icône" [156] est généralement utile pour véhiculer une information dite *structurée*. Avec une bonne conception, des structures ou des catégories d'informations peuvent par conséquent être fournies à l'utilisateur. Cela peut se faire via le sens visuel en utilisant les *icônes*, par l'audition en utilisant les *icônes auditives* et les *earcons*, comme nous l'avons vu dans le chapitre précédent (Chapitre 3, p. 51), mais également par notre thermoperception via le concept d'*icônes thermiques* introduit dans [182]. Pour le sens du toucher, MacLean a défini les icônes haptiques en 2003 comme "[...] de brefs signaux générés par ordinateur, proposés à l'utilisateur au moyen de la force ou de vibrations dans le but de transmettre des informations telles que la notification d'événement, du contenu ou un état (...) Les icônes synthétiques peuvent être perçues passivement ou activement [...]". Ce concept a ensuite été affiné un an plus tard via le concept de *tactons*, qui a été introduit en 2004 par Brewster et Brown [25]. Ils sont définis comme "*des messages structurés et abstraits pouvant être utilisés pour communiquer des concepts complexes aux utilisateurs de manière non visuelle*". La recherche a montré que de la même manière que les icônes (visuelles) sont un bon moyen pour fournir de l'information spatiales aux utilisateurs, leur équivalent tactile, à savoir les tactons, semblent être un bon moyen pour transmettre de l'information temporelle. Leur conception est assez similaire à celle utilisée pour les earcons car les deux concepts sont composés de signaux sonores.

Ils sont conçus via les paramètres de la perception passive, fournis dans le Tableau 4.1 ci-après dans ce chapitre.

Alerter l'utilisateur et fournir de l'information spatialisée

De nombreuses études ont été publiées au cours des dernières années, étudiant des moyens de fournir de l'information spatiale aux utilisateurs par le biais de feedback vibrotactiles. Dans le contexte de la conduite, plusieurs études utilisent des patterns vibratoires dans le but d'indiquer au conducteur des directions ou des obstacles à éviter. Petermeijer *et al.* ont montré dans [130] que les feedback vibrotactiles haptifiés dans le siège du conducteur sont un bon moyen de signifier que le contrôle manuel doit être repris (dans le cas d'une conduite hautement automatisée). Ils ont également montré que les stimuli vibrotactiles statiques entraînent des temps de réaction plus rapides que les stimuli dynamiques. Cette étude et ses résultats sont utiles à notre propos car les auteurs postulent que les conducteurs ne sont pas concentrés sur leur sens visuel puisque dans le cadre de la conduite autonome, des études ont montré qu'ils peuvent être afférés à d'autres tâches que la conduite (manger, parler, écouter de la musique, etc). Ils considèrent donc le cas où le conducteur est privé de son sens visuel et, par conséquent, plus alerte concernant ses autres sensations (comme le toucher). Il existe également de nombreuses études dans le domaine aéronautique. Van Erp *et al.* ont conçu un système fournissant aux pilotes des feedback vibrotactiles par l'intermédiaire d'une ceinture pour donner les directions des repères de balisage, et donc des indices d'orientation [52, 168]. Une autre étude porte sur les pilotes d'hélicoptères pour les aider à effectuer des manœuvres en vol stationnaire [134].

Interaction kinesthésique

Nous avons préalablement défini le terme anglophone *kinesthetics*. L'interaction kinesthésique fait référence à la prise en compte de notre corps, ou d'une partie de celui-ci, dans le but d'interagir avec les systèmes. Ce terme a été formellement défini pour la première fois en 2008 par Fogtman *et al.* [55], et est basé sur des travaux antérieurs datant de 1999 [104] considérant le corps comme un dispositif d'interactions (Full-Body Interaction [128]). Fogtman s'inspire également des travaux de Dourish [48] et Klemmer [95] sur la *cognition incarnée*, qui est un modèle dans lequel la représentation de l'espace dépend de nos sensations perceptives et motrices. La cognition incarnée peut être significativement améliorée dans un environnement immersif en considérant les sensations provenant du système somatosensoriel. L'interaction kinesthésique pourrait également renforcer cette sensation en permettant à l'utilisateur d'agir aussi naturellement que possible, et donc, dans une perspective écologique, de renforcer son sentiment d'immersion. Elle révèle le potentiel du corps dans la conception d'interaction en favorisant l'interaction naturelle et les interfaces basées sur le mouvement. Le concept d'interaction kinesthésique est basé sur notre proprioception, pour créer un cadre interactif qui améliore l'implication de l'utilisateur et la sérénité du système.

Dans leur étude, les auteurs identifient trois axiomes principaux qui devraient être pris en compte lors de la conception d'une interface kinesthésique, à savoir [55]: la composante *physiologie*, l'*expérience kinesthésique* et les *technologies interactives*. L'aspect physiologique implique

la prise en compte par l'utilisateur de son propre corps ainsi que de ses propres mouvements dans l'espace. Une interface doit donc prendre en compte la place prise par le corps et ses mouvements, et promouvoir la perception de l'espace. L'expérience kinesthésique est directement liée à certains aspects physiologiques et fait référence à la compréhension, consciente ou non, par l'utilisateur des aspects kinesthésiques de l'interface. Le dernier aspect à considérer concerne les technologies interactives employées pour concevoir une interaction kinesthésique. Ces 3 thèmes, associés à sept paramètres (voir Tableau 4.2) définissent un cadre conceptuel pour construire une interaction kinesthésique.

4.1 Introduction

Haptics is a vast term which is often employed in a non-specific way within several fields (e.g., medicine and physiology, physics, human-computer interaction). In physics, the term *haptics* refers to a branch of classical mechanics whose purpose is to study the relations that exist between movement and its causes, i.e., forces and torques. This word was imagined in 1892 by German psychologist Max Dessoir in order to have a term related to the sense of touch, in the same way as we already had for sight (“optics”) and for hearing (“acoustics”). Its lexical roots come from ancient Greek terms such as *haptomai* (“I touch”), *haptós* (“palpable”) or *haptikós* (“suitable for touch”). Nowadays, the term “haptics” is often employed to designate the fields which is explored and exploited by the sense of touch and kinesthetic (by analogy with acoustic and optic)—even if this latter is a different field in itself. As sonification is the auditory counterpart of visualization, the term *haptics* is sometimes endowed with the term *haptification* [147, 42, 165, 126], which is, for touch sensory channel, what sonification and visualization are for auditory and visual sensory channels, respectively.

In computer science, the word is rather used under the theme of haptic *technologies*, and designates a branch of HCI whose goal is to study the contribution of the perception resulting from our sensory organs related to touch and our proprioception, within the framework of an interaction with systems. While the term *haptics* is related to the sense of touch, the term *kinesthetics* is related to proprioception and forces. It refers to the forces applied to the body, its movements and its perception in space. We can also speak of *kinesthetic communication*, and we talk about *kinetics* technology when the concerned device rely on body movements, specifically (such as force feedback devices or Microsoft Kinect, see Figure 4.8). Therefore, and with an abusive language, sometimes the formula *haptic technologies* may refer to all technologies, concepts and devices that use our tactile perception, in the form of vibrations or forces, as well as our movements and sense of balance, in order to interact with systems. An haptic or kinesthetic interaction can be designed in both directions: in return, the systems can also measure the forces that are applied by the users.

There are several terms that are frequently used in the context of haptics technologies. For example the term *force feedback*, which refers to the artificial application of a force that should be perceived by the users during an interaction. This is the case in many gaming devices such as joysticks, in the aeronautics field with Boeing sticks, or the *PANToM* device that will be presented in more detail in this chapter [110] (see Section 4.2). The term *tactile* is often used as part of an haptic interaction or feedback through the skin using our touch (i.e., cutaneous) sensation. In a more specific way, the term *vibrotactile* is used when using vibrations (via mechanical movements, air pressure, ultrasounds or electricity) in the context of an haptic interaction (or feedback) on the skin. We also qualified as *haptic display* a framework that haptifies input data via a specific process or an algorithm. The aim could be to haptify a text into braille for example [102] (see Figure 4.1), or to haptify a graphical environment [146]. The process to haptify data is called *haptic rendering*, which have been defined by Salisbury as “[...] the process of computing and generating forces in response to user interactions with virtual objects” [148].

This HCI field is, for several years now, largely explored with many publications in main con-

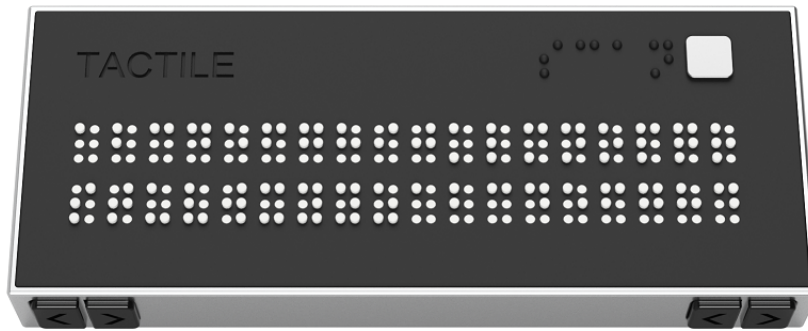


Figure 4.1 — Prototype version of the text-to-braille “tactile” device using a camera to capture text to translate, currently in development at MIT under the courtesy of Microsoft [102].

ferences (IEEE World Haptics Conference, 3DUI¹, TEI², ICM³, CHI, UIST, DIS⁴) and journals (ToH⁵, TOCHI, HCI, IJHCS). In the same way as for the previous chapter, we do not propose in this one an extensive bibliographic work on computer interfaces that use haptics because this represents a complete and vast topic in itself. We rather focus on a presentation of the domain and its history, the so-called founding devices, and the main current application areas.

As we will explain in the second part of this manuscript, during our experiments we coupled haptics with another interaction modality based on hearing. This is why, in this section we present the key haptic devices while focusing on the research questions to which they answer. In this perspective, we focus on the use of HMI devices that rely on haptics in a broad sense, to create feedback, feedforward, or even interaction between systems and users, by means of touch or, sometimes, kinesthetics. After having proposed a brief history of the consideration of haptics and kinesthetics as a research area in computer science (Section 4.2), we develop the main topic of this chapter. Specific areas were selected for their relevance, their maturity or for the current trends they arouse in HCI research. This is done along the following sections: i) *The use of haptics in Augmented, Virtual and Mixed realities* (Section 4.3), ii) *Tactile stimulation to provide information* (Section 4.4), and iii) *Kinesthetic Interaction* (Section 4.5).

Sections ii) and iii) will narrow the discussion around the questions that concern this study more specifically. Finally, we summarize the various information introduced in this chapter that will be used later in the document (Section 4.5).

4.2 History of haptics in computer science and early prototypes

Numbers of haptic devices have emerged since the beginning of the field, and it would be nearly impossible to list them all in a simple section relating the history related to the use of haptics in Computer Science. This is why we mainly focus in this section on the key devices and techniques

¹IEEE Symposium on 3D User Interfaces (3DUI)

²ACM International Conference on Tangible, Embedded and Embodied Interactions (TEI)

³ACM International Conference on Multimodal Interaction (ICMI)

⁴ACM Conference on Designing Interactive Systems (DIS)

⁵IEEE Transactions on Haptics (ToH)

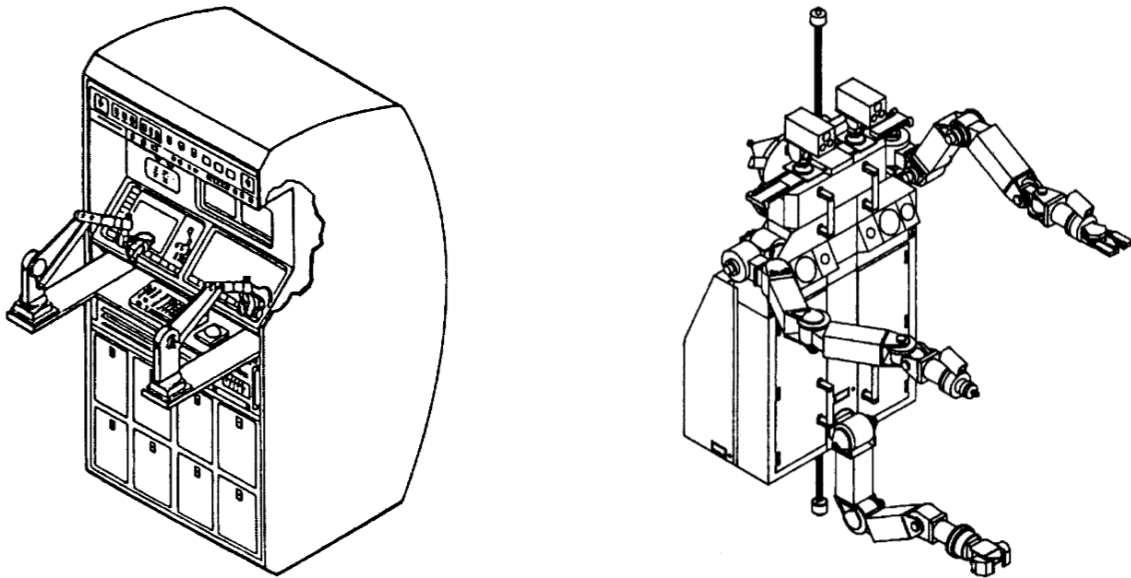


Figure 4.2 — The NASA’s Telerobotic Flight Servicer (TFS) (right) and its workstation (left), allowing an astronaut to control the system remotely (1990) [5]. The astronaut is here seen as a teleoperator who manipulates objects from a distant position.

that have been developed. This brief history covers the period from the mid-20th century to the mid-2000s. We will focus on more actual or advanced techniques and devices in the following sections.

Haptics technologies have been studied for many years. The desire to consider haptic feedback to interact with systems began in the 1950s, within a scientific community eager to implement means of *teleoperation* [163]. This term, which also encompasses the notion of *telepresence*, was introduced to qualify any interaction mean that allow the control of a remote system, *telepresence* being the fact to communicate sufficient information to the *teleoperators* in order to allow them to be sufficiently physically involved in the task. At this time, these concepts were an ideal to aim for researchers and practitioners from physics (especially from nuclear field), oceanography, space and also military fields. The telepresence concept has further led at NASA to build a remote assembling control system based on kinesthetic interaction and called Flight Telerobotic Servicer (TFS) [173, 5]. This robotic system was used to help astronauts to assemble the space station (see Figure 4.2). These ideas gradually took shape until the beginning of the 1970s when some “tangible” studies began to appear.

Probably the first haptic device using electricity comes from a U.S. patent which was filed by Thomas D. Shannon in 1972. It consists in a “*tactile communication device [...] intended to be used in pairs, to establish or permit tactile communications between two or more parties*” [152]. Basically, it was the idea of a telephone equipped with vibrotactile feedback. Anyway, two years earlier, one of the first consideration of haptics for a possible contribution to HCI (or Human–Machine Interaction) came from Carl E. Sherrick. At this time, a strong research trend, especially coming from the psychology community, was on the study of how humans perceive vibrotactile stimuli. In his paper [154], Sherrick studied the effects of the spacing on the body, the quality and the intensity of the limen (i.e., the “*interval between stimulus onsets at which the observer*

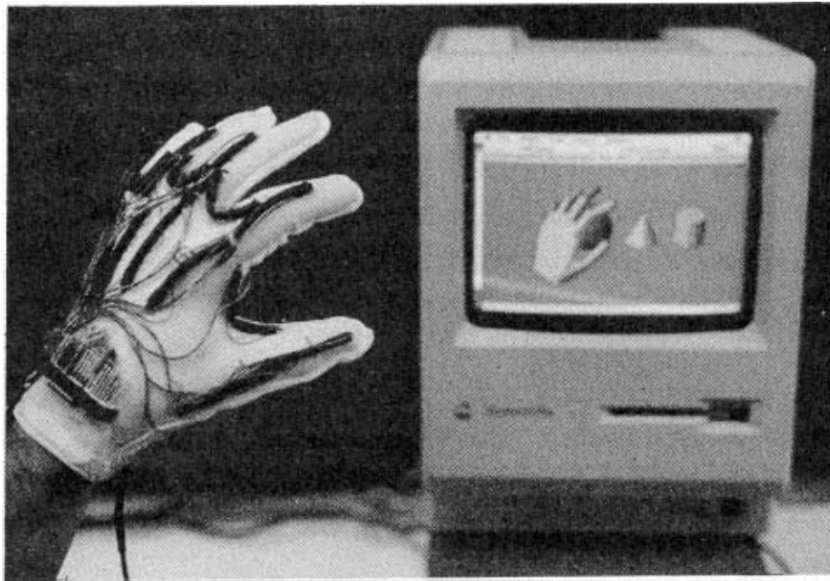


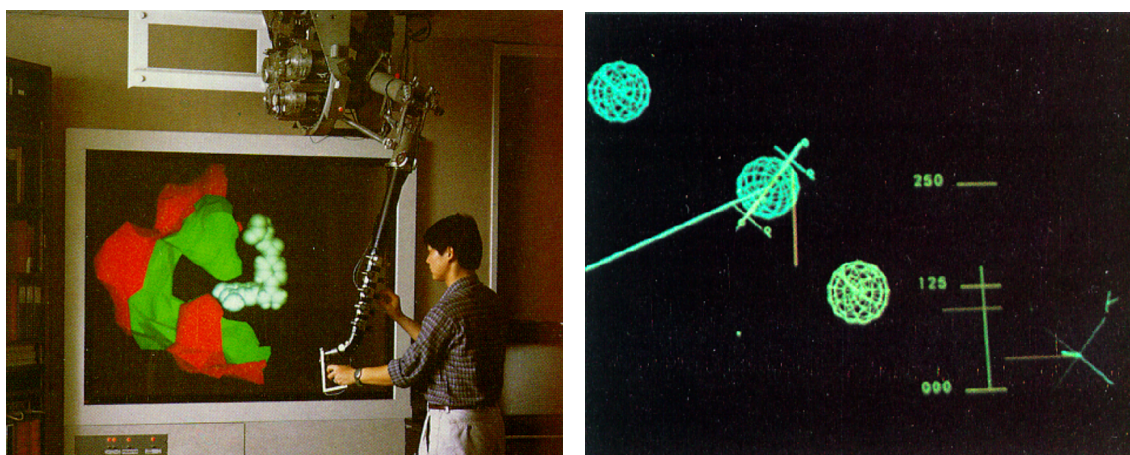
Figure 4.3 — The DataGlove device, presented by Zimmerman and Lanier in 1986, providing to the system real-time hand orientation and position, and gestures of the wearer [189].

can correctly identify the stimulus that came first on 75% of all trials”) between two ordered stimuli. Having previously found that this limen is about 20 ms on average for hearing, vision and touch [82] (but also for a combination of these modalities) he found that there are some limitations concerning the spacing of the two stimuli and also concerning their temporal disposition. Twelve years later, Sherrick *et al.* additionally found that the back could be a suitable body area to provide tactile stimulation [39].

These findings shall be interpreted considering the type of haptic feedback (electric, mechanical, etc) and the body’s area that is studied. For example in 1987, Frisken-Gibson *et al.* presented an haptic device used to study non-visual exploration of HCI with the fingertip, and therefore on a tiny area of the body, using non-vibrating haptic cues [57]. Made of a 8×8 array of solenoids, it was used to translate visual images into haptic cues by controlling the height of each pin (the same principle was later reconsidered for the design of the Tactile device presented in Section 4.1, Figure 4.1). Moreover, the users could zoom in and out the image using a mouse, and then choose to haptify specific areas. Using a technique which is similar to braille language, this device could help people with visual impairment.

On a different note but still considering haptics in HCI, probably the first kinesthetics device to use for hands was introduced in 1986, when Zimmerman and Lanier proposed a new “*hand to machine interface*” in the form of an haptic glove, namely *DataGlove* [188, 189]. Combining ultrasonic and magnetic flux sensors, this device was designed to provide real-time information to the system about hand position and orientation of the wearer’s hand, like the Leap Motion do nowadays [174]. It was also equipped with piezoceramic benders to provide tactile feedback in the glove (see Figure 4.3).

These few examples are certainly not exhaustive considering the period from the early 70’s to the early 90’s, however publications concerning haptic devices or HCI studies are relatively few



(a) The GROPE-III device with a user operating the mechanical arm in front of the explored molecular 3D model. This is one of the first example an haptic device to explore 3D structures.

(b) Visualization used in coupling with the arm on a secondary screen held by the user in his/her free hand to provide him/her the current possibilities of movements.

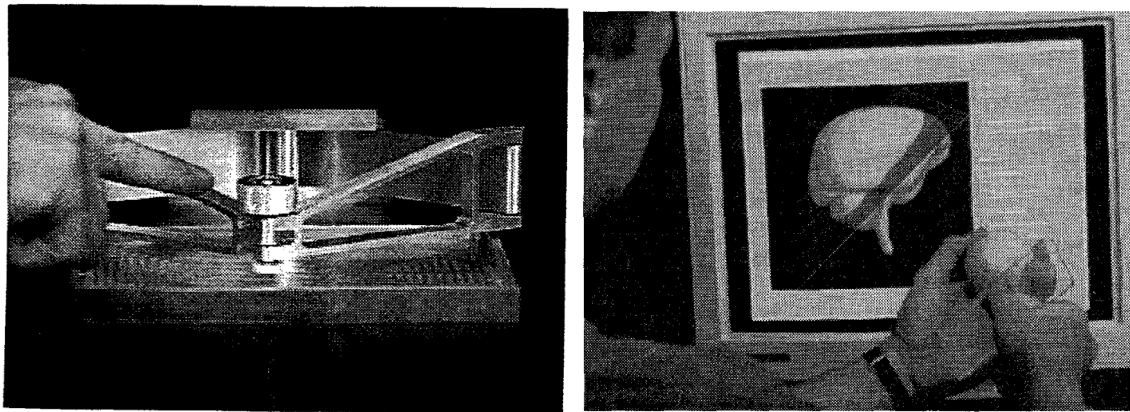
Figure 4.4 — The GROPE-III haptic device presented in the early 90's by Brooks *et al.* to help chemists to interact with 3D graphical models of molecules using a 6 Degrees of Freedom (DoF) mechanical arm [28].

through this period and these selected technologies were novel at their times. In 1990, Brooks *et al.*, through their project officially started in 1967, presented the GROPE-III haptic device [28] (previous models I and II are also described in the paper). This device is often cited when talking about fonder interactive devices because it is an important step in the literature, and for haptic technologies in a more general point of view. It can be seen as a real adaptive solution to telepresence and teleoperations. It provided scientists (more specifically chemists) with a mean to interact with 3D graphical models of molecules (see Figure 4.4). Designers who used it found that, coupled with visualization, this type of haptic display can significantly improve the exploration of impenetrable objects through the use of force feedback. Shortly after its development the device was widely used by scientists who found in it a very effective way to study the structure of molecules.

Starting from the mid-90s we saw that the number of publications was increasing significantly⁶. In 1994, Ramstein and Hayward introduced a device called *Pantograph* that allowed blind people to access graphic interfaces [135] (see Figure 4.5), with quite the same goal as the abovementioned work from Frisken [57]. The main idea was to display on the finger an haptification of the concerned HMI though 2 DoF. During the same year and at the same conference (i.e., CHI'94), Hinckley *et al.* presented a haptic interface that helped neurosurgeons to graphically define cutting plans [80]. The key point is that they successfully used direct manipulation principles [87], by physically manipulating a tangible object from the real world to define relationships in 3-dimensional virtual world.

Still in 1994, another key device was introduced by Thomas H. Massie and Kenneth Salisbury.

⁶This was found by searching for the keyword *haptics* in the ACM and IEEE digital libraries and also confirmed by Stone in [163].

(a) *Pantograph* haptic device [135].

(b) Passive interface props for neurosurgical manipulations [80].

Figure 4.5 — Mid 90's haptic devices examples.

The *PHANToM* device [110]⁷ is a kinesthetic device in the form of a mechanical arm coupled with fine-tuned force feedback. Users who manipulate it can feel forces materializing the collision with the handled remote or virtual object, increasing precision in their interactive experience. Its main function is quite similar to the one of GROPE devices, however the *PHANToM* device may be cheaper, smaller, more convenient to install and use, and incorporates vibrotactile feedback. Salisbury *et al.* also introduced the next year a panel of appropriated software techniques to consider [148]. Nowadays it represents an entire category of device since it was declined into several models and was successfully used in many field (sculpture, VR, or surgery, for examples [163]).

**Figure 4.6** — An early model of *PHANToM* device (left), and a more recent one (right) [110].

Projects such as GROPE and *PHANToM* were one of the first examples of devices to use for what is now called *haptic visualization*. As for sonification with sound, here the data remain unperceived from the start, and are transformed into something that is palpable. Jonathan C. Robert defines it as follows: “*In the instance of haptic visualization, the underlying model is an*

⁷*PHANToM* is standing for *Personal Haptic iNterface Mechanism*.

abstract concept that both holds the data and the mapping of that data into a tangible form” [144]. Since them many devices have been design to explore data [125].

Since the mid-90s, haptic and kinesthetic technologies have been deeply considered for HCI purposes. For example, the field of video games has been the vector for some major innovations, such as vibrotactile feedback, incorporated in console controllers or in the form of kits (see Figure 4.8). Designed to provide new sensations to players, they help to increase their immersion. Another common example from our everyday life is our smartphones. We have the possibility to be notified in a tactile way through vibrators since the mid-90s, early 2000s. However, currently one of the main application for haptics technologies remain immersive environments such as VR, AR and MR.

4.3 The use of haptics in Augmented, Virtual and Mixed realities

Virtual realities (VR, AR, MR) have long remained at the experimental stage because of a lack of computing power. Today, this barrier has completely disappeared with the emergence of new generations of CPUs and GPUs. Commercial considerations has taken precedence with the appearance of devices such as the Oculus Rift, the HTC Vive or the Sony PlayStation VR. VR has now become an everyday object. AR and MR tend to be as important as VR nowadays, however, some technological barriers still exist. In particular, we can talk about the relatively limited field of view that is offered by devices such as the Microsoft HoloLens. Anyway, they remain a major focus of study for HCI community.

These technologies, grouped under the term XR, represent—with others—the future of HCI. The will to immerse yourself in virtual worlds, whether for data analysis purposes [50], studies in other scientific fields such as space [103, 111] or psychology and medicine [108, 145], interactive visualization [86], or entertainment [72], is more and more justified. Immersion improves performance, and synthetic worlds allow us to place ourselves in contexts that we could not find or build in RR [51]. One of the ways to increase this immersion is to be inspired by Nature, by considering all our perceptive characteristics. This is how haptic technologies fit into this area. The number of published studies implementing our haptic and kinesthetic senses within virtual environments is increasing every year. The applications are very varied, and here we could cite hundreds of different studies. However, we will present some of the most recent ones to give an actual overview of this area.

With Elastic-Arm for example [2], the authors present an ingenious system for integrating passive haptic feedback into the user’s arm into a virtual immersive environment. The principle is based on a simple elastic placed between users’ hand and their shoulder, allowing them to feel a progressive resistance as they reach out their arm. The authors propose different use cases in which such device could make a contribution, such as the selection of distant objects, the perception of the limits of the immersive environment or even the perception of variable levels of effort. They also identify areas that could benefit, such as ergonomics and medical rehabilitation.

The recurring problem of developing fine haptic feedback on the fingers skin was studied in

2017 by Talvas *et al.* in [166]. With Soft Fingers, the authors have recently presented a promising model for the precise and interactive manipulation of virtual objects in an immersive environment. In a similar desire to perceive objects to the user by skin stimulation at the end of the fingers, Benko *et al.* introduced in 2016 their NormalTouch and TextureTouch systems [17]. Made up of mechanical actuators adapting to the surfaces encountered by the users, they allow them to feel objects when they touch their virtual surface (see Figure 4.7). Another example of device is Haptic Revolver [178], presented by Whitmire *et al.* in 2018. It is a haptic controller allowing tactile-kinesthetic perception in a virtual immersive environment. With such a device, the users can feel a texture using a wheel that moves and turns under their fingers. Several wheels are available to transmit different skin sensations. Moreover, Haptic Revolver is also tracked in space, which allows the geometry of the explored object to be associated with the haptified texture. This desire to make the user feel cutaneous sensations at the end of the fingers is very present in the literature of the field. Indeed, without an efficient system of haptic rendering for this very sensitive area, virtual immersive environments would not allow us to appeal to the haptic and tactile-kinesthetic perceptions, which represent very important parts of our perception of the world [66].

Other devices or principles use vibrotactile sensations to reproduce tactile-kinesthetic or haptic perception in VR. Using the vibrotactile features provided via the immersive headset they used for their study and (kinesthetic) cutting plans, Prouzeau, Cordeil *et al.*, using Scaptics [132], proposed a way to explore the density of clusters in scatter plots visualizations under immersive environments. Nevertheless, the use of vibrotactile feedback is far from being only found in abstract data visualization or within the currently highly studied immersive environments. Other more discreet applications use this type of perceptions to transmit concrete information to the users, and have been doing so for several years now.

4.4 Tactile stimulation to provide information

Haptic technologies are therefore numerous and contribute to many areas of HCI. Skin (tactile) stimulation, represents a big topic within the field. In this section we present the main technologies to implement tactile feedback. We will then see how they can be used to transmit information to the user. The last two paragraphs present some concepts to provide information of direction, or to attract users' attention.

4.4.1 Tactile devices and current technologies

The most part of techniques for creating stimulation on the users' skin rely on mechanical devices. These techniques can be purely mechanical, for example smartphones vibrators, vibration kits integrated into joystick game consoles, or VR interactors, to name a few (see Figure 4.8). With this kind of device, vibrations are usually obtained from a small weight put in movement using electrically controlled magnets. This rotation creates a fast imbalance and, thus, perceivable vibrations. Mechanical stimulations can also be obtained using sound transducers. These devices are in some ways almost infra-bass speakers, which should be fixed on the support that we want to make vibrating (preferably made of wood to provide a better medium to carry the signal). These



(a) The NormalTouch haptic device providing cutaneous sensation on the fingers by means of force feedback and vibrations.



(b) The Haptic Revolver providing texture sensation on geometrical objects explored by the user.

Figure 4.7 — NormalTouch [17] (top) and Haptic Revolver [178] (bottom) devices provide tactile-kinesthetic perception into immersive virtual environments.

devices are often used in cinemas or other audio-visual installations. It is this kind of device that we used in the experiments described hereafter in this manuscript.

More recently, technologies has emerged to create tactile sensations in the air using ultrasounds [84]. This kind of feedback is called mid-air ultrasonic haptic feedback [183]. The best known of these devices is Ultrahaptics [32]. This is a matrix of 12×12 or more ultrasonic transmitters. The physical principle involved, which is quite complex, is based on the fact that when ultrasound signals of high intensity intersects with fine adjustments, a cutaneous sensation of tingle can be felt at the location of these intersection points. With the help of this type haptic displays, 3D objects can then be haptified in the air. However, the technology remains not mature enough to feel vibrations with sufficient intensity [184, 37]. Nevertheless, this could be used for particular purposes. For example, it could be used to create affective computing since humans also communicate emotions through touch [122]. We also can quote tactile feedback by electrovibration. Introduced in 2010, TeslaTouch [16] is based on this principle to add skin sensations at the end of

fingers on touch surfaces. This device allows to haptify many different textures regarding what is displaying on the touch screen (wood, metal or whatsoever). Finally, we can also find in the literature devices that create true or illusory tactile sensations with the help of the suction effect [73], or with the help of elements that can be electrically contracted [107].

4.4.2 Tactile sensations to provide structured information: the concept of tactons

We have seen previously that structured information such as computer interfaces can be haptified at the fingertips [135]. Since then, different HCI concepts built on haptic (cutaneous) sensations to provide structured information to users have been made (see also the following section). We will not make an exhaustive list here, however we can mention what we call *haptic icons* (or *tactile icons*) and the family of underlying concepts. The concept of “icon” [156] is generally useful to vehicle the idea of *structured information*. Using the right design, structures or categories of information can be provided to the user. It can be done through the visual sense using the well-known *icons*, through hearing using *auditory icons* and *earcons*, as we saw in the previous chapter on sonification (Section 3.3), and even through our thermoception via *thermal icons* [182]. For the sense of touch, MacLean defined *haptic icons* in 2003 as “*brief computer-generated signals, displayed to a user through force or tactile feedback to convey information such as event notification, identity, content or state [...] Synthetic icons may be experienced passively or actively*” [105]. They can be designed to build haptic language that convey interpersonal messages.

This concept was refined a year later with the concept of *tactons*, which have been introduced in 2004 by Brewster and Brown [25]. They are defined as “*structured, abstract messages that can be used to communicate complex concepts to users non-visually*”. The research shown that when visual icons are good to provide spatial information to the users, their tactile equivalent, namely tactons, are good to convey temporal information. Their design is quite similar to the one used for earcons because they are both made of sound signals. They are designed through the parameters of cutaneous (passive) perception, given in Table 4.1.

As for earcons, tactons can be of *one-element earcons* type, *compound* type when they are composed of multiple one-element tactons, *hierarchical* type when they are composed in a hierarchical way with a tree structure relying parent tactons to their children, or of *transformational* type when they are generated along multiple parameters. They should also be learned; but learning time is often quite short. Using the abovementioned design space, we considered the concept of tactons for the next chapters to provide information to the users (see Chapters 5, p. 93, and 8, p. 145).

4.4.3 Tactile sensations to alert users and to provide spatial information

Numerous studies and systems have been presented in recent years to communicate spatial information to users via vibrotactile feedback. For example, we can mention FeelTact [53] device. In the form of a bracelet, it was used to transmit information to its carrier using vibrations. A large number of vibratory patterns (i.e., tacton) could be created using a specific IDE, which provided a rich mean of dialog that can be useful for disabled people.

Table 4.1 — Guidelines with a set of parameters for the design of tactons [25].

Parameter	Designation	Definition
<i>Frequency</i>	T_1	Perceivable frequencies are in the range of 20 – 1000 Hz, with a maximum perception at 250 Hz approximately [70]. A change in amplitude lead to a change in frequency (as for audition), and not much than 9 different values can be haptified to optimize the perception.
<i>Amplitude</i>	T_2	Interactions with frequency parameter led re-searchers to use only one parameter related to amplitude, with a maximum of 28 dB [39].
<i>Waveform</i>	T_3	Corresponds to the timbre in audio perception; experiments suggested that our cutaneous perception of wave forms is limited. We cannot distinguish more differences than between a square wave, a sine wave and a triangle wave.
<i>Duration</i>	T_4	Stimuli lasting less than 0.1 second are perceived as little stings on the skin, whereas stimuli of longer duration are perceived as more fluid cutaneous sensations [71].
<i>Rhythm</i>	T_5	Refers to a combinations of stimuli with different duration (see parameter T_4).
<i>Body location</i>	T_6	When it is possible (when the transducers used permit it), different body locations can be considered for tactons. Fingers are the most sensitive parts, but when hands are required for other tasks, other locations such as back, the thigh or the abdomen can also be considered with good results [39].
<i>Spatiotemporal patterns</i>	T_7	Refers to a combination of T_5 and T_6 . Transducers arrays, for example, can be used to display spatially and temporally coordinated haptic patterns.

In the context of driving, several studies use vibrotactile patterns to indicate directions or obstacles to be avoided to the driver. In their study, Schwalk *et al.* [151] studied subjective and behavioral measurements regarding recognition, adequacy and workload of tactile patterns presented to participant through a driving seat. In the same optic, Petermeijer *et al.* shown in [130] that vibrotactile feedback provided in the driver seat can convey a take-over request (within highly automated driving task). They also shown that static vibrotactile stimuli bring to faster reaction times than dynamic ones. This study and their results are useful for our purpose because the authors postulate that the drivers are not focused on their visual sense since in the framework of the autonomous driving, studies have shown that they would be acceded to other tasks than the driving or the monitoring of the driving (such as eating, talking, listening to music, etc.). Hence, they considerate the case in which drivers are deprived of their visual (and auditory) sense, and, therefore, better able to “listen” to their other sensations (such as touch). Besides, participants seemed to not recognize directional cues when they are provided through their seat.

Dynamic vibrotactile feedback in the context of driving has been studied by Meng *et al.* [113], relying on the fact that auditory stimuli used to attract attention are more effective when they come to the user than when they leave. This principle seems to work also for tactile stimulations to alert of external events. We can also mention the work of Gray *et al.* [67], which suggests that the effectiveness of dynamic vibrotactile feedback depends on the link between the event and the alert and the direction of the stimulus. Still to alert the user, Jensen *et al.* built a customizable automotive steering wheel equipped with vibrotactile feedback to use when obstacle to avoid [89]. It was used in the aim of providing the user additional information concerning him/her driving environment. The results suggest that vibrotactile feedback provided in the steering wheel let the drivers avoiding 62 % more obstacles with a 10 meters increase in the reaction distance to the obstacle. Ho *et al.*, for their part, studied vibrotactile feedback to present spatial information to the driver (e.g., spatial obstacle to avoid), and showed that participants reacted faster to spatial stimuli [83].

There also have many studies in the aeronautical field. Van Erp *et al.* designed a system providing the pilots with vibrotactile feedback through a belt to give waypoint directions, and thus orientation cues [52, 168]. Another study investigate helicopter pilots to help them performing hover maneuvers [134].

4.5 Kinesthetic Interaction

We have previously defined the key words kinesthetic and its place within HCI field (see Figure 4.8). Kinesthetic interaction (KI) refers to engaging our body, or a part of, in a context of interaction with systems. This term was formally defined for the first time in 2008 by Fogtmann *et al.* [55], and is based on previous work, including the Gesture-based Interaction introduced in 1999 by Long *et al.* [104] that consists in considering the body as an input device for interaction, and the Full-Body Interaction [128]. Fogtmann also draws on the work of Dourish [48] and Klemmer [95] on *embodied cognition*, which is a model in which the representation of one’s own space



(a) The Sony DualShock 1 game controller used in early versions of PlayStation 1 console (1997). (b) The Sega vibratory kit which could be plugged into the Dreamcast game controller (2000). (c) The new Microsoft Kinect device, used to capture body's movements (2019).

Figure 4.8 — Examples of everyday life entertaining devices providing (a), (a) vibrotactile feedback and (c) kinesthetic interaction.

depends on our perceptual and motor sensations⁸. KI reveals the body's potential in interaction design by promoting natural interaction and motion-based interfaces. It is based on our kinesthetic sensations and proprioception, to create interactive frameworks improving user's implication and serendipity. A precise definition of the term KI is proposed in the study from Fogtmann *et al.*:

“KI works as a theoretical foundation and vocabulary for describing the body in motion and how it conditions our experience of the world in the interactions with and through interactive technologies [...] KI is when the body in motion experiences the world through interactive technologies.”

In their study [55], the authors identify three main axioms that should be considered when designing a kinesthetic interface, namely *physiology*, *kinesthetic experience* and *interactive technologies*. The physiological aspect involve users' consciousness of the position of their own body as well as their own movements in space. An interface shall therefore consider the place taken by the body and its movements and promote the perception of space. Kinaesthetic experience is directly related to some physiological aspects and refers to the user's understanding, consciously or not, of the kinesthetic aspects of the interface. The final aspects to consider concerns the interactive technologies used to provide kinesthetic interaction into the interface. These 3 themes, associated to seven parameters (see Table 4.2) define a framework to build kinesthetic interaction.

⁸This concept comes from social psychology. We can also talk about *embodiment* [43] or *sense of presence*. More precisely, there is a recent consensus in the cognitive science community that the perception of one's own body in space critically depends on multisensory integration [100]. Embodiment can be significantly increased in an immersive environment by considering sensations from the somatosensory system. In HCI, this concept could be considered, for example, to reduce lag in the display and then to improve the fidelity of the projected scene. Kinesthetic interaction could enhance this feeling as well, by permitting the user to act as natural as possible, and therefore, in an ecological perspective, enhance his feeling of immersion.

Table 4.2 — Guidelines with a set of parameters for the design of kinesthetic interactions [55].

Parameter	Designation	Definition
<i>Engagement</i>	K_1	This parameter describes a KI concept that can be easily memorized by users and that facilitates serendipity through their body movements.
<i>Sociability</i>	K_2	Describes the fact of considering one body among others in the case of a multi-user system that consider KI.
<i>Movability</i>	K_3	This parameter must be considered to be aware, within the design phase, of the fact that the body, during a KI with a system, can move freely or is spatially constrained.
<i>Explicit motivation</i>	K_4	This parameter means that the system must explicitly describe to users how to interact.
<i>Implicit motivation</i>	K_5	This parameter must be taken into account when the KI is well opened, without restriction on the movements themselves (not only spatial).
<i>Expressive meaning</i>	K_6	The expressivity parameter is well taken into account when the KI is perceptibly related to the result.
<i>Kinesthetic empathy</i>	K_7	Is where specific and controlled movement patterns are affected by the relation to other people and stimuli from the surrounding environment.

Wrap-up message

In this chapter we have committed to present the field of haptics in HCI. Haptic technologies are widely considered for interaction. They can be divided into 2 main categories, namely *haptic interfaces* when it comes to touch, and *kinesthetic interfaces* when it comes to body's movements. They provide a more natural framework for communication, bringing a tangible dimension through the senses of touch and proprioception. They also contribute to immersion by promoting sensory stimulation. Actually, our everyday sensations tell us a great deal about our environment and it would be a shame to deprive ourselves of it in the context of interactions. They can sometimes increase the overall performance of users by helping to focus on their task and thus increasing their embodiment, or by taking advantage of the fact that haptic reflexes, for example, are shorter than visual ones. Moreover, when the visual sense is defective, they represent a good alternative to overcome such a lack and still transmit useful information to the user, particularly via tactile interfaces.

Researchers have been interested in the sense of touch for a relatively long time, the first ideas dating from the mid-1900 century, and the first “true” HCI studies dating back to the early 1970s. Since then, they have contributed to areas as diverse as scientific visualization, surgery, medical imaging, geology, physics, space construction, military, automotive and video games, among others. Their contribution is very often appreciated by experts, and application fields are almost as varied as there are sub-domains in HCI: data mining, affective computing, XR, orientation, language, Information Visualization, to name a few. Haptic technologies that are used to achieve this are also numerous: vibrotactile feedback using mechanical vibrations, air pressure or ultrasounds, force feedback, kinesthetic interaction, tangible or deformable interfaces, etc.

Previously, we have seen that there are 3 types of perceptions relating to the sense of touch: cutaneous or passive perception, tactile-kinesthetic or active perception, and haptic perception. Current haptic interfaces mainly involve the latter, which integrates the notions of passive cutaneous perception, voluntary (therefore active) exploration, as well as kinesthetic perception and proprioception (consciousness of one's own body and of his own movements). However, the principle of kinetic exploration is an interesting option to immerse the user. In a synthetic setting like that of remote control towers, this could contribute to embodied cognition by favoring a “natural” gesture. Moreover, in a wish to alert the user, as may be the case in aeronautics and more specifically ATC, passive perception shall also be considered. For this option, tactons represent a useful framework because they allow to haptified structured information.

This leads us to answer the question of the form that haptic technologies could take in RVT context, having in mind to enhance the overall ATCo's experience, and performance. We also have to identify the situations in which such technologies could be use. This will be the topic of the next part.

PART II

AUDIO FOCUS MODALITY: INVOLVING AUDIFICATION, PASSIVE PERCEPTION AND KINESTHETIC INTERACTION

Chapter 5

Designing Audio Focus interaction modality

Résumé (Français)

Première campagne expérimentale (Conception de la modalité interactive Audio Focus)

Méthode de conception

Grâce à une approche centrée sur l'utilisateur, nous avons pu à la fois définir un besoin et proposer des solutions avec l'aide d'experts du domaine, impliqués dans le processus de conception. Cette approche a été divisée en trois phases principales décrites dans la Figure 5.1 ci-après. Nous avons tenté d'identifier un cas d'utilisation présentant un intérêt particulier pour les professionnels. Dans ce but, nous avons été aidés par deux contrôleurs aériens. Le premier travaille comme contrôleur d'approche à l'aéroport de Roissy-Charles de Gaulle tout en étant également instructeur à l'ENAC au sein des parcours d'apprentissage du contrôle aérien. Le second, quant à lui, travaille comme contrôleur d'approche à l'aéroport de Muret et, également, comme pilote instructeur de voltige. Le but de la première phase du processus de conception fut de discuter de manière informelle, lors de réunions, avec ces contrôleurs pour tenter de comprendre les généralités de leur tâche de travail quotidienne. Parallèlement à ces réunions, nous avons par ailleurs suivi une formation d'une semaine d'initiation au contrôle aérien à l'ENAC, au cours de laquelle nous avons eu la possibilité d'aborder la théorie relative aux différents types d'ATC et de l'appliquer au travers de divers travaux pratiques. Après plusieurs réunions, il est apparu que l'une des difficultés récurrentes des contrôles d'approche et d'aéroport est la localisation des avions. Dans certaines conditions, notamment lorsque la visibilité est mauvaise, leur détection s'avère difficile, voire impossible si l'on ne s'appuie que sur la vision pour les localiser. Dans de telles situations, l'Interaction Homme-Systèmes peut être en mesure de fournir un confort supplémentaire aux opérateurs afin de préserver leurs habitudes et ainsi maintenir un niveau de sécurité acceptable dans les tours de contrôle déportées.

Ce contact visuel, de manière naturelle, est d'une importance non négligeable pour les contrôleurs aériens dans la mesure où il contribue à leur représentation mentale du terrain contrôlé,

tout en fournissant une information plus lisible, instantanée, par rapport aux instruments auxquels ils ont accès (les radars, notamment). Au cours de la deuxième phase du processus, après avoir sélectionné cette situation critique, nous nous sommes attachés à formaliser plus en détails ce que cela signifiait pour les contrôleurs aériens. Les situations où la visibilité est mauvaise ou nulle sont courantes dans les tours de contrôle. Il semble également que ce genre de situation soit relativement problématiques pour les contrôleurs. Dans une tour physique, les contrôleurs utilisent leur vision pour rechercher visuellement les aéronefs avec lesquels ils sont en communication. Cela leur permet d'obtenir une mesure subjective plus précise de leur position dans la zone contrôlée, mais également d'anticiper leurs temps de passage. Dans le cas d'un brouillard épais qui empêcherait un contact visuel avec les avions, ou simplement en cas de perte du signal vidéo dans un environnement de tour de contrôle déporté, les contrôleurs aériens avec lesquels nous avons travaillé ont trouvé utile d'avoir un outil qui leur permettrait d'agir comme (ou approximativement comme) s'ils disposaient d'une visibilité sur l'environnement de l'aéroport. Le principal cas d'utilisation que nous avons considéré pour notre étude a été, par conséquent, la perte de visibilité sur la zone contrôlée.

Observations

Plus précisément, les discussions qui ont eu lieu avec les contrôleurs impliqués dans le projet lors des différentes réunions nous ont principalement conduits aux trois constatations suivantes :

- Les contrôleurs aériens utilisent souvent leur corps au sein de leur espace de travail : lorsque les instruments ne sont plus suffisants, pour des raisons de précision ou pour se faire une représentation plus fidèle de la situation de contrôle, ils se lèvent de leur siège et se rapprochent des fenêtres de la tour dans le but de scanner l'environnement aéroportuaire à la recherche d'avions ou autres objets.
- Ils utilisent souvent des éléments du paysage pour améliorer leur représentation mentale de l'emplacement des aéronefs. Cela exige une connaissance approfondie des environs de l'aéroport (c'est en partie la raison pour laquelle ils sont formés en fonction du terrain auquel ils sont affectés). Par exemple, dans le cas de l'aéroport de Muret, l'un des contrôleur impliqué dans le processus de conception nous a expliqué que lorsqu'il obtient un contact visuel avec un avion entrant dans le circuit de piste, souvent, il projette mentalement sa position au sol pour ensuite comparer ce point avec des éléments connus du paysage et ainsi être en mesure d'évaluer de manière plus précise sa distance à la tour.
- En cas de panne d'équipement ou de perte de visibilité, les contrôleurs aériens utilisent des procédures prédéfinies en partie basées sur les observations précédentes, c'est-à-dire sur leur connaissance du paysage et leur propre perception de l'environnement contrôlé. Par exemple, lorsqu'un aéronef en vol à vue ne peut plus émettre de message radio, le contrôleur en charge peut demander au pilote d'effectuer un passage bas au-dessus de la tour pour qu'il puisse ainsi être entendu et par conséquent localisé avec une plus grande justesse.

Améliorer la détection d'éléments spatialisés en situation de non visibilité

La solution que nous avons imaginé pour palier au problème de non visibilité est une modalité d'interaction que l'on a nommé *Audio Focus*. Celle-ci est basée sur les concepts précédemment introduits d'audification, d'interaction kinesthésique et de stimulation vibrotactile (lorsqu'elle est couplée à des vibrations). Elle permet aux contrôleurs aériens de construire leur propre représentation mentale du voisinage aéroportuaire en se basant sur l'audio et non la vision. Plus précisément, cette modalité interactive permet d'évaluer la position spatiale des aéronefs à proximité de l'aéroport en effectuant des mouvements de tête faisant varier le niveau sonore des sources sonores leur étant associées. Étant donné que la vision n'est plus stimulée, l'ouïe prend ainsi le relais et devient le canal sensoriel sur lequel les contrôleurs se basent dans leur évaluation de leur environnement.

Son principe repose premièrement sur la spatialisation de sons de moteur d'avions conventionnels, reliés aux positions des avions à proximité de l'aéroport. L'orientation et la position de la tête de l'utilisateur est récupérée à l'aide du périphérique HoloLens de Microsoft.¹ Lorsqu'une source sonore (i.e., un avion) est alignée avec la tête de l'utilisateur, alors le volume sonore lui étant associé est augmenté, tandis que le volume des sources sonores étant situées en dehors de cet axe est diminué. La distance entre l'avion concerné et le point de vue de l'utilisateur dans la tour est également matérialisée par le gain associés aux sources sonores (plus un avion est proche de la tour de contrôle, plus le gain de la source sonore lui étant associée est élevé). Le but final de cette modalité d'interaction est d'aider les contrôleurs à localiser les avions dans un environnement en 3 dimensions. Ainsi, la modalité *Audio Focus* favorise le filtrage spatial” [6] car les niveaux sonores relatifs des sources sonores sur et hors de l'axe sagittal de l'utilisateur sont ajustés afin de permettre aux participants de “jouer” avec les sons perçus. Par conséquent, si un son semble beaucoup plus fort que les autres, cela signifie qu'un avion se trouve devant l'utilisateur (voir Figure 5.3) : les mouvements des utilisateurs ont une incidence sur le volume sonore des avions. Par conséquent, une telle modalité d'interaction favorise la sérendipité : les utilisateurs ont la possibilité d'augmenter leurs chances de localiser un avion dont ils ne connaissent pas la position à l'avance. Le Tableau 5.1 explique comment les différents paramètres de l'interaction kinesthésique (voir Chapitre 4, p. 71) ont été pris en compte dans la conception de cette modalité.

Nos aptitudes à localiser les sons le long de l'axe horizontal sont plutôt bonnes [115], et la modalité *Audio Focus* repose sur cette donnée. Avec la technique présentée précédemment, nous proposons aux utilisateurs un moyen de détecter les aéronefs sur le plan horizontal, cependant rien n'est fait pour leur donner des indications quant à leur position verticale. Les technologies haptiques, plus précisément la stimulation vibrotactile, ont été considérées et ajoutées à la modalité *Audio Focus* à cette fin. Des stimulations vibrotactiles sont utilisées pour la soutenir, et ne sont activées que pour les sources sonores alignées avec la tête des utilisateurs. L'objectif est d'augmenter la sélectivité verticale en ajoutant des informations leur notifiant si l'aéronef avec lequel ils sont alignés est situé dans les airs ou au sol. Tout comme [159] et [130], les vibrations sont ici con-

¹Le champ magnétique induit par les transducteurs utilisés pour les stimulations vibrotactiles – voir ci-après – déréglait les stations inertielles que nous avons testées; nous avons donc été obligés d'utiliser un autre périphérique et il s'est avéré que l'HoloLens était insensible à ces interférences magnétiques

sidérées dans le but de décharger le canal visuel en déportant l'information spatiale vers le canal sensoriel du toucher (perception passive). Pour réaliser ces vibrations, deux transducteurs vibrotactiles ont été fixés sur une chaise en bois (voir Figure 6.3, p. 118). Le premier a été fixé sous l'assise de la chaise, et le second derrière son dossier. Si la tête de l'utilisateur est orientée vers un avion, le premier vibre si cet avion est situé au sol; logiquement, le second vibre si l'avion est situé dans les airs. Les patterns vibratoires employés ont été conçus en tenant compte des paramètres fournis par Brewster et Brown dans [25]. La Table 5.2 reprend les choix qui ont été faits pour chacun de ces paramètres.

5.1 Introduction

In Parts II and III of this manuscript, we present in detail how we have used some of the concepts introduced in the previous Part I – [Theoretical backgrounds](#) (p. 35) to develop HCI solutions that could be adapted to the practice of remote airport ATC. Therefore, after having introduced some concepts for the design of non-visual interaction, namely sonification techniques (Chapter 3, Section 3.3, p. 60), the consideration of tactile stimulation in HCI (Chapter 4, Section 4.4, p. 82) and kinesthetic interaction (Chapter 4, Section 4.5, p. 86), we began to build on this knowledge to create new interactive techniques adapted to ATC and to the ATCos' working task. Two experimental campaigns were planned within the project (see introduction Chapter). The present part concerns the first campaign. We present Audio Focus (a modality tested during the first experimental campaign) (AF), an interaction based on sound spatialization, audification, proprioception and passive perception. Designed based on a user-centered study, it addresses a specific need to improve the ATCos' experience in remote tower environment.

Some situations in which visual contact with aircraft appears to be impossible are of particular interest to ATCos. If the control conditions are met (see Chapter 1, Section 1.2.4, p. 20), they must manage the air configuration on the controlled airport without being able to rely on their vision. In such situations, HCI seems to be able to provide additional comfort to operators in order to maintain an acceptable level of safety in remote tower environments. Using a user-centered approach, we were able to both understand and define the need, and propose solutions by involving a sample of end users in the process of designing new interaction modalities. This chapter describes this approach (Section 5.2) and formulates working hypothesis that we rely on during the design phase (Section 5.4). Then, we explain how we followed the guidelines mentioned previously (Chapters 3 and 4) to design a new interaction modality based on hearing and touch (Section 5.5). We finally provide a wrap-up message at the end of the chapter (5.5.2).

5.2 Design approach

Based on the idea that end-users (in our case, ATCos) are in the best position to help us to design, evaluate and use these new interaction modalities, we involved some of them in our design process and during the two experimental campaigns. This process was divided into 3 main phases, as described in Figure 5.1. Following several discussions and interviews with professional ATCos through an iterative user-centered process, it appeared that one of the recurring problems in airport ATC is the visual location of aircraft. In particular conditions, especially when visibility is poor, aircraft detection could be difficult or even impossible using only vision.

To design these modalities, we first looked at situations of particular interest, both for controllers themselves and for the specific context of Remote Control. The first ATCo involved in the design process was working as an approach ATCo at Roissy-Charles de Gaulle airport while also occupying a position of instructor at ENAC. The second was working as an approach ATCo at Muret airport and is, as well, an aerobatic instructor pilot. We conducted several meetings and discussions with these two professionals to try to understand the generalities of the Air Traffic Controllers' working task. This has taken the form of several meetings and informal discussions.

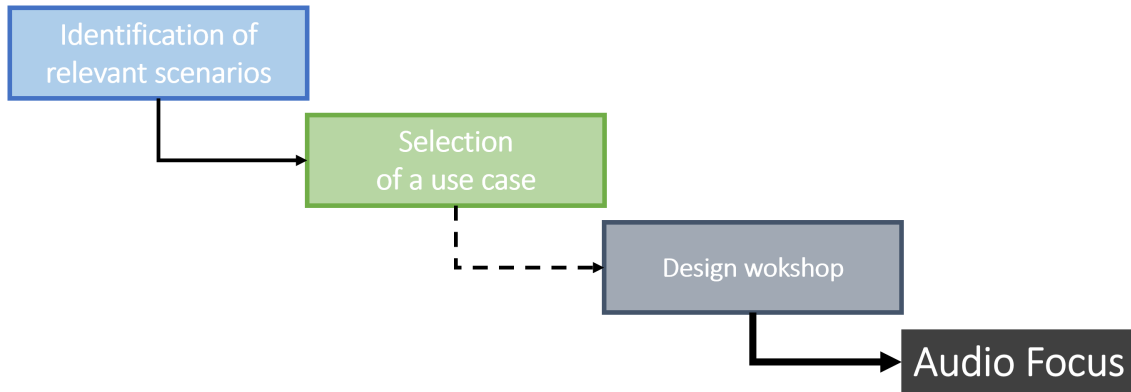


Figure 5.1 — The design process we followed for the design of Audio Focus interaction modality.

In parallel to these meetings, we also followed a one-week training course on ATC offered by the ENAC, in which we had the possibility to discuss the theory of the different types of ATC (see Section 1.2.3, p. 19), and to get into shape through realistic practical work. This has allowed us to have a much more precise vision of the working tasks and to understand the essential elements as well as the potentially risky situations and some of the needs of operators.

During the second phase, we selected a critical situation and formalized in more detail what this meant for ATCos. Situations of poor or with no visibility are likely to be common in tower ATC. It also appeared that this kind of situation seems to be problematic for ATCos. Of course, in tower ATC, ATCos eventually use their vision to physically search for aircraft they are in communication with. This allows them to have a better idea of their positions within the controlled area. Also, it allows them to anticipate passage times of aircraft and, thus, improve their mental representation of the current control configuration. In the case of heavy fog, which prevents such visual information on aircraft from being available, or simply in the case of loss of video signal in remote control tower environment, the ATCos with whom we worked found it useful to have a tool that could allow them to act as, or at least approximately as, if they were in good visibility conditions. Hence, the principal use case we considered for our study was the lack of visibility on the controlled area.

5.3 Specific observations resulting from the user-centered approach

More precisely, the discussions that were held with the ATCos involved in the project during the various meetings defining the 3 phases of the design process explained above led us to the 3 following main observations:

- Tower ATCos often use their bodies in their work: when instruments are no longer sufficient, for precision reasons or to have a more *embodied* representation of the current situation, they get up from their seats to get closer to the windows and then start to scan the airport environment in search of aircraft or other objects.

- They often use landscape elements to improve their mental representation of aircraft location. This requires an advanced knowledge of the airport vicinity (which is why ATCos are trained according to the terrain they are assigned to). For example, in the case of Muret airport, one of the ATCo involved in the design process explained to us that when an aircraft enters the runway circuit and a visual contact with it is possible, he often mentally projects its position on the ground and then compare this resulting point to known elements of the landscape to better evaluate its distance from the tower.
- In the case of an equipment breakdown or loss of visibility, they use predefined procedures based on the 2 previous observations, i.e., using their own embodied perception of their surroundings and their knowledge about the airport vicinity. For example, when a VFR aircraft can no longer transmit through radio means, the ATCo may instruct the pilot to make a low pass over the tower so that it can be heard, and therefore located.

Hence, we have identified specific situations in which ATCos use methods that can be described as *unconventional*, in which they involve their knowledge and perception in a broad way. Still with the help of the ATCos involved in the design process, we were able to identify a particular use case, namely the non-visibility conditions. Since ATCos rely mostly on sight, and in this particular case they are deprived of it, they no longer have any concrete means to apply this kind of unconventional methods. This is a particular situation in which HCI can improve their comfort at work, by offering them a way to use the processes to which they are accustomed, even in the event of a degraded situation, e.g., in the event of a lack of visibility on controlled aircraft and more generally in the controlled area.

5.4 Informal hypotheses

In fact, 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 goniometer². In addition, some low-traffic areas are not equipped with radars. As discussed before (see Chapter 2, Section 2.2.2.4, p. 44), 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 to 2 degrees [115]. However, one significant type of location error in space occurs when sound sources are almost aligned with the same azimuth (see the *cone of confusion* effect schematized in Figure 2.2, p. 44). The interaction modality presented here has been designed taking these factors into account in order to assist the search for aircraft localization in poor visibility conditions. We used a sound interaction based on hearing and touch channels as information vectors to enhance this selection process. In degraded visibility conditions, the contribution could be fourfold:

²A goniometer is a device or sensor used to measure angles. Historically, it has been the basic instrument in tower ATC for representing the position of controlled aircraft in the airport vicinity.

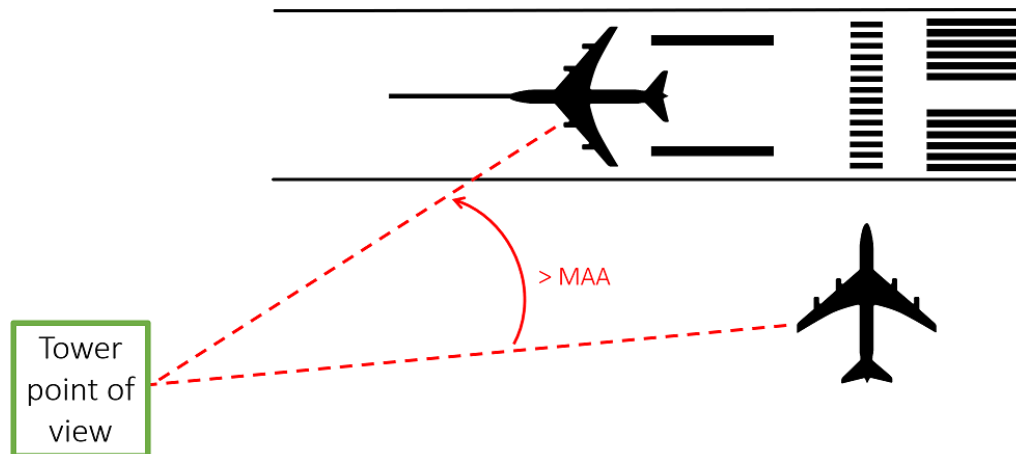


Figure 5.2 — Audio Focus interaction modality is based on the good horizontal resolution of the human ear (MAA) to distinguish between 2 sound sources located on the same horizontal plane.

- to reproduce sounds that exist in real towers using spatial sound sources, and then
- to enhance the users' immersion by slightly increasing their sound levels (compared to a real control tower),
- while trying to increase ATCos' performance using new interaction techniques acting on these spatial sound sources;
- in addition, this could decrease visual channel bandwidth, which is often overloaded, especially for those professions generating a high mental load such as ATC [159, 112].

These 4 points were our informal working hypothesis for Audio Focus modality during its design process.

5.5 Improving spatial sound source location in poor visibility conditions

In this section we present the interaction and feedback techniques which have been designed to overcome this visibility issue. As said in [22], our work might belong to the category of “enactive interfaces”, since they are part of “*those that help 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 Audio Focus interaction modality asks the user to do. Therefore, we rely on the design principles defined in [55] and mentioned in Chapter 4 of this document (Section 4.5, p. 86) concerning KI (see K_i parameters), as well as the audification principle in Chapter 3 (Section 3.3.1, p. 60), and tacton-related design principles set out by Brewster and Brown [25] and mentioned in Chapter 4 (see T_i parameters, Section 4.4.2, p. 84)

5.5.1 Auditory channel and proprioception

As seen previously, we overcome the visibility issue by designing Audio Focus (AF), which is based on KI and spatial sound sources. These latter are engine sounds coming from small types of aircraft. They are linked to each static position of aircraft in the airport vicinity. Users head orientation and position have been retrieved using a Microsoft HoloLens device³. 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 [170], and 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 ATCos to locate the related aircraft in a 3-dimensional environment. The AF interaction modality can be qualified with “spatial filtering” [6] as 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. Therefore, if a sound appears to be much louder than the other ones, it means that an aircraft is in front of them (Figure 5.3): users’ movements act on the sound. In this way, we could qualify this treatment as a simple *audification* treatment. Besides, the term “act” is here used to designate information provided by the user to the system (act of doing). In this way, AF interaction modality enhance the level of *serendipity*: by using it, users increase their chances of locating an aircraft for which they do not know in advance its position. Table 5.1 resume how KI guidelines set out by [55] were considered for the design of this modality.

In their systematic state of the art on sonification, Dubus and Bresin analyzed that sound, because of its intrinsically time-related nature, is a well-suited medium to communicate information for time-related task such as monitoring or synchronization [49]. Hence, it appears that sound is well adapted to searching tasks using KI guidelines, which are inherently linked to time. This modality provides ATCos with means of constructing a mental representation of the airport vicinity. More precisely, it makes it possible, when there is no visual cue, to assess the spatial position of aircraft in the airport vicinity by making head movements to audibly search for them instead of trying to have an impossible visual contact. A sound source playing a generic aircraft engine sound is assigned to each aircraft in the airport vicinity⁴, and spatially positioned according to the actual position of the related aircraft. Since vision is no longer stimulated, hearing takes over and becomes the sensory channel on which ATCos rely on. AF principle is similar to the one proposed by Savidis *et al.* [150], and is based on the strong correlation between head position and visual attention [162], as well as on the increase in the sound level associated to aircraft. When the visibility is poor and does not allows the ATCos to see the aircraft, they can move their head to make the sound volume associated to each aircraft varying.

³Magnetic field induced by the tactile transducers used for the vibrotactile patterns (see Chapter 6 Section 6.2.8, p. 116) was powerful enough to make every inertial unit out of order. We then chose to use the HoloLens device because it was the only one available to us and is not sensitive to this type of magnetic field.

⁴It should be noted that this type of sound cannot normally be heard in a physical control tower; it is therefore an augmentation.

Table 5.1 — First experimental campaign: How KI design space [55] was considered for the design of Audio Focus modality (see reference Table 4.2).

Parameter	How it is applied
K_1 (engagement)	<i>The body language is natural and therefore the interaction is easily memorable, since the principle of AF interaction only asks the users to move their head to search for aircraft.</i>
K_2 (sociability)	<i>Remote Control Tower environments (single and multiple) are managed by a single operator, so this parameter could not be considered for AF interaction.</i>
K_3 (movability)	<i>The head movements required for AF interaction are completely free; they are only constrained by the natural movements of the head.</i>
K_4 (explicit motivation)	<i>We will see later in this document that users had a training phase before starting the experiment to learn the different interaction modalities. However, AF interaction requires only natural movements: from the moment the users understand the direct link that exists between the increase in sound levels and their own head movements, the interaction can be qualified as learned.</i>
K_5 (implicit motivation)	<i>Users can do very small movements to accurately appreciate the location of aircraft. They can also move their head forward or backward to perceive relative distances more precisely.</i>
K_6 (expressive meaning)	<i>AF interaction have a good reactivity. Actually, head movements and related sound level increases are well synchronized (this synchronization seems to be in the order of 10 milliseconds range and no user seemed embarrassed with a potential lag; none of them alerted us to a potential desynchronization in the sound increase).</i>
K_7 (kinesthetic empathy)	<i>For the same reason than for K_2 parameter, K_7 could not be taken into account for AF interaction.</i>

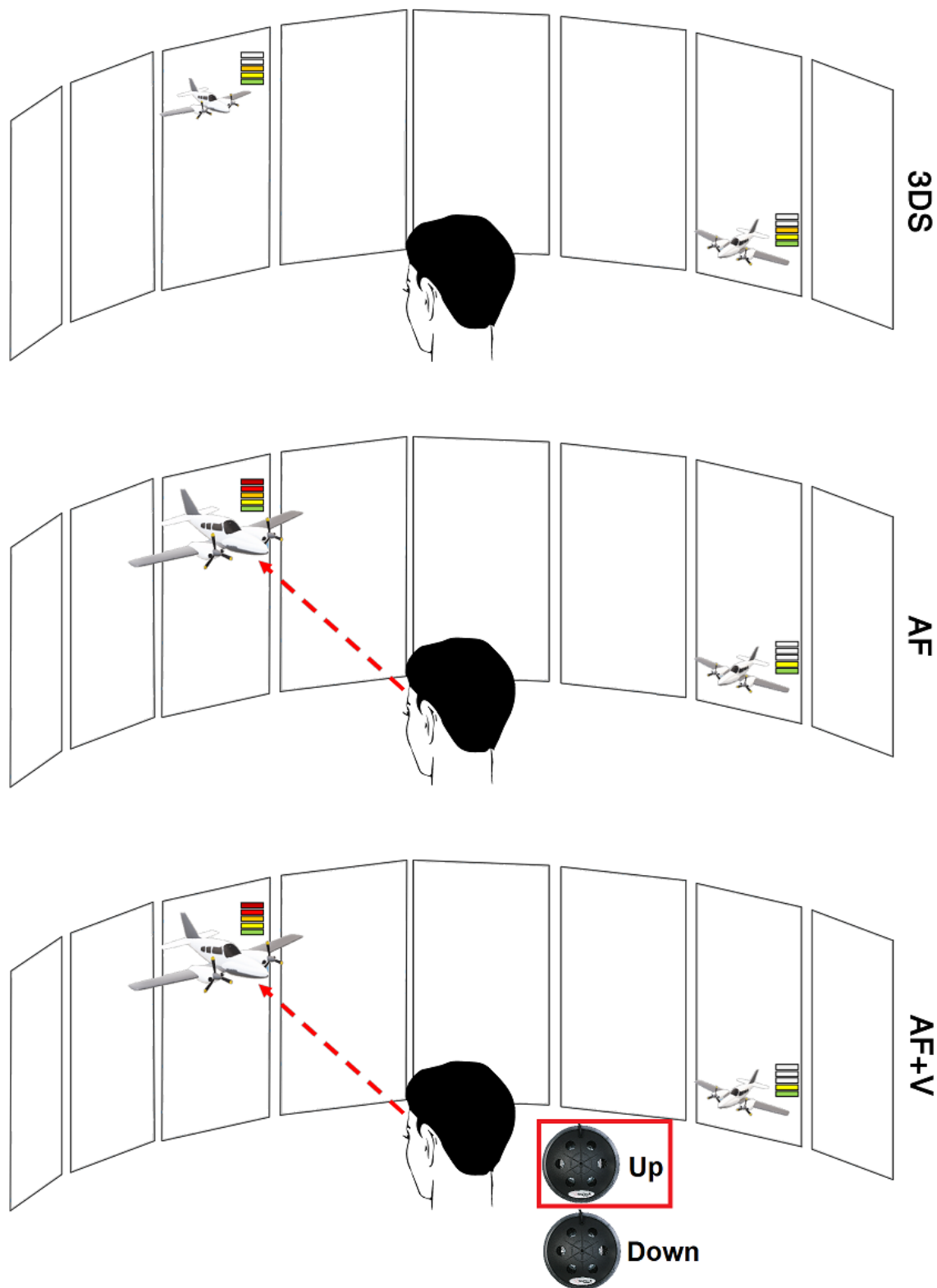


Figure 5.3 — Audio Focus interaction modality design principle — On top, AF modality is not activated; the sound sources associated to aircraft are not amplified. In the middle, AF interaction modality is activated: by moving their head, users amplify the sound sources which are aligned with their sagittal axis. At the bottom, AF interaction modality is coupled with vibrotactile feedback: when an aircraft is aligned with the users' head sagittal axis, the corresponding transducer vibrates (behind the back of the chair for aircraft located in the air, under the sit for aircraft located on ground).

5.5.2 Touch channel: passive perception through vibrotactile feedback

Humans are good at locating sounds along the horizontal axis [115], and AF interaction relies on this aptitude. With the technique presented before, we provide the users with a mean to detect aircraft on the horizontal plane, but they have no clue to resolve their vertical position. Therefore, haptics, and more precisely vibrotactile feedback, has been added to AF modality for that purpose. Vibrotactile feedback are 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 users if the aircraft they are currently aiming at is located in the air or on the ground. In the same way as [159] and [130], the vibrations are here considered for the purpose of unloading the visual sensory channel by presenting spatial information through the sensory channel of touch via passive perception.

To present this feedback to the users, 2 vibrotactile transducers have been fixed on a wooden chair (see Figure 6.3, p. 118): the first one was fixed under the seat of the chair, and the other one was fixed behind its back. Providing the user's head is oriented toward an aircraft, the first vibrates if this aircraft is located on the ground (which is materialized by the *Down* modality), and the second vibrates if the aircraft is located in the air (*Up* modality). Audio Focus coupled with Vibrotactile feedback (a modality tested during the first experimental campaign) (AF+V) interaction gives the system the same input as AF interaction (e.g., users' head orientation), however it gives the users another input (e.g., amplification of sounds, as for AF interaction, but here coupled with vibrotactile feedback). These vibratory patterns have been designed considering the guidelines provided by Brewster and Brown in [25]. The Table 5.2 resume the choices we made for each parameter to consider.

Wrap-up message

In this chapter we have exposed the method we followed to isolate a need, formulate a solution with the help of the professionals for whom it was intended for, and then design it. To do this, we have relied on concepts and methods formulated by different authors and explained in the previous chapters. We have thus characterized kinesthetic aspects through subjective parameters such as engagement, sociability, movability, motivation, expressivity and empathy. The haptic aspects, more precisely those related to passive perception, were expressed using more formal parameters such as frequency, amplitude, waveform, duration, rhythm and body location.

The AF interaction modality has therefore been designed through various concepts such as audification, kinesthetic interaction and tactons, and integrates these three aspects, which have been developed in detail in this chapter. The following chapter describes the strategy we envisaged to evaluate this new interaction modality, as well as the resulting data and their analysis.

Table 5.2 — First experimental campaign: How tactions design space [25] was considered for the design of Audio Focus modality (see reference Table 4.1).

Parameter	How it is applied
T_1 (frequency)	<i>Vibrations are made using tactile transducers (see Section 6.2.8), which are almost infra-bass speakers. For this reason, we considered very low frequencies, and the sound signal were at 55 Hz (for the Up, and Down modalities).</i>
T_2 (amplitude)	<i>Vibrotactile patterns are normalized at 0 dB, however the signal was adjusted on-demand for each participant during the evaluation phase by turning up or down the volume button of each transducers.</i>
T_3 (waveform)	<i>For maximum perception through our device, after several tests, the smoothed square waveform has been selected.</i>
T_4 (duration)	<i>The vibrotactile patterns used for Up and Down modalities last 40 ms, and are looped while the user's head is aligned with the concerned aircraft.</i>
T_5 (rhythm)	<i>Vibrotactile pattern used for the Up modality is composed of a group of 10 ms smoothed square waves clocked at 55 Hz (A musical note) followed by 10 ms of silence; this couple of two blocks is then repeated one time. Vibrotactile pattern used for the Down modality is composed of a group of 20 ms smoothed square waves clocked at 55 Hz followed by 20 ms of silence (see Figure 5.4).</i>
T_6 (body location)	<i>Vibrotactile pattern used for the Up modality is located in the back of the user (more precisely just below the shoulder blades, but this may vary from one user to another, depending on their size), and the one used for the Down modality is located under the seat (more specifically at the thighs).</i>
T_7 (spatiotemporal patterns)	<i>The vibrotactile patterns used for AF interaction modality is represented along only one spatial dimension; hence, T_7 parameter is expressed through T_5.</i>

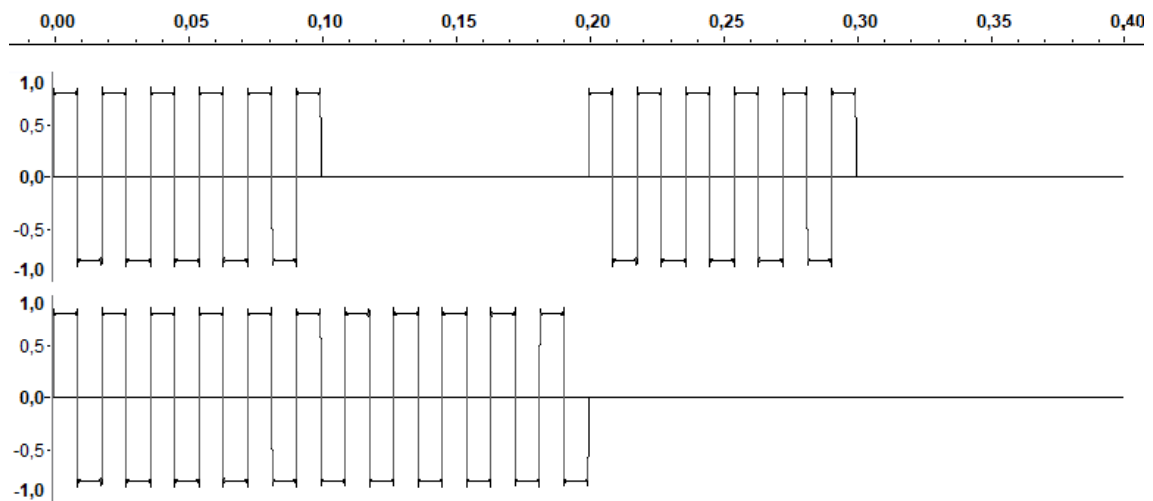


Figure 5.4 — Audio signals used for creating tactons that are involved in AF interaction modality. Top signal is used for the *Up* modality and bottom signal is used for the *Down* modality.

Chapter 6

First experimental campaign: Evaluation of Audio Focus interaction modality

Résumé (Français)

Première campagne expérimentale (Évaluation d'Audio Focus)

Participants et tâche expérimentale

Vingt-deux contrôleurs aériens volontaires de nationalité Française et exerçant dans des aéroports différents ont participé à l'expérience (8 femmes et 14 hommes). L'âge moyen était de 40,68 ans ($SD = 8$) et l'expérience professionnelle moyenne de 10,48 ($SD = 6,87$). Onze contrôleurs (4 femmes et 7 hommes) ont formé le groupe *A*, le reste a formé le groupe *B*. Comme tous les contrôleurs sont soumis à des tests médicaux dans le cadre de leurs profession, aucun d'entre eux n'avait de problèmes auditifs. En particulier, ils n'ont signalé aucune atténuation dans leur bande passante auditive ni aucun déséquilibre entre leur deux oreilles.

Les participants ont été invités à donner des informations sur leur perception de la localisation des avions à proximité de l'aéroport grâce à leur ouïe et à leur sens du toucher, dans 2 conditions de visibilité (bonne et mauvaise) et pour chacune des 3 modalités d'interaction testées (voir plus loin). Les avions étaient visibles dans de bonnes conditions de visibilité, mais ne l'étaient plus lorsque les conditions de visibilité étaient mauvaises.

Conditions expérimentales

Trois différents aspects ont été manipulés au cours de l'expérience : le type d'interaction, le niveau de difficulté et les conditions de visibilité. Une phase de pré-test a permis de quantifier ces conditions expérimentales. Le premier est le facteur *Modalité*, qui correspond au type d'interaction utilisé. Celui-ci pouvait prendre les 3 valeurs suivantes: *3DS* (son spatialisé seul), *AF* (modalité Audio Focus), ou bien *AF+V* (AF couplée à des vibrations). *3DS* est la valeur référence puisque les sons sont naturellement spatialisés dans une tour de contrôle physique. Dans un contexte de contrôle déporté, ces sons doivent être générés pour construire un environnement réaliste. Le nombre de sources sonores simultanées représente le facteur *Difficulté*, et peut prendre les valeurs : *Facile* (1 avion dans l'espace aéroportuaire contrôlé), *Moyen* (2 avions simultanés associés à 2

sources sonores différentes ayant des positions séparées dans l'espace aéroportuaire contrôlé), ou *Difficile* (3 avions simultanés associés à 3 sources sonores différentes ayant des positions séparées dans l'espace aéroportuaire contrôlé). Le dernier facteur, *Visibilité*, correspond aux conditions de visibilité, pouvant prendre les valeurs *Bonne* ou *Mauvaises*.

Hypothèses

Nous nous attendions à ce que les participants localisent les avions de manière plus précise lorsqu'ils utilisaient la modalité Audio Focus dans des conditions de visibilité mauvaise. Nous nous attendions également à ce que leur précision soit d'autant plus grande lorsqu'ils utilisaient la modalité Audio Focus couplée à des vibrations. Les temps de réaction n'ont pas été pris en compte pour ces hypothèses car la modalité Audio Focus nécessite des mouvements de la tête, ce qui la rend naturellement plus lente que le son spatial seul. Trois hypothèses de travail ont par conséquent pu être formulées, à savoir : "*les participants, placés dans des conditions de visibilité mauvaise, localisent les avions avec plus de précision...*

- $H_{1,1}$: AF Vs. 3DS \rightarrow ...lorsqu'ils utilisent la modalité AF, comparé à la modalité 3DS;
- $H_{1,2}$: AF+V Vs. 3DS \rightarrow ...lorsqu'ils utilisent la modalité AF+V, comparé à la modalité 3DS;
- $H_{1,3}$: AF+V Vs. AF \rightarrow ...lorsqu'ils utilisent la modalité AF+V, comparé à la modalité AF".

Discrimination de l'espace aéroportuaire, IHM de réponse et combinatoire

Pour cette campagne expérimentale, et compte tenu des simplifications statistiques, l'aéroport de Muret a été divisé en 5 zones distinctes, à savoir (voir Figure 6.1) : la zone de décollage (*Take-off*, en l'air au-dessus des pistes), de vent traversier (*Crosswind*, opposée aux pistes, en l'air en face de la tour), de base (en l'air, opposée à la zone de décollage par rapport aux pistes), les parties Est et Ouest de la piste de décollage (*Runway east* et *west*). Les sources sonores, correspondant aux avions à localiser ont été placées au centre de chacune de ces zones. Une zone ne pouvait contenir plus d'un avion à la fois. Le nombre total d'avions correspondait au niveau de difficulté.

Il y avait 3 possibilités pour la modalité utilisée (3DS, AF, AF+V), 2 possibilités pour les conditions de visibilité (Bonne, Mauvaise), et 3 niveaux de difficulté (Facile, Moyen, Difficile). Le nombre de combinaisons pour 5 zones aéroportuaires parmi 1, 2 ou 3 avions simultanés (niveaux de difficulté) est donné par les coefficients binomiaux C_k^M ¹ : $C_1^5 = 5$ (Facile), $C_2^5 = C_3^5 = 10$ (Moyen et Difficile). Cependant, certaines de ces combinaisons sont redondantes. Par exemple, pour un participant donné, il n'est pas obligatoire de tester la combinaison 1.2a et également la combinaison 1.2b en raison de leurs positions symétriques : la première est utile pour tester la perception d'un avion situé à gauche, alors que la seconde est utile pour tester la perception d'un avion situé

¹ k parmi N .

à droite . Si l'on considère que les participants n'ont pas de déséquilibre auditif, ces deux exemples sont équivalents. C'est pourquoi 2 séries d'épreuves ont été créées, dans lesquelles le même nombre de combinaisons par niveau de difficulté a été sélectionné de manière équilibrée :

- Groupe $A = \{ 1.1, 1.2.a, 1.2.b, 1.3.a, 1.3.b, 2.2, 2.3.a, 2.4.a, 2.5.a, 2.6.a, 3.2, 3.3.a, 3.4.a, 3.5.a, 3.6.a \}$;
- Groupe $B = \{ 1.1, 1.2.a, 1.2.b, 1.3.a, 1.3.b, 2.1, 2.3.b, 2.4.b, 2.5.b, 2.6.b, 3.1, 3.3.b, 3.4.b, 3.5.b, 3.6.b \}$.

De multiples combinaisons de positions statiques d'avions ont été affichées selon les 5 zones précédemment présentées. Tous les participants ont entendu les mêmes configurations, dans un ordre aléatoire pour chaque *Modalité*, niveau de *Difficulté* et condition de *Visibilité*. En cliquant sur la ou les zones correspondantes sur une interface spécialement créée à cet effet, les participants ont indiqué l'origine du ou des sons qu'ils entendaient (voir Figure 6.1).

Protocol expérimental

L'ensemble du protocole expérimental envisagé à été schématisé en détails en Figure 6.5. Après un texte de bienvenue (voir annexe 10.8), l'expérience a débuté par une phase d'apprentissage qui consistait en une présentation des stimuli dans chaque combinaison possible. Le but était d'habituer les participants à "jouer" avec les sons par l'intermédiaire de mouvements de leur tête, comme s'ils devaient chercher quelque chose. Cependant, comme expliqué précédemment, cette recherche se fait ici en utilisant l'ouïe au lieu de la vue. À la fin, toutes les combinaisons d'avions potentielles ont été présentées. Au besoin, les blocs d'entraînement pouvaient lancés une seconde fois. La formation était terminée lorsque les participants s'étaient familiarisés avec tous les cas expérimentaux. L'expérience à proprement parler pouvait ensuite commencer. Elle a été divisée en blocs correspondant à chaque modalité. Les 6 blocs de modalité ont été présentés 4 fois à chaque participant, avec une pause de 5 minutes au milieu de l'expérience. À la fin de l'expérience, les participants devaient remplir un questionnaire en ligne afin de donner leurs impressions sur la facilité d'utilisation, la fatigue ressentie, les performances et la charge de travail perçues.

Mesures

La performance des participants a été mesurée objectivement à l'aide de 2 variables comportementales différentes (variables dépendantes) : la *précision*, quantifiant le nombre de bonnes réponses, c'est-à-dire le nombre d'avions correctement localisés parmi les 5 possibilités (zones) offertes, et le *temps de réaction*, qui correspond au temps pris par les participants pour localiser le ou les avions en question, à partir du moment où une nouvelle combinaison leur était proposée, jusqu'au moment où le bouton Valider était pressé sur l'IHM de réponse.

Nous avons également effectué des mesures neurophysiologiques afin de disposer d'une autre mesure objective de la performance au travers d'un *indice de charge mentale*. La performance subjective, quant à elle, a été obtenue en demandant aux participants de remplir un ensemble de

Table 6.1 — Première campagne expérimentale : Synthèse des résultats obtenus lors de la première campagne expérimentale concernant la modalité d'interaction Audio Focus. La colonne de droite indique comment le symbole > doit être compris.

Variable	Résultat simplifié	Commentaire
<i>Précision</i>	AF+V > AF > 3DS	Plus précise
	AF+V > AF	Plus rapide
<i>Temps de réaction</i>	3DS > AF+V	Plus rapide
	3DS > AF	Plus rapide
<i>Utilisabilité</i>	AF+V > AF > 3DS	Plus facile à utiliser
<i>Fatigue</i>	AF+V > AF > 3DS	Engendre moins de fatigue
<i>NASA-TLX Demande mentale</i>	AF+V > 3DS	Charge mentale plus faible
<i>NASA-TLX Performance</i>	AF+V > AF > 3DS	Plus efficace
<i>NASA-TLX Effort</i>	AF+V > AF > 3DS	Moins fatiguant
<i>NASA-TLX Frustration</i>	AF+V > AF > 3DS	Moins frustrant
<i>Charge mentale (EEG)</i>	3DS > AF+V	Engendre moins de charge mentale

questionnaires en ligne, directement après la fin de l'expérience, et comprenant : deux questionnaires relatifs à la *Fatigue* perçue et à l'*Utilisabilité* des différentes modalités proposées, et un questionnaire de type NASA Task Load Index également pour chacune des modalités.

Résultats

Les résultats obtenus ont été rassemblés et synthétisés dans le tableau 6.1 ci-après. Il résume les principaux résultats qui ont été trouvés lors de cette campagne expérimentale, dans des conditions de visibilité mauvaises. Le message principal est que la modalité Audio Focus, associée ou non à une stimulation vibrotactile, en plus d'avoir été bien reçue par les participants, leur a fourni un moyen efficace pour localiser les avions lorsqu'ils n'avaient aucune visibilité sur l'aéroport. En fait, alors que les temps de réaction s'avèrent être légèrement plus lents lorsqu'ils utilisent cette modalité, leurs réponses sont beaucoup plus précises. Parallèlement, les résultats subjectifs nous ont donné un aperçu de la façon dont les participants ont reçu ce genre d'outil. Globalement, les modalités proposées ont été reçues de manière très positive, cependant, la charge mentale objective (donnée neurophysiologiques) s'est avérée significativement supérieure lorsque les modalités d'interaction étaient activées.

6.1 An experiment to evaluate Audio Focus modality

AF has been designed to facilitate the work of ATCos when they are immersed in a RCT environment. As explained previously (Chapter 5, p. 93), its principle is based on a natural behavior of ATCos when they seek for visual contact with aircraft they are currently managing. AF interaction modality uses this visual principle and adapts it in audio dimension. This was applied to the case of loss of visibility, which had been previously identified with the help of professionals of the field. Using this principle, when meteorological conditions do not allow visual contact with aircraft, or if, for example, the data link to the remote terrain is interrupted, AF interaction modality allows ATCos to create their own mental representation of the current control situation at the controlled airport by audio means. Relying on head movements, ATCos can apprehend the position of the different aircraft located around the runway by acting on the sound volume associated with each of them (see Chapter 5, Section 5.5, p. 100). In the same way as when they are searching for aircraft visually in a “normal” control environment, this HCI concept offers them the possibility to look up and make circular movements as if they had to search for aircraft with their eyes.

As part of the MOTO project, we were able to evaluate this interaction modality and thus quantify its contribution. This chapter describes our experimental strategy for this purpose. Section 6.2 describes the method used, namely the description of the group of participants, the detailed experimental protocol, the combinatorics of the trials, the description of the proposed experimental task, the equipment used and the different metrics used. Section 6.3 describes in detail the statistical analysis that were carried out on the collected data. Section 6.4, 6.5 and 6.6 respectively describes behavioral, subjective and neurophysiological results. Finally, a synthesis is proposed in section 6.6.

6.2 Method

In this section we describe in detail the experimental protocol we designed for the evaluation of Audio Focus interaction modality. First, we present the group of participants (Section 6.2.1), we relate important ethical considerations and give an overview on the experimental task (Section 6.2.2). Then, we explain how the experiment was designed (Section 6.2.3) and we set out our working hypotheses in a more formal way (Section 6.2.4). After that, the airport discrimination strategy is explained (Section 6.2.5), as well as a specific HMI used during the experiment to collect participants’ answers (Section 6.2.6), the metrics (Section 6.2.7) and apparatus (Section 6.2.8). Finally, we describes combinatorics aspects (Section 6.2.9) before finishing with a detailed presentation of the entire experimental protocol (Section 6.2.10).

6.2.1 Participants

Twenty-two volunteers French tower ATCos coming from different airports took part in the experiment (8 females and 14 males). The mean age was 40.68 years ($SD = 8$). Their professional experience was diversified 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 6.2.2). Eleven controllers

(4 females and 7 males) formed group *A*, and the remainder formed group *B* (see Section 6.2.9). 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.

6.2.2 Nature of the task and ethics consideration

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 (i.e., good and poor) and for each of the 3 interaction modalities tested (see next section). Aircraft were visible in good visibility conditions, while they were not in poor visibility conditions. Since the simulation asked the participants to perform an ATM-like task, few constraints have been imposed in order to avoid any confusing effect.

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. Consequently, as these data were anonymous, participants were also informed of the impossibility of destroying the related information in a targeted way (i.e., by giving first and last names) if they wanted to. All the participants signed a Consent Form (see Appendix 10.8) to make it clear that they agreed with the conditions of the experiment. A local ethical committee approved this experiment before its execution.²

6.2.3 Experimental conditions

Three different aspects were manipulated during the experiment: interaction type, difficulty level, and visibility conditions. A pre-test phase helped to quantify these experimental conditions (see Section 6.2.10). The first one is the *Modality* factor, which is the current feedback and/or interaction type and which can be (see Figure 5.3, p. 103):

- Spatial sound only (a modality tested during the first experimental campaign) (3DS) for the baseline,
- AF,
- AF+V.

3DS is considered as a baseline since sounds are naturally spatial 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:

²This committee is named CERNI which is incorporated into the Research House of the Federal University of Toulouse; see <https://www.univ-toulouse.fr/recherche-dynamique/ethique-et-integrite-scientifique> for more details.

- *Easy* with 1 engine sound,
- *Medium* with 2 simultaneous engine sounds coming from 2 different sound sources having separated positions in the airport vicinity), or,
- *Hard* with 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 6.2), which can be:

- *Good* visibility, when all aircraft are visible, or,
- *Poor* visibility, when there is fog and no aircraft is visible.

6.2.4 Hypotheses

We expected participants to locate aircraft more accurately using AF modality in poor visibility conditions, even more so when coupled with vibrotactile feedback. Reaction times were not taken into account for these hypotheses since AF interaction modality requires movements to make of the head, which is a slower use than spatial sound only. Therefore, 3 working hypotheses have been made:

Experimental hypotheses for AF modality

When placed under poor visibility conditions, we expected that participants could locate aircraft more precisely (i.e., greater accuracy)...

- $H_{1.1}$: AF Vs. 3DS \rightarrow ...when they are using AF modality compared to 3DS modality;
- $H_{1.2}$: AF+V Vs. 3DS \rightarrow ...when they are using AF+V modality compared to 3DS modality;
- $H_{1.3}$: AF+V Vs. AF \rightarrow ...when they are using AF+V modality compared to AF modality.

6.2.5 Airport discrimination using selectable areas

At an airport, aircraft follow a runway circuit. For this experimental campaign, and considering statistical simplifications, Muret airport³ has been separated into 5 distinct areas. From the point of view of the control tower, 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 (including the runway), and a more distant one, located in the air after the runway. Regarding this approach and how spatial sounds

³This is the airport we considered for this experimental campaign; it is located near Toulouse, in the south west of France.

could be manipulated, the choices we made for these 5 areas are the following legs composing the airport circuit (see Figure 6.1):

- *Take-Off*;
- *Crosswind*;
- *Downwind*;
- *Base and Final*;
- *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. We will see a little further that this number corresponded to the difficulty level during the experiment.

6.2.6 Answering interface

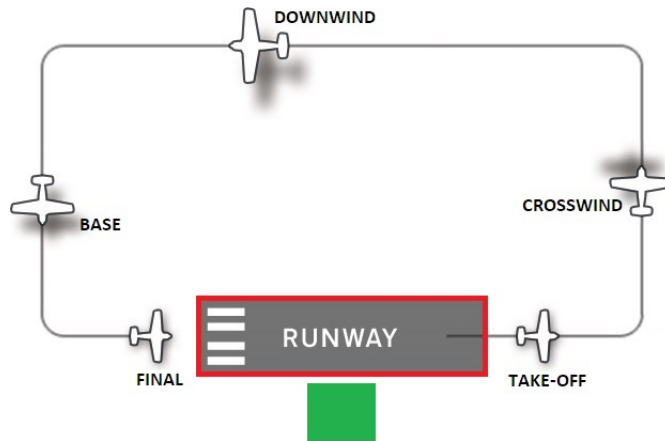
Multiple static aircraft combinations have been displayed through Flight Gear⁴ according to the 5 areas previously presented. All the participants heard the same configurations, in a random order for each *Modality*, *Difficulty* level and *Visibility* factors (i.e., conditions). By clicking on the corresponding area(s) through a specific 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, approximately on a virtual position materialized with a green square on Figure 6.1.

The distance between the center of this HMI and each of its button was constant to minimize and standardize their movements while answering. In addition, every button had the same color in order to not influence their answers. Buttons had two states: selected (dark grey) or not selected (light grey). 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 corner for right-handed participants, or at the bottom left corner for left-handed ones. This HMI was displayed on a tablet, which was used by the participants during the experiment (see Figure 6.1).

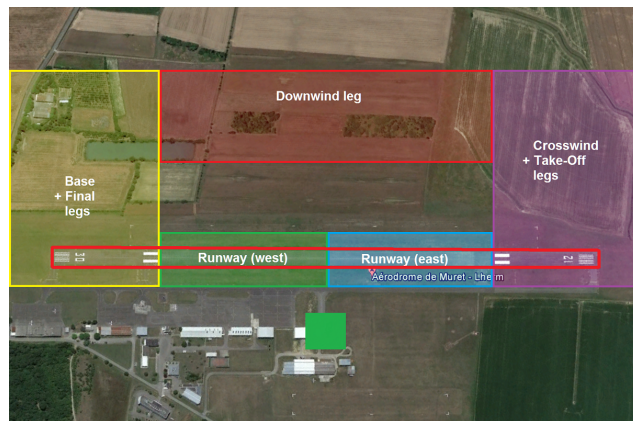
6.2.7 Metrics

Participants' performance has been measured objectively using 2 different behavioral scores (dependent variables, see Section 6.3.1):

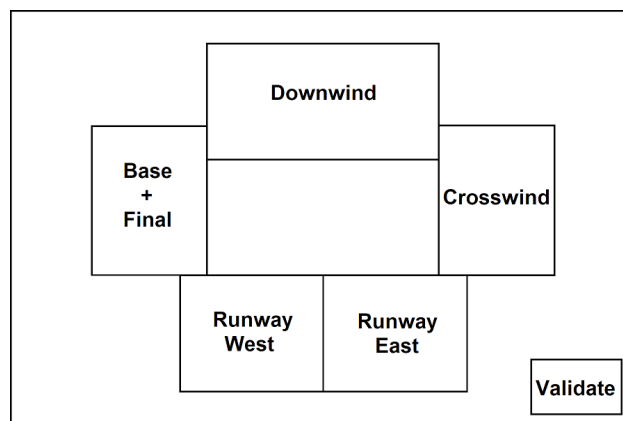
⁴Flight Gear is a realistic open flight simulator, freely available and which can be easily configured to be used along some other software modules; see <https://www.flightgear.org/> for more detail.



(a) A diagram of a typical runway circuit, which separates the four legs composing an approach segment: *base*, *final* and *take-off*, *crosswind* legs on either side of the airport circuit, *downwind* leg on the farthest side of the airport circuit in front of the control tower (green square), and the two parts composing the runway (squared in red, *east* and *west*).



(b) A satellite view of the Muret airport. Each area is represented with the same disposition as in (a).



(c) The layout for the HMI used to collect participants' answers during the first experimental campaign. The 5 buttons in the center corresponds to the 5 areas discriminating the airport environment (see (b)).

Figure 6.1 — Airport circuit discrimination using 5 distinct areas. On (a) and (b) the runway is highlighted in red lines, and the position of the control tower is represented with a plain green square. This last position corresponded to the virtual position of the participants during the experiment. (c) is the resulting layout used for the answering HMI.

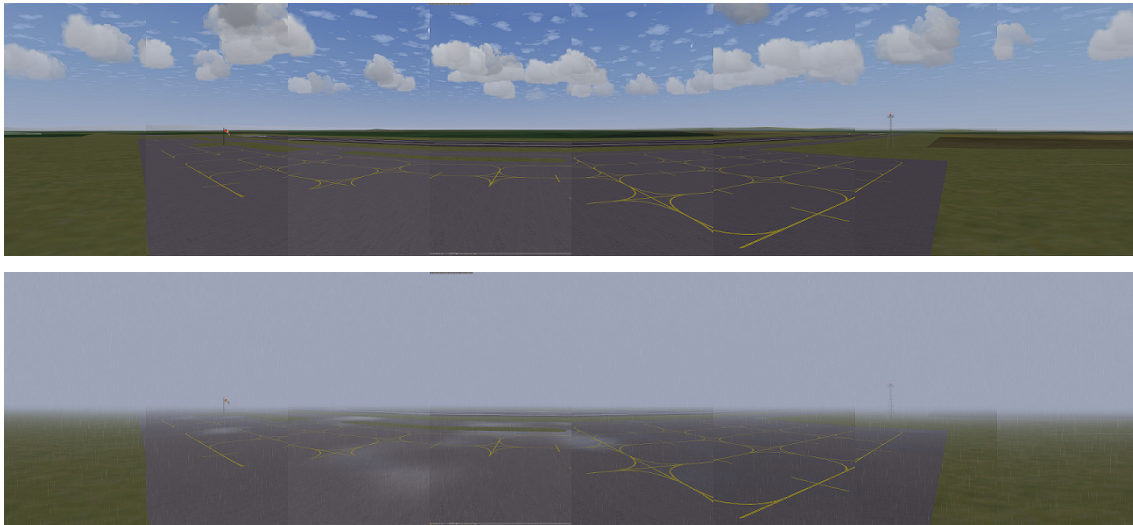


Figure 6.2 — Screenshots of the Flight Gear displays used during the study. The two meteorological situations are good visibility (top), and poor visibility (bottom) with fog. No aircraft was visible in this last condition.

- *Accuracy* in the task, which quantifies the number of right answers, i.e., the number of correctly located aircraft among the 5 possibilities (areas) offered by the experimental design;
- *Reaction time*, which is the time taken by the participants to locate aircraft, from the moment when a new combination was displayed until the moment when the Validate button was pressed on the answering HMI.

We also carried out neurophysiological measures to provide another objective measurement of the performance using a *Mental workload* index (see Section 6.3.3). Subjective performance was obtained by asking the participants to fill an online set of questionnaires directly after the end of the experiment (see Section 6.3.2), comprising:

- *Fatigue* questionnaire,
- *Usability* questionnaire,
- and *NASA-TLX* questionnaires.

6.2.8 Apparatus

The setup was composed of 8 UHD Iiyama Prolite X4071 screens (40 inches), an Alienware Area51 computer equipped with 2 NVidia GeForce GTX 1080 graphic cards, two other computers and screens to display air and ground radar HMIs, and a wooden chair on which 2 Clark Synthesis T239 Gold tactile transducers have been attached (see Figure 6.3, one behind the back to code the *Up* modality, and another one under the seat to code the *Down* modality).

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 speakers of the 8 UHD screens used to provide a

panoramic view on the airport vicinity. This solution provides a *physical* spatialization of sound via the physical positions of the 8 speakers.⁵ 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.⁶

The 3D environment was created using Flight Gear open flight simulator (see Figure 6.2 and 6.6). The different software modules for the augmentations, i.e., 3DS and AF interaction modality coupled or not with vibrotactile feedback, were written in C# language using Microsoft .Net framework SDK 4.6 and Direct Sound library. Network communications were developed using ENAC Ivy bus technologies [30, 33], which provides a high level means of communication using string messages and a regular expression binding mechanism (see Figure 6.3).

6.2.9 Combinatorics of the trials

There were 3 possibilities for the modality used (3DS, AF and AF+V), 2 possibilities for the visibility conditions (Good and Poor), and 3 difficulty levels (Easy, Medium and Hard). The number of combinations for 5 airport areas among 1, 2 or 3 simultaneous aircraft (difficulty levels) is given by binomial coefficients⁷ C_k^N :

$$\begin{cases} \text{Easy:} & C_1^5 = 5, \\ \text{Medium:} & C_2^5 = 10, \\ \text{Hard:} & C_3^5 = \binom{5}{2} = 10. \end{cases}$$

However, some of these combinations are redundant. For example, for a given participant it is not mandatory to test combination 1.2a and combination 1.2b (see Figure 6.4), because of their symmetrical positions: the first one is used to 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. Considering that participants have no hearing imbalance between their left and right ears, these two examples are equivalent. This is why 2 sets of trials have been created, in which the same numbers of combinations per difficulty level were selected in a balanced:

- set $A = \{ 1.1, 1.2.a, 1.2.b, 1.3.a, 1.3.b, 2.2, 2.3.a, 2.4.a, 2.5.a, 2.6.a, 3.2, 3.3.a, 3.4.a, 3.5.a, 3.6.a \}$;
- set $B = \{ 1.1, 1.2.a, 1.2.b, 1.3.a, 1.3.b, 2.1, 2.3.b, 2.4.b, 2.5.b, 2.6.b, 3.1, 3.3.b, 3.4.b, 3.5.b, 3.6.b \}$.

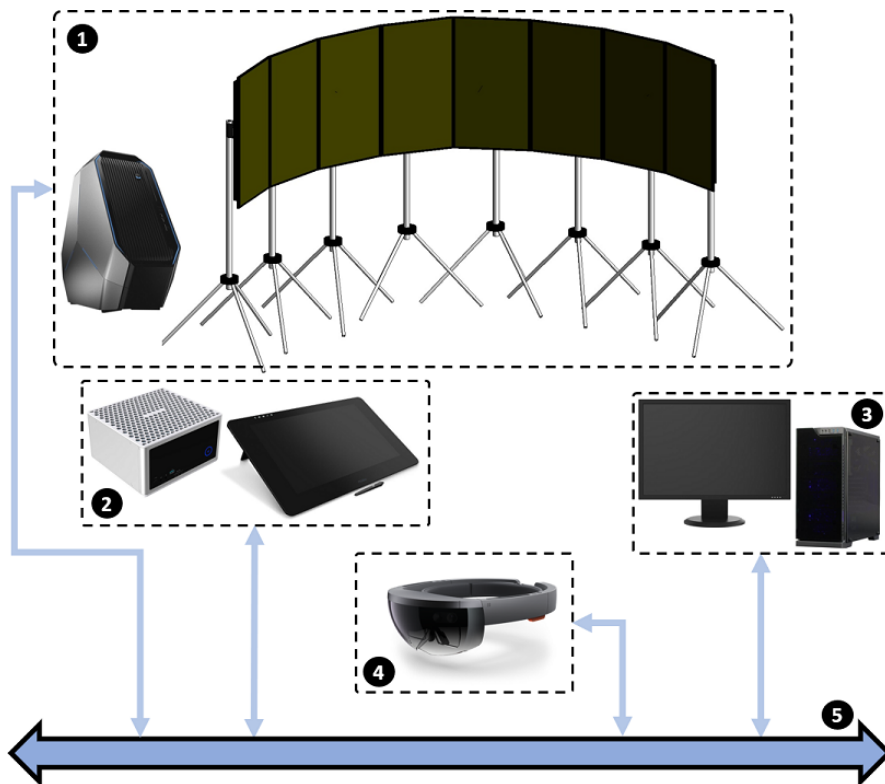
⁵This solution was mainly chosen because we had demonstration constraints; this way, the platform could be seen and heard by several people at the same time when it was demonstrated.

⁶This can be seen as unsuitable and expensive, however, as explained previously in Chapter 5, Section 5.5.1, p. 101, we could not use an inertial unit because of the magnetic field induced by the 2 transducers. The HoloLens was only used to measure the participants' head orientation, and its augmented reality features were not considered for this experiment.

⁷ k among N .



(a) The “*haptic chair*”, on which 2 tactile transducers have been attached to spread localized vibrotactile feedback under the seat and behind the back of the users. More precisely, these tactile feedback were located a little below the shoulder blades (depending on the size of the participants), and under the thighs.



(b) ❶ Digital visuals for the airport vicinity are rendered using an Alienware computer and 8 40 inches UHD screens, ❷ ground radar view is computed and rendered using a Zotac Mini PC and a Wacom tablet, ❸ air radar was computed and rendered using generic desktop computer and display, ❹ head position and orientation were retrieved using a Microsoft HoloLens AR headset, and ❺ all these equipment were connected to each other’s through ENAC’s Ivy software bus [30, 33].

Figure 6.3 — (a) Haptic chair and (b) complete apparatus used for the first experiment.

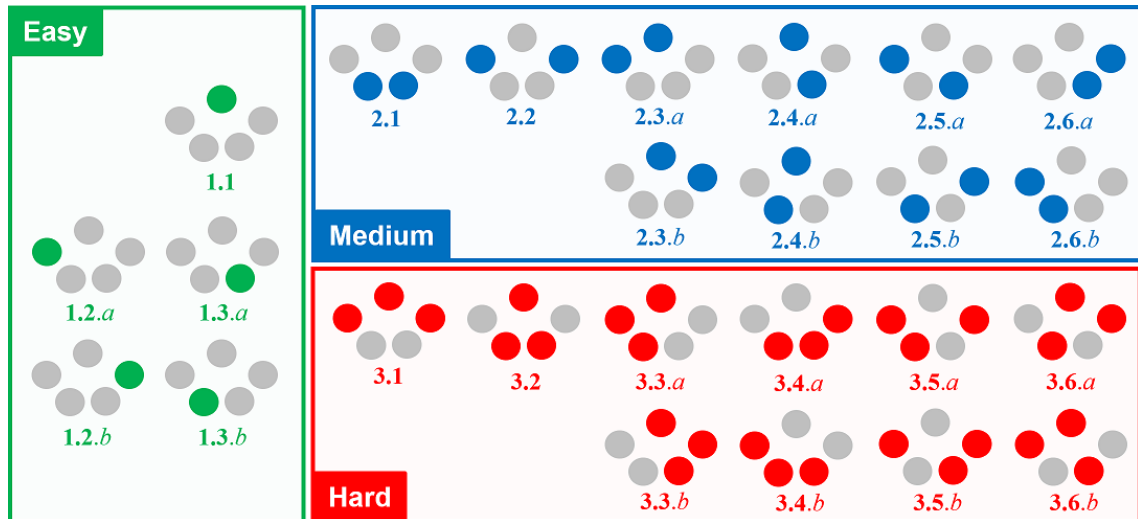


Figure 6.4 — The combination used to place the sound sources during the experiment. The dots correspond to the 5 areas. Grey ones denote unused ones. Difficulty level 1 (Easy) is in green (1.X.y), difficulty level 2 (Medium) is in blue (2.X.y), and difficulty level 3 (Hard) is in red (3.X.y). For a given combination, *a* and *b* designate the possible symmetrical combinations.

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 4 times to each participant, which led to a total of 360 trials for each participant.

6.2.10 Experiment protocol

Figure 6.5 hereafter provides details on the entire protocol designed for this first experimental campaign. After a welcome to each participants (see the welcome text page in Appendix 10.8), the experiment started with a training phase which consisted of a presentation of the stimuli (spatial sound sources playing engine sounds) in each different combination. To do so, 2 experimental blocks have been made: one under good visibility conditions, and another one under poor visibility conditions. Each of these 2 experimental blocks contained 3 other sub-blocks, one for each of the 3 modalities tested (3DS, AF and AF+V). The goal was to accustom the participants to “play” with the sounds by moving their head like if they had to look for something. However, as explained previously in this manuscript, the search is done here using hearing instead of sight. Each of these sub-blocks contained 3 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 experimental cases.

Then the experiment started after this first training phase, and it was divided into blocks for each modality. Each of these blocks occurred with good and poor visibility conditions, which means a total of 6 distinct blocks (3DS good, 3DS bad, AF good, ..., AF+V bad). Within these blocks, the fifteen combinations were randomly presented to each participant, following the set

A or B that had been randomly (but equally) assigned to them (see previous section). Finally, the blocks were randomly ordered too. The 6 modality blocks were presented 4 times to each participant, with a 5-minute break in the middle of the experiment. At the end the participants completed an online questionnaire in order to give their personal feelings about the usability, fatigue, performance felt, and perceived workload (see Section 6.3.2).

6.3 Data analysis

In this section we describe which type of data were acquired during the experiment, and how they were statistically analyzed. Behavioral analysis are explained in Section 6.3.1, Subjective analysis are explained in Section 6.3.2, and neurophysiological measurements and analysis are explained in Section 6.3.3.

6.3.1 Behavioral data

The two behavioral variables which were acquired correspond to the 2 variables described in Section 6.2.7: *Accuracy*, which is defined by a percentage of correct aircraft localization by the participants for the five airport areas, and *Reaction time*, in milliseconds.

The Accuracy variable was normalized using Arcsine transform [181]⁸, and Reaction time using Log transform [143]. For each variable, a two-way ANalysis Of VAriance (ANOVA) with repeated measures ($CI = .95$) 3×2 (Modality [3DS, AF, AF+V] \times 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 a 3-level interaction analysis (i.e., $3 \times 3 \times 2$ (Modality [3DS, AF, AF+V] \times Difficulty [1, 2, 3] \times Visibility [Good, Poor]), which would mean a theoretical number of 18 conditions per participant among a practical number of 22 participants), difficulty levels were averaged for each Modality and Visibility factor.

6.3.2 Subjective data

After the experiment, participants were immediately asked to fill in an online questionnaire which was divided into 5 main parts (see Appendix 10.8). 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. A one-way ANOVA with Modality factor and *Usability score* implemented as within factor have been investigated. 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 \times Visibility \times Difficulty level combinations. The fourth part consisted in the *NASA Task Load Index (NASA-TLX)* questionnaires [75, 76], one for each modality, which were used to estimate the cognitive workload. NASA-TLX questionnaire is composed of 6 parts to assess the:

⁸Given a data value $v \in \{0, 1\}$, these transformation consists in computing $\arcsin(\sqrt{v})$.

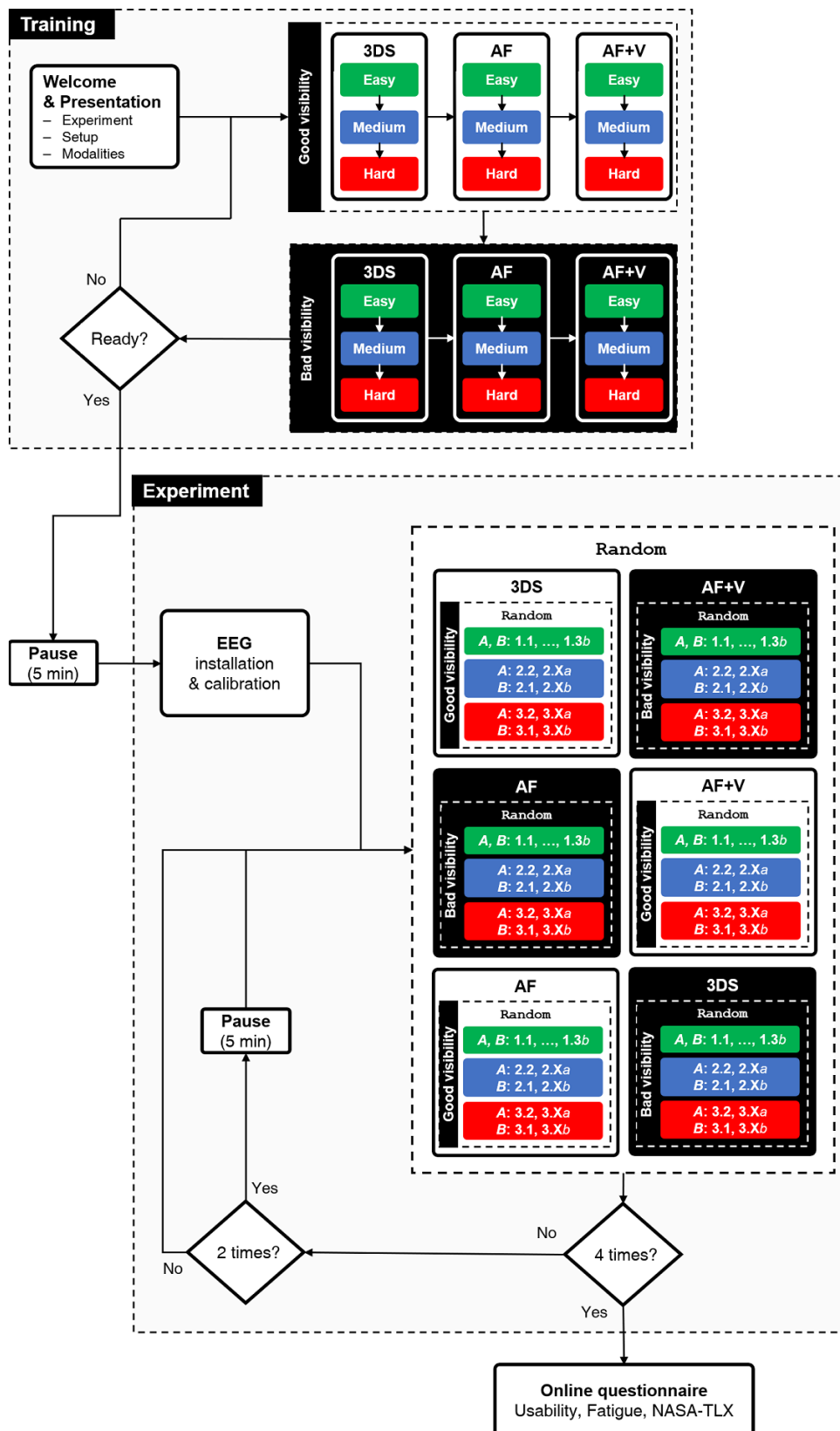


Figure 6.5 — Detailed diagram for the protocol which was designed to evaluate Audio Focus interaction modality during the first experiment (if required, see Sections 6.2.9 and 6.2.10 for a detailed explanation).

- Mental demand: *How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?*
- Physical demand: *How much physical activity was required? Was the task easy or demanding, slack or strenuous?*
- Temporal demand: *How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?*
- Performance: *How successful were you in performing the task? How satisfied were you with your performance?*
- Effort: *How hard did you have to work (mentally and physically) to accomplish your level of performance?*
- Frustration: *How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?*

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 departed 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 6.5.1). One-way ANOVA with repeated measures ($CI = .95$) with *Modality* factor [3DS, 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 [3DS, AF, AF+V] \times 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.

6.3.3 Neurophysiological data

Electroencephalography (EEG) is one of the most commonly used techniques when it comes to collect relevant data related to brain activity and mental workload variations. This is mostly due to its relatively low cost and the higher temporal resolution that is offered to practitioners, comparing to other brain imaging techniques such as fMRI or MEG⁹, for example. This type of measurements have the advantage to be accessible in real time if necessary during an experiment, and to

⁹fMRI, standing for *Functional Magnetic Resonance Imaging*, is a brain imaging technique based on the detection of changes in blood flow, which is linked with neural activation. MEG, standing for *Magnetoencephalography*, is another neuroimaging technique that relies on the magnetic field induced by natural electrical activity of neurons.

not impact the experimental task, unlike subjective measures, which necessarily require the participants' intervention since they aim at collecting their own, personal perception. During the first experimental campaign, neurophysiological measurements were carried out by researchers from the University of La Sapienza in Roma, who were also part of the MOTO project, in order to quantify the mental workload of the participants. Unlike NASA-TLX assessments, which provide a subjective measure of mental workload, the neurophysiological measurements have allowed us to obtain an objective measure of mental workload. We describe here the procedure which was envisaged by those researchers. For more details, please refer to the EMBC¹⁰ publication [10].

This objective neurophysiological mental workload was measured using W_{EEG} index, which has been computed from the participants' EEG activity and other physiological parameters such as Galvanic Skin Response (GSR). For each participant, scalp EEG signals have been recorded using a digital amplifier with wet electrodes covering the frontal and parietal sites (see the white cap on images in Figure 6.6). In order to compute the W_{EEG} index, several steps have been performed, consisting of processing (e.g., filtering) the EEG signal to remove any artefacts, noise or other interfering signals that may affect the overall quality of the signal. More precisely, the researchers who were in charge of this measurement described this process as follows: “*neurophysiological mental workload index has been computed from the EEG activity for each [participant], as the ratio between frontal theta and parietal alpha frequency bands contributions [...] In particular, for each [participant], scalp EEG signals have been recorded by the digital monitoring beMicro amplifier (EBNeuro system) with a sampling frequency of 256 (Hz) by 13 Ag/AgCl passive wet electrodes covering the frontal and parietal sites (Fpz, AFz, AF3, AF4, Fz, F3, F4, Pz, P3, POz, PO3, PO4) referenced to both the earlobes and grounded to the left mastoid, according to the 10–20 standard [91]. In order to compute the W_{EEG} index the following steps have been performed. First, EEG signals have been band-pass filtered with a fifth-order Butterworth filter [1–30Hz] and segmented in 2-seconds long epochs, shifted of 125 ms [11]. Artefacts contributions that could affect the morphology of theta and alpha bands (e.g., eyes blinks and saccades, muscular artefacts, amplifiers saturations) have been removed by following specific procedures available in the EEGLAB toolbox [45]. All the EEG epochs marked as artefact have been rejected from the EEG dataset with the aim to have an artefact-free EEG signal from which to estimate the brain variations along the different modalities [...] At this point, the Power Spectral Density (PSD) was calculated for each EEG epoch using a Hanning window [11] of the same length of the considered epoch (2 seconds). Then, the EEG frequency bands of interest have been defined for each participant by the estimation of the Individual Alpha Frequency (IAF) value [96]. In order to have a precise estimation of the alpha peak and, hence of the IAF, each [participant] has been asked to keep the eyes closed for a minute before starting with the experiment. Finally, the theta rhythm ($IAF - 6 \div IAF - 2$), over the EEG frontal channels (Fpz, AFz, AF3, AF4, Fz, F3 and F4), and the alpha rhythm ($IAF - 2 \div IAF + 2$), over the EEG parietal channels (Pz, P3, P4, POz, PO3 and PO4) have been considered as variables for the W_{EEG} index estimation [...]” The resulting signal have been normalized using the z-score method [187] to obtain the W_{EEG} values.*

Then, 2 different experimental conditions were considered for the neurophysiological analysis:

¹⁰IEEE Engineering in Medicine and Biology Society (EMBC)

Normal, which is when AF interaction modality was not activated (in other words, this corresponds to 3DS condition), and *Augmented*, which was when AF+V interaction modality was activated.¹¹ Finally, a paired Student's *t*-test ($CI = .95$) has been performed to investigate differences between these 2 factors.

6.4 Behavioral results

In this section we provide a detailed presentation of the behavioral results, namely concerning Accuracy (Section 6.4.1) variable and Reaction times (Section 6.4.2). The reader can refer to Section 6.3.1 to see how Accuracy and Reaction time data have been analyzed.

6.4.1 Accuracy

Analysis revealed main effects for Modality and Visibility factors and a Modality \times Visibility interaction. Detailed results are reported in Table 6.2.

Table 6.2 — First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with *Accuracy* dependent variable implemented as within factor (Audio Focus modality).

Variable	ddl	<i>F</i>	<i>p</i>	η_p^2
Modality	2, 42	116.93	< .0001	.85
Visibility	1, 21	409.93	< .0001	.95
Modality \times Visibility	2, 42	215.52	< .0001	.91

Tukey's HSD post-hoc analysis showed the following main results (Figure 6.7). For the Modality factor main effect: participants were more accurate using AF+V modality ($M = 1.36$, $SD \pm .02$) than using AF modality ($M = 1.19$, $SD \pm .01$, $p < .001$) and 3DS modality ($M = 1.11$, $SD \pm .02$, $p < .001$). They were also more accurate using AF modality than using 3DS modality ($p < .001$).

For the Modality \times 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 3DS modality ($M = 1.43$, $SD = .02$, $p < .001$). In poor visibility conditions, participants gave more accurate answers using AF+V modality ($M = 1.28$, $SD \pm .03$) than using AF modality ($M = .9$, $SD \pm .02$, $p < .001$) and 3DS modality ($M = .78$, $SD \pm .02$, $p < .001$). Participants were also more accurate using AF modality than using 3DS modality ($p < .001$).

6.4.2 Reaction time

The analysis revealed main effects for Modality and Visibility factor and a Modality \times Visibility interaction. Detailed results are reported in Table 6.5.

¹¹For simplification purpose, AF condition, i.e., without haptic feedback, was not considered for these neurophysiological assessments.



Figure 6.6 — Photograph of the sandbox used for the first experiment while a participants are answering. We can see that each participant wears an EEG on their head in the form of a cap with electrodes.

Table 6.3 — First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with *Reaction time* dependent variable implemented as within factor (Audio Focus modality).

Variable	ddl	F	p	η_p^2
Modality	2, 42	51.86	< .0001	.71
Visibility	1, 21	16.55	< .001	.44
Modality \times Visibility	2, 42	18.43	< .0001	.47

Tukey’s HSD post-hoc analysis showed the following main results (Figure 6.7). For the Modality factor main effect: participants were faster using AF+V modality ($M = 3.88$, $SD \pm .02$) than using AF modality ($M = 3.93$, $SD \pm .02$, $p < .01$), but slower than using 3DS modality ($M = 3.8$, $SD \pm .01$, $p < .001$). They were also slower using AF modality than using 3DS modality ($p < .001$).

For the Modality \times Visibility interaction, in good visibility conditions, participants were slower using AF+V modality ($M = 3.85$, $SD \pm .02$) than using 3DS modality ($M = 3.81$, $SD \pm .02$, $p < .05$). They were also slower using AF modality ($M = 3.89$, $SD \pm .02$) than using 3DS modality ($p < .001$). In poor visibility conditions, participants were faster using AF+V modality ($M = 3.92$, $SD \pm .02$) than using AF modality ($M = 3.97$, $SD \pm .02$, $p < .01$), but slower than using 3DS modality ($M = 3.79$, $SD \pm .02$, $p < .001$). They were also slower using AF modality than using 3DS modality ($p < .001$).

6.5 Subjective results

Subjective results are divided into 4 groups: results from *General questions* (Section 6.5.1), results from questionnaire related to *Usability* (Section 6.5.2), results from questionnaire related to *Fatigue* (Section 6.5.3), and results from *NASA-TLX* questionnaires (Section 6.5.4). The reader can refer to Section 6.3.1 to have a look at how subjective measurements have been analyzed.

6.5.1 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 positively. 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? (3DS, AF or AF+V)*”, 95.7% of the participants answered AF+V, while 4.3% 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,*”

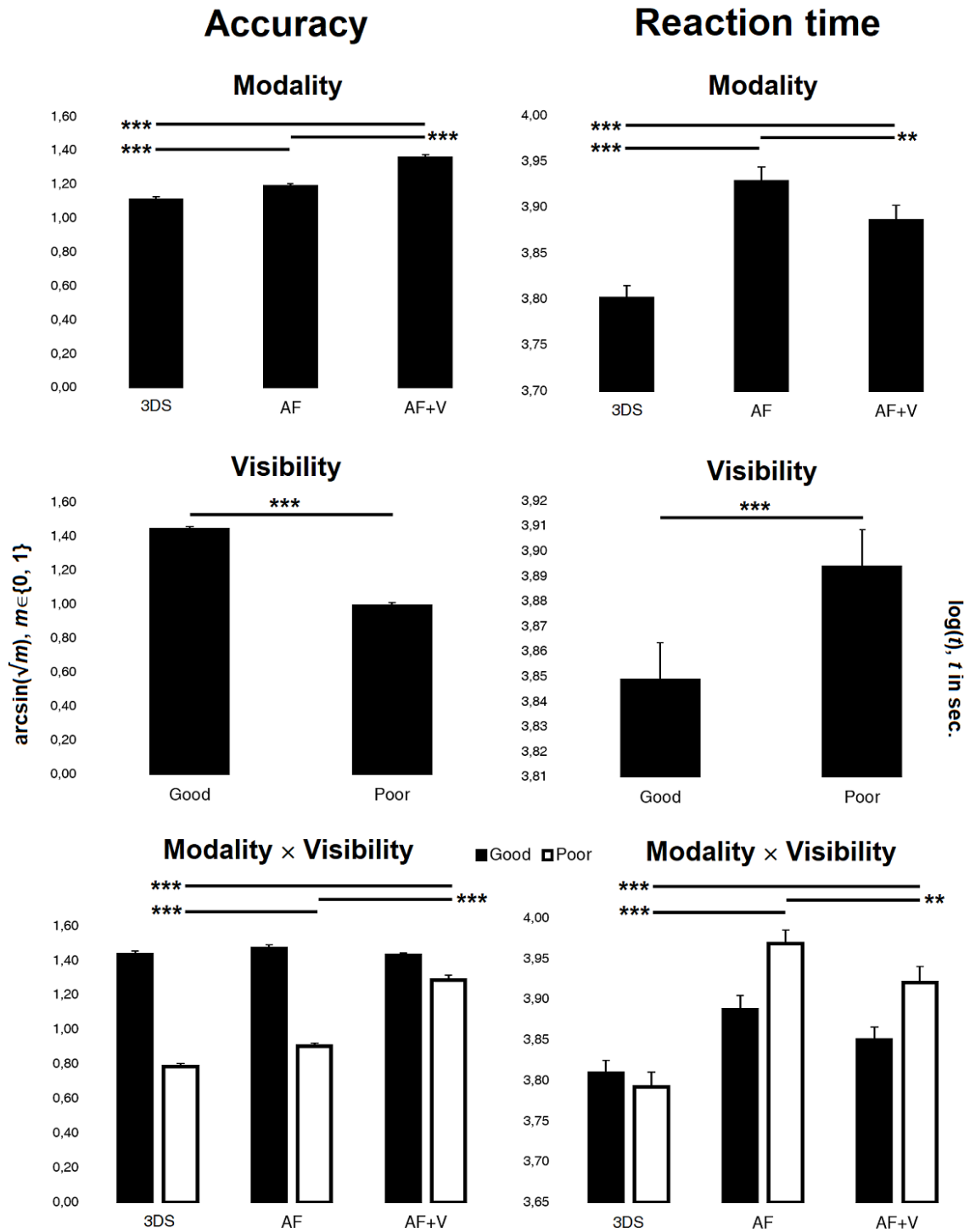


Figure 6.7 — Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with *Accuracy* (left, Arcsine normalization) and *Reaction time* (right, Log normalization) dependent variables implemented as within factor for Audio Focus interaction modality. Error bars are standard errors.

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 ($SD = 1.71$).

6.5.2 Usability

A main effect was found for the Modality factor. Detailed Usability results are reported in Table 6.4

Table 6.4 — First experimental campaign: Results from one-way ANOVA with Usability score dependent variable implemented as within factor (Audio Focus modality).

Variable	ddl	F	p	η_p^2
Modality	2, 42	11.36	< .001	.35

Tukey’s HSD post-hoc analysis revealed that AF+V modality was considered more usable by the participants ($M = 1.28$, $SD \pm .04$) than AF modality ($M = .98$, $SD \pm .04$, $p < .001$) and 3DS modality ($M = .96$, $SD \pm .08$, $p < .001$). No significant difference was found between AF and 3DS modalities (Figure 6.8).

6.5.3 Fatigue

Main effects for Modality and Visibility factors and a significant Modality \times Visibility interaction were found. Detailed Fatigue results are reported in Table 6.5

Table 6.5 — First experimental campaign: Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with Fatigue score dependent variable implemented as within factor (Audio Focus modality).

Variable	ddl	F	p	η_p^2
Modality	2, 42	60.23	< .0001	.74
Visibility	1, 21	138.09	< .0001	.87
Modality \times Visibility	2, 42	67.23	< .0001	.76

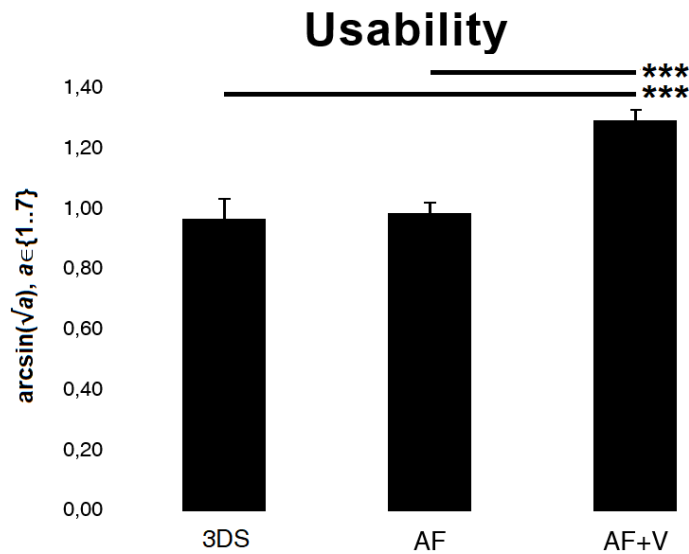


Figure 6.8 — Results from one-way ANOVA with *Usability score* dependent variable implemented as within factor for Audio Focus interaction modality. Error bars are standard errors.

Tukey's HSD post-analysis revealed the following main results (Figure 6.9). For the Modality factor main effect: participants have felt less fatigue using AF+V modality ($M = 1.41$, $SD \pm .03$) than using AF modality ($M = 1.21$, $SD \pm .02$, $p < .001$) and 3DS modality ($M = 1.13$, $SD \pm .02$, $p < .001$). They also perceived less fatigue using AF modality than 3DS modality ($p < .05$).

For the Modality \times Visibility interaction, no significant results were found in good visibility conditions. However, in poor visibility conditions, participants perceived less fatigue using AF+V modality ($M = 1.33$, $SD \pm .04$) than AF modality ($M = .96$, $SD \pm .03$, $p < .001$) and 3DS modality ($M = .8$, $SD \pm .04$, $p < .001$). They also perceived less fatigue using AF modality than 3DS modality ($p < .001$).

6.5.4 NASA Task Load Index

Results are reported for each NASA-TLX parts in Table 6.6 and Figure 6.10. No significant results were found concerning Physical demand and Temporal demand parts.

Table 6.6 — First experimental campaign: Results from NASA-TLX with each NASA-TLX sections implemented as within factor (Audio Focus modality).

NASA-TLX	Variable	ddl	F	p	η_p^2
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

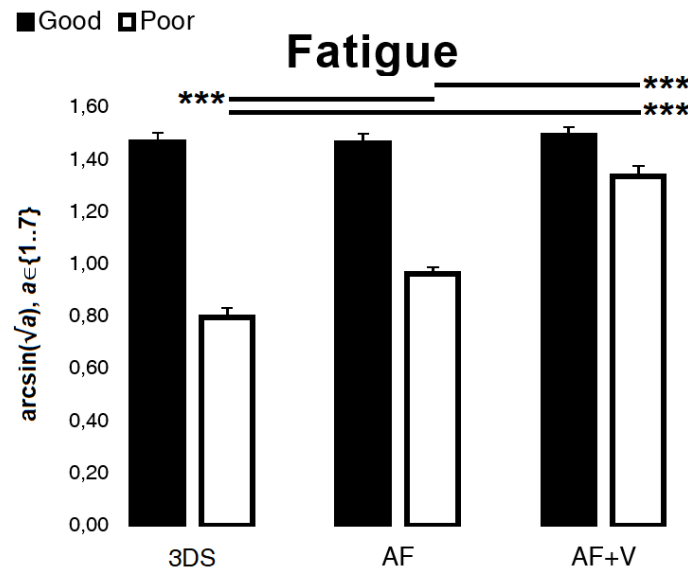


Figure 6.9 — Results from 3×2 (Modality [3DS, AF, AF+V] \times Visibility [Good, Poor]) ANOVA with *Fatigue score* dependent variable implemented as within factor for Audio Focus interaction modality. Error bars are standard errors.

For the Mental demand part, a main effect was found on the Modality factor [$F(2, 42) = 8.94$, $p < .001$, $\eta_p^2 = .3$]. Tukey's HSD post-hoc analysis revealed participants felt that AF+V required a smaller mental demand ($M = .86$, $SD \pm .04$) than 3DS modality ($M = 1.03$, $SD \pm .06$, $p < .001$). Also, AF modality required a smaller mental effort ($M = .93$, $SD = .04$) than 3DS 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 [$F(2, 42) = 55.97$, $p < .0001$, $\eta_p^2 = .73$]. Tukey's HSD post-hoc analysis revealed that participants felt more efficient using AF+V modality ($M = 1.31$, $SD \pm .05$) than AF modality ($M = .95$, $SD \pm .03$, $p < .001$) and 3DS modality ($M = .8$, $SD \pm .03$, $p < .001$). They also felt more efficient using AF modality than 3DS modality ($p < .05$).

For the Effort part, a main effect was found on the Modality factor [$F(2, 42) = 13.01$, $p < .0001$, $\eta_p^2 = .38$]. Tukey's HSD post-hoc analysis revealed that AF+V modality less effort from the required the participants ($M = .78$, $SD \pm .04$) than AF modality ($M = .99$, $SD \pm .03$, $p < .001$) and 3DS modality ($M = 1.02$, $SD \pm .05$, $p < .001$). No significant difference was found between AF and 3DS modalities.

Finally for the Frustration section, a main effect was found on the Modality factor [$F(2, 42) = 24.53$, $p < .0001$, $\eta_p^2 = .54$]. Tukey's HSD post-hoc analysis revealed that participants felt less frustrated using AF+V modality ($M = .57$, $SD \pm .04$) than AF modality ($M = .86$, $SD \pm .05$, $p < .001$) and 3DS modality ($M = .94$, $SD \pm .06$, $p < .001$). No significant difference was found between AF and 3DS modalities.

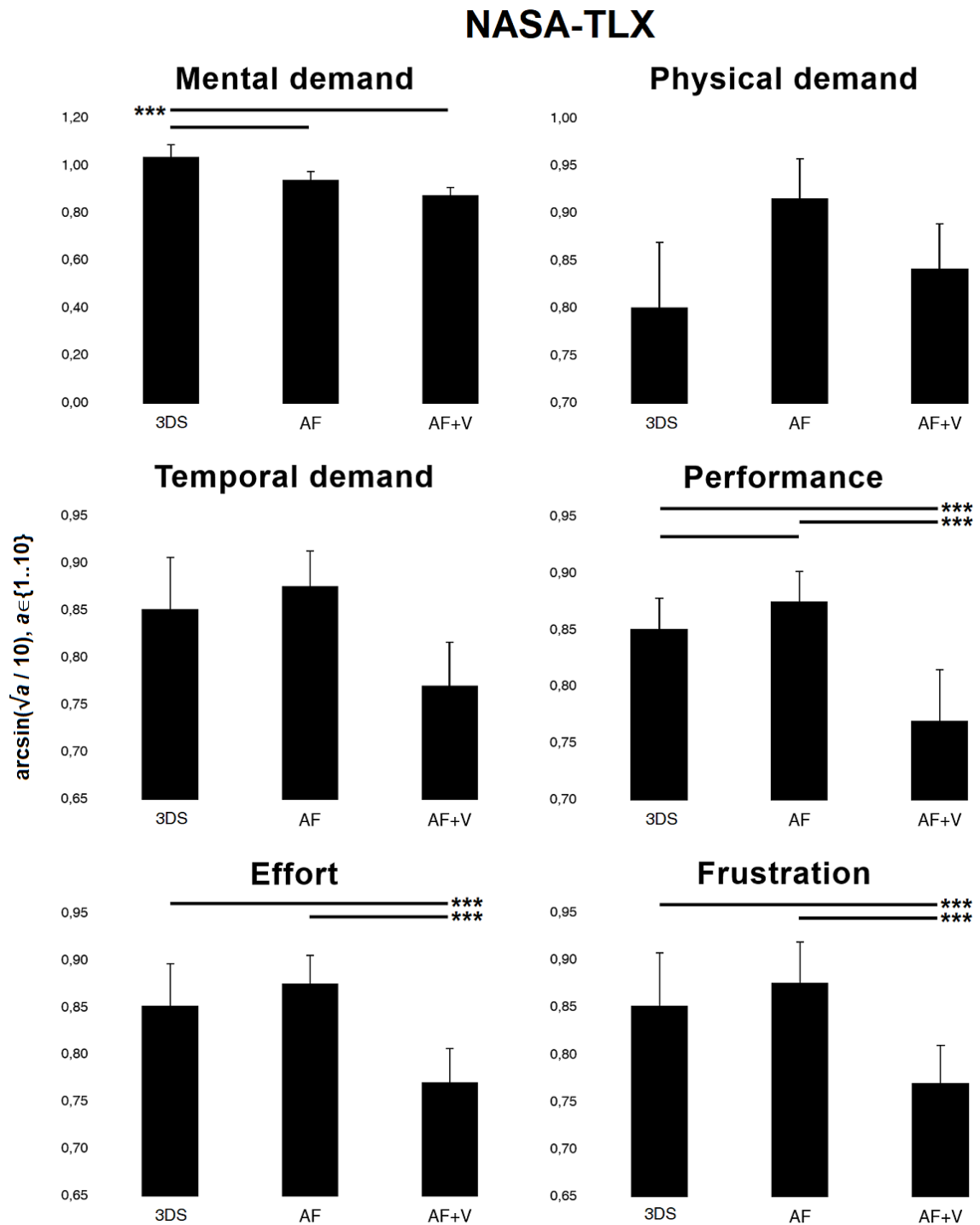


Figure 6.10 — Results from NASA-TLX with each NASA-TLX sections implemented as within factor for Audio Focus interaction modality. Error bars are standard errors.

6.6 Neurophysiological results

Readers can refer to Section 6.3.3 for more details concerning the analysis that have been led on neurophysiological data. These statistical analysis performed among the experimental conditions, revealed a significant decrease of the experienced workload during the *Augmented* condition with respect to the *Normal* one ($p < .01$). In other words, the statistical analysis highlighted a significant decrement in W_{EEG} scores when AF+V modality was activated (see Figure 6.11).

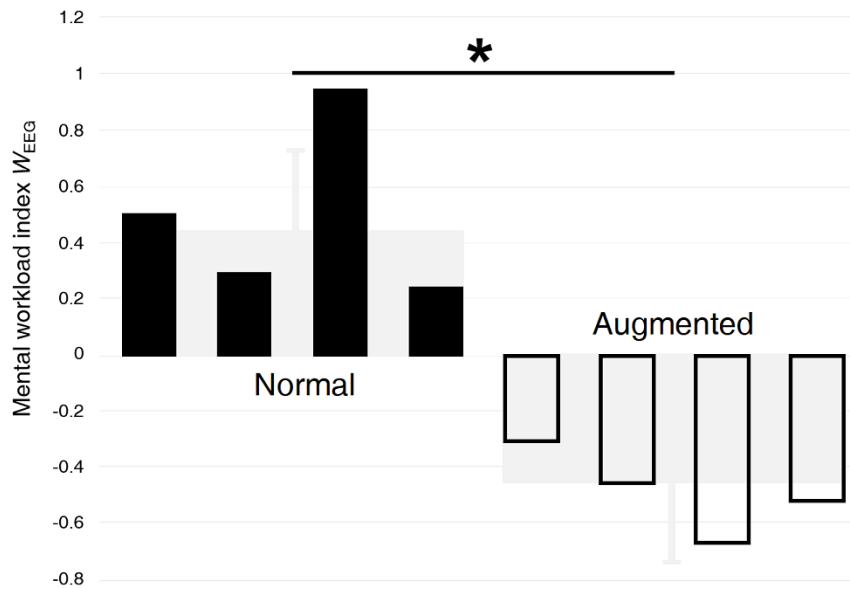


Figure 6.11 — W_{EEG} scores and standard deviation related to the workload experienced by the ATCo, for each modality and repetition.

Wrap-up message

In this chapter, we presented the entire experimental approach which was considered for the evaluation of Audio Focus interaction modality. After having formalized our working hypothesis (see $H_{1.1}$, $H_{1.2}$ and $H_{1.3}$, Section 6.2.4) and described the experimental task, we shown that a complete experimental bench have been set up with the goal to be as close as possible to the working conditions of a RCT environment. However, mainly because of the statistical positions of aircraft presented to the participants, the experimental task was more a laboratory task than an ecological one. Hence, AF interaction modality have been tested in several aspects. To address our problematic, a complete and exhaustive (from the combinatorics point of view) protocol have been designed.

Data from different types were collected among a population of 16 professional ATCos: *Accuracy* and *Reaction times* as behavioral measurements, as well as answers from *Fatigue*, *Usability* and *NASA-TLX* questionnaires were collected as subjective measures. A *mental workload index* W_{EEG} was also computed from neurophysiological data as an objective neurophysiological value related to the performance. To clarify our findings, we propose a synthesized view on the results mentioned previously. Table 6.7 hereafter summarizes the main results which were found during

Table 6.7 — First experimental campaign: Summary of the results obtained during the first experimental campaign regarding the Audio Focus interaction modality. Right column gives an indication how symbol > should be read.

Variable	Simplified result	Comments
<i>Accuracy</i>	AF+V > AF > 3DS	More accurate
	AF+V > AF	Faster
<i>Reaction Time</i>	3DS > AF+V	Faster
	3DS > AF	Faster
<i>Usability</i>	AF+V > AF > 3DS	More usable
<i>Fatigue</i>	AF+V > AF > 3DS	Generate less fatigue
<i>NASA-TLX Mental demand</i>	3DS > AF	More demanding
	3DS > AF+V	More demanding
<i>NASA-TLX Performance</i>	AF+V > AF > 3DS	More efficient
<i>NASA-TLX Effort</i>	AF+V > AF	Less effort
	AF+V > 3DS	Less effort
<i>NASA-TLX Frustration</i>	3DS > AF+V	More frustrating
	AF > AF+V	More frustrating
<i>Mental workload W_{EEG}</i>	3DS > AF+V	Generates less workload

this first experimental campaign under poor visibility conditions. The main message that emerges from these results is that AF interaction modality, coupled with vibrotactile feedback or not, in addition to having been well received by professional ATCos participants, has provided them with an effective way to locate aircraft when they had no visibility over the airport environment. Actually, while Reaction times were slightly slower when using AF interaction modality, they provided much more accurate answers. At the same time, subjective results gave us an overview on how the participants received this kind of uncommon tool: they globally received it positively. However, our mental workload assessment revealed that they were more cognitively charged when using these augmentation means. These results will be discussed in the following section.

Chapter 7

First experimental campaign: Discussion

Résumé (Français)

Première campagne expérimentale (Discussion)

Les résultats sont conformes à nos hypothèses : lorsqu'ils utilisaient la modalité Audio Focus (AF), couplée à des vibrations ou non, les participants ont localisé de façon plus précise les avions qui avaient été disposés dans les différentes zones de l'aéroport. Par ailleurs, les temps de réaction étaient légèrement plus élevés lorsque la modalité AF était utilisée dans des conditions de visibilité mauvaise : alors que les participants ont mis en moyenne environ 6,3 secondes pour localiser des avions avec l'aide du son spatialisé uniquement (3DS), ils ont mis environ 9,4 secondes lorsqu'ils utilisaient la modalité AF, et environ 8,5 secondes lorsque que cette modalité était couplée aux vibrations (AF+V). Néanmoins, ces mesures de temps moyen peuvent être considérées comme relativement courtes par rapport à la tâche de travail globale des contrôleurs aérien. Cette durée d'exécution supplémentaire peut tout à fait s'expliquer par le fait que ce type d'interaction oblige l'utilisateur à faire des mouvements de tête pour localiser les sources sonores, alors que cette localisation est quasiment naturellement instantanée avec le son spatialisé uniquement. Ce résultat peut être considéré comme négligeable lorsqu'il est confronté aux résultats relatifs à la précision des participants : les avions ont été localisés à 49% en moyenne avec la modalité 3DS, 61% en moyenne avec la modalité AF, et 90% en moyenne lorsque cette dernière était couplée aux vibrations.

Les résultats subjectifs ont révélé une nette préférence pour la modalité AF+V : lorsque la visibilité était mauvaise, les participants ont eu plus de facilité à localiser les avions en utilisant cette modalité. De même, la fatigue perçue a été réduite. Cependant, les participants ont estimé que le son spatialisé était moins demandeur, physiquement, par rapport aux deux autres modalités. Comme expliqué précédemment, ceci est facilement compréhensible étant donné que la modalité AF nécessite d'effectuer des mouvements. La charge mentale perçue a été plus faible lorsque la modalité AF était couplée aux vibrations. De plus, la modalité AF a été perçue comme étant la plus efficace, la moins frustrante et la moins exigeante en terme d'effort à fournir pour la tâche. Tous ces résultats suggèrent qu'une fonctionnalité telle qu'Audio Focus peut être utile dans un contexte de contrôle à distance. En effet, selon les participants, 81% d'entre eux n'ont pas réussi au moins une fois à localiser un avion au cours de leur expérience professionnelle, ce qui justifierait l'emploi

d'une telle fonctionnalité. La modalité AF+V peut être considérée comme une fonctionnalité interactive, naturelle et intuitive. Nous constatons une nette préférence en faveur de cette modalité (95,7%), ce qui suggère que son concept est généralement considéré comme utile, surtout lorsqu'il est couplé à des vibrations.

Contrairement aux résultats subjectifs, qui ont montré que les participants ressentaient une charge mentale plus faible en utilisant la modalité AF+V, les résultats neurophysiologiques ont montré une tendance statistiquement différente de celle provenant des réponses aux questionnaires de type NASA TLX. En fait, cette tendance n'est pas significative pour la mesure subjective, ce qui n'est pas le cas de la mesure neurophysiologique. Par conséquent, il semblerait que les participants n'étaient pas conscients de leur charge mentale, sachant que leur performance, en termes de précision, était meilleure lorsqu'ils utilisaient la modalité AF+V, générant objectivement la plus grande charge cognitive (du point de vue des résultats neurophysiologiques).

De manière générale, les technologies telles que les cartes collaboratives, dans le domaine militaire, ou les bandes électroniques, dans le control du trafic aérien, ne sont pas toujours bien accueillies par leurs utilisateurs finaux car elles représentent un changement significatif par rapport aux autres technologies qu'elles visent à remplacer ou simplifier [35]. Cela ne semble pas être le cas de la modalité Audio Focus. Avec l'aide de contrôleurs aériens, Audio Focus a été conçue en tenant compte du fait que les contrôleurs recherchent souvent visuellement les avions qu'ils doivent contrôler. De ce point de vue, elle ne demande pas aux utilisateurs de modifier profondément leurs habitudes. De plus, nous savons que les interfaces supportant plusieurs modalités diminue l'ambiguïté et réduit le nombre d'erreurs [35, 124]¹. Nos résultats confirment cette tendance avec l'utilisation combinée de sons spatialisés et de stimuli tactiles localisés.

Cette recherche ouvre de nouvelles perspectives pour améliorer la qualité et l'efficacité des IHMs dans le contrôle aérien, dans le but final d'améliorer la performance des utilisateurs et d'accroître le niveau de sûreté. Par conséquent, des informations perceptuelles supplémentaires qui pourraient être fournies via le son spatialisé devraient être étudiées pour le contrôle du trafic aérien à distance. Dans notre étude, cela a été considéré comme une base de référence (modalité 3DS). Cependant, l'ajout d'information sonore à un contexte dans lequel les opérateurs n'y sont pas forcément habitués pourrait être perçu comme inconfortable. Une solution à cet inconvénient potentiel pourrait être de l'intégrer en tant que fonction activable à la demande. De plus, cette étude représente une perspective nouvelle quant à la contribution des mesures neurophysiologiques. En conclusion, les résultats obtenus suggèrent que la modalité Audio Focus constitue une forme d'interaction bien perçue par ses utilisateurs (expérimentaux). Il a été prouvé qu'un son spatial interactif couplé à des vibrations améliore le processus de sélectivité. Néanmoins, une nouvelle étude mériterait à être réalisée dans un cadre plus proche des contraintes opérationnelles afin de confirmer ces résultats dans un environnement plus écologique. De plus, une telle technique peut facilement être exportée au-delà de l'ATM, en particulier le handicap visuel, les réalité virtuelle, augmentées et mixtes ou les jeux vidéo. Certaines études sur le handicap visuel ont d'ailleurs déjà introduits des concepts d'interaction similaires [119, 40, 41].

¹Dans notre cas il s'agit plus de *stimuli multisensoriels* que de modalités interactives à proprement parler.

7.1 Introduction

In this first experimental campaign, 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 to see the aircraft. This specific use case of poor visibility conditions, as well as the design of Audio Focus interaction modality, have been made with the help of professional from the field, being both Tower ATCos (see Chapter 5, p. 93). AF 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.

We designed a complete experiment to perform performance, subjective and neurophysiological measurements that make us able to quantify the contribution of the AF interaction modality (see Chapter 6, p. 107). To do this, we have discretized the environment of the controlled airport (Muret) to be able to make accurate measurements. Through an experimental protocol, which was provided to 22 professional ATCos, we were able to propose a laboratory task involving an exhaustive set of trials via various combinations of Modalities (3DS, AF and AF+V), Difficulty levels (Easy, Medium and Hard) and Visibility conditions (Good or Poor). This allowed us to evaluate the participants' performance when using the Audio Focus interaction modality, in terms of Accuracy, Reaction Time, subjective questionnaires, and Neurophysiological measurements.

In this last chapter related to the first experimental campaign, we discuss the results exposed in the last sections of the previous chapter (Chapter 6, Sections 6.4, 6.5, and 6.6). In the first Section 7.2 we discuss the behavioral results. In the same way, subjective results are discussed in Section 7.3. Neurophysiological results are compared to subjective and behavioral ones in Section 7.4. We then have a word in Section 7.5 about the limitations of the study. A synthesis is proposed in Section 7.6 which make the link with the formal hypothesis previously formulated in Chapter 6, Section 6.2.4 (p. 113), before finishing with a conclusion on the first experimental campaign in Section 7.7.

7.2 Behavioral results

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 3DS. However, this was achieved with relatively longer reaction times. More precisely, reaction times were slightly higher with AF interaction modality in poor visibility conditions. Whereas participants ATCos have taken 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. Nevertheless, these averaged time measurements can be seen as relatively short regarding the ATCos' task. This additional time could be explained by the fact that this sort 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 3DS modality, 61% using AF modality and finally, 90% using AF+V modality (see Figure 7.1).

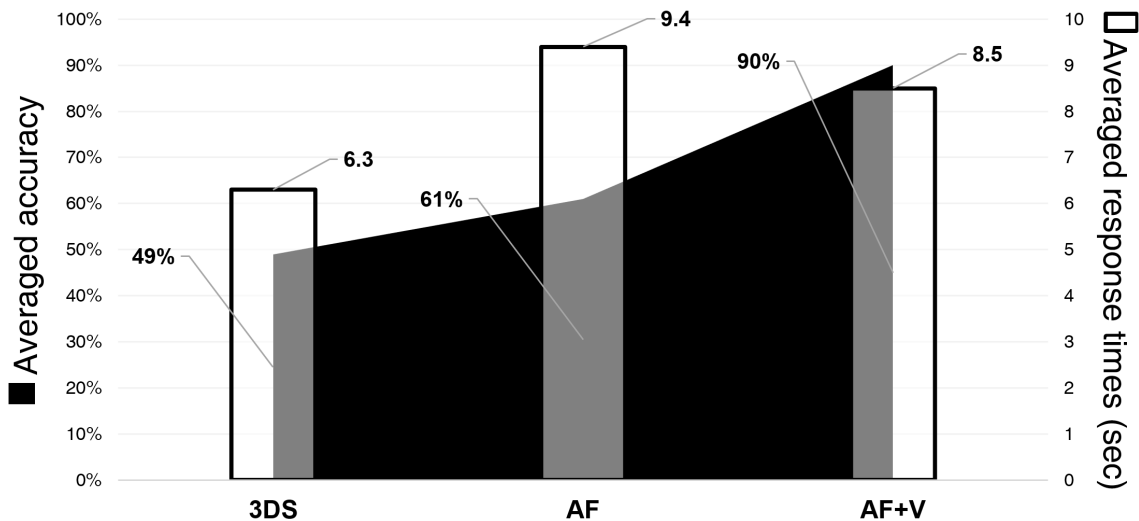


Figure 7.1 — Summary of the averaged behavioral values resulting from the statistical analysis of Accuracy (bars) and Reaction time (plain black area) measurements.

It can be seen very clearly on these curves that the use of the AF interaction modality significantly increases the accuracy in aircraft location when comparing the results with spatial sound alone. The addition of a way to better discriminate space horizontally, as proposed with this modality, can therefore be seen as an aid to ATCos when visibility conditions are poor. Moreover, when vibrotactile feedback is activated to make it easier to discriminate between aircraft on the vertical axis, the accuracy is further increased to reach an average of 1 unlocalised aircraft over 10. Incidentally, Reaction times are significantly longer when participants use AF interaction modality. As said before, these additional times, compared to the reaction times obtained with spatial sound alone, are negligible if we take into account the temporal dimension of the ATCos' working task. Indeed, control actions are most often distributed over time, and relatively spaced from each other. Time pressure comes when the frequency, and therefore the number of aircraft, increases. However, a feature such as AF must be seen as an on-demand feature. In this way, ATCos could use it when they feel they have enough time.

7.3 Subjective results

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 with spatial sound alone. 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 professional working experience to locate an airplane. When it occurred, the solutions 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 [85], 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 [35]. This is not the case with AF interaction, which seems to be acceptable for the participants who tested it. With the help of ATCos during its design process, AF interaction was designed with the consideration that ATCos often search for aircraft they are controlling in their field of view. From this perspective, AF interaction 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 [35, 124]. Our results confirm this tendency with the combined use of spatial sounds and vibrotactile feedback.

7.4 Neurophysiological, behavioral and subjective results comparison

As explained in Chapter 6, Section 6.6 (p. 132), neurophysiological measurements have been done considering *Normal* condition (without augmentation, using only spatial sound) and *Augmented* condition (using AF+V interaction modality). Results highlighted a clear advantage for the Augmented conditions, with respect to the Normal one. More precisely, as observed with behavioral and subjective results, and neurophysiological workload index analysis have confirmed such behavior. Furthermore, W_{EEG} index showed a significant negative correlation, confirming that a higher performance is followed by a lower *experienced* workload and vice-versa.

On the contrary, as subjective results showed that participants felt a lower mental demand using AF+V modality than using 3DS (see Chapter 6, Section 6.6, p. 132), neurophysiological results measures showed a trend which is statistically different from this subjective results from obtained with NASA-TLX questionnaires. Achieved results suggested that neurophysiological indicators could provide lower intra-subjects variability with respect to subjective methods (i.e., NASA-TLX), since the same trend resulted not significant for subjective measure, and significant for neurophysiological one. Therefore, it would seem that the participants were not aware of their mental load, knowing that their performance, in terms of Accuracy, was best when they used the interaction modality (i.e., AF+V) generating the greatest cognitive workload.

7.5 Limitations

Subjective results shall be interpreted with care because of some limitations in the way the questionnaire was delivered to the participants. Actually, they were not completed during or immediately after each condition (i.e., after each trial, see protocol diagram in Figure 6.5, p. 121), 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. To give a concrete example, if a participant felt 40% frustrated with 3DS modality, they might then want to re-evaluate the score when they have to evaluate AF interaction modality regarding their perceived frustration. 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. Indeed, there is potentially a bias linked to possible forgetting after experiencing the different conditions. Actually, it could have caused the loss of recency effect and affected the final results. However, this bias should impact all the conditions in an equivalent way because of the randomization of the modality presentation. In fact, the participants had the same number of trials, difficulty levels, visual conditions and modalities, but in a different order from one another.

Furthermore, and as mentioned earlier in this manuscript, the experiment which was proposed during this first campaign was more like a laboratory experiment than an ecological situation. Indeed, in order to test AF interaction modality in an exhaustive way, we have chosen to discriminate the airport environment in 5 distinct zones (see Chapter 6, Section 6.2.5, p. 113) and to place aircraft in these zones in a static way. This has allowed us to collect behavioral data in a secure way. Since the participants were free to move in order to be able to use the AF interaction without restriction, it would have been very complicated, if not impossible, to dynamically evaluate this interaction modality and collect data that could be compared between participants.

7.6 Synthesis

In this section we synthesize the hypotheses that were made in the previous Chapter 6 (Section 6.2.4, p. 113). Table 7.1 below summarizes these 3 hypotheses and reports their status (validated, partially validated or rejected) after analysis of the results. It has to be mentioned that these hypotheses only concern Accuracy measurements. We expected that Reaction time measurements would be slower because of the nature of AF interaction, asking the participants to make movements with their head, which takes some time compared to the almost instantaneous nature of spatial sound alone.

7.7 Conclusion

This research opens new perspectives to enhance HCI quality and efficiency in ATC with the final aim of improving user performance and increasing safety. 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 (3DS modality). However, adding sound information to a context (e.g., ATC) where operators are not always used to deal with it, could

Table 7.1 — First experimental campaign: Synthesis of the results regarding the hypothesis formulated in Chapter 6.

Hypothesis	Status	Conclusion
$H_{1.1}$	Validated	Accuracy results were better using AF interaction modality than using 3DS.
$H_{1.2}$	Validated	Accuracy results were better using AF+V interaction modality than using 3DS.
$H_{1.3}$	Validated	Accuracy results were better using AF interaction modality coupled with vibrotactile feedback than without.

be seen as uncomfortable. A solution to this potential drawback could be to integrate it as an on-demand feature. Furthermore, this study represents an example of a new perspective in using neurophysiological measures, in which the objective is to assess the quality of human-machine interaction by the comparison of different technological solutions, with the final aim of enhancing user's performance and increasing safety.

To conclude, 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, XR in general or video games. Some studies in visual handicap have already suggested other interaction concepts close to the AF interaction modality [119, 40, 41]. 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 VR and AR environments. Finally, Audio Focus could also be a playful concept of interaction in the field of video games.

PART III

MULTIMODAL EXPERIENCE IN ECOLOGICAL REMOTE CONTROL TOWER ENVIRONMENT

Chapter 8

Designing multimodal augmentations for an ecological RCT environment

Résumé (Français)

Seconde campagne expérimentale (Évaluation)

Méthode de conception

L'approche de conception que nous avons envisagée pour l'élaboration des 3 modalités présentées ci-après est, de la même manière que pour la conception de la modalité Audio Focus, une approche centrée utilisateur (Chapitre 5, p. 93). Ainsi, l'identification des cas d'utilisation a également été réalisée avec l'aide des mêmes professionnels du domaine. Au cours de plusieurs réunions, nous avons discuté de situations potentiellement dangereuses qui pourraient survenir dans l'exercice des contrôles aériens d'approche et d'aéroport. Nous avons ensuite sélectionné des situations identifiées lors de la première phase. Cette phase a été guidée par le réalisme des situations identifiées, que seuls les professionnels du domaine impliqués dans le processus de conception pouvaient évaluer, ainsi que par leur faisabilité technique au regard des contraintes du projet, que nous seuls, concepteurs, pouvions juger. Cette phase a été la plus longue et s'est déroulée sur plusieurs semaines au travers de multiples réunions. Enfin, la dernière phase a consisté à considérer chacune de ces situations potentiellement dangereuses, pour ensuite imaginer des solutions IHM adaptées pouvant améliorer la performance mais également l'immersion des contrôleurs. Pour cette phase, deux nouveaux contrôleurs se sont joints à nous occasionnellement, tous deux sensibilisés au domaine des facteurs humains. Cette phase s'est déroulée parallèlement à la seconde : dès que nous identifions une situation d'intérêt, nous nous concentrons sur la proposition d'une solution IHM appropriée. Finalement, à la fin de cette troisième phase, les 3 cas d'utilisation et leurs solutions IHM (c'est-à-dire les modalités d'interaction) étaient identifiées, conçues et prêtes à être évaluées (voir chapitre suivant).

Identification de situations d'intérêt et solutions IHM associées

Événements non autorisés au sol

Une situation à laquelle les contrôleurs impliqués dans le processus de conception ont fait référence à plusieurs reprises lors de nos réunions fut celle des événements non autorisés. Cette formulation fait référence à des événements anormaux du point de vue des contrôleurs, généralement des événements imprévus. La plupart du temps, il s'agit d'actions effectuées par les pilotes sans autorisation préalable. En général, ils démarrent leurs moteurs, effectuent des essais moteur ou même se déplacent sans pour autant en avoir informé les contrôleurs (au minimum). Dans des circonstances normales, le contrôleur en charge répond aux demandes des pilotes et leur autorise ou interdit l'exécution de ce type d'action. Si ces actions sont effectuées sans autorisations préalables, le risque est qu'elles engendrent des situations dangereuses car les contrôleurs n'en auront pas conscience. Il s'agit du premier cas d'utilisation considéré pour notre étude. Il a été intitulé "*Déplacement non autorisé au sol*", et abrégée en utilisant le terme SPATIAL pour rendre compte de l'aspect spatialisé tout autour de la tour de contrôle de ce genre d'événement.

Dans le but de prévenir le contrôleurs aérien d'événements de type SPATIAL, nous avons conçu la modalité SSA (pour "*Alerte Sonore Spatialisée*", ou *Spatial Sound Alert*). Il s'agit d'un avertissement sonore déclenché de manière localisée, sur l'azimut (du point de vue de la tour) sur lequel l'événement anormal en question s'est produit. La modalité SSA est ainsi conçue pour attirer l'attention des contrôleurs en direction d'une zone spécifique. De la même manière que la modalité Audio Focus, la modalité SSA nécessite d'exécuter des mouvements avec la tête. Cependant, il n'est pas requis ici d'effectuer ces mouvements dans le but de faire fonctionner la modalité. Lorsque la tête des utilisateurs est alignée avec l'azimut de l'événement spatial, l'alerte sonore s'arrête (Figure 8.3). La Table 8.1 ci-après montre comment les paramètres de conception des earcons ont été utilisés dans la conception de la modalité SSA.

IncurSION de piste

L'une des situations les plus dangereuses auxquelles un contrôleur peut être confronté est l'incurSION sur piste. Cette situation se produit lorsqu'un avion entre sur la piste alors qu'un deuxième est sur le point d'atterrir (voir Figure 8.2). Inversement, elle peut également se produire lorsqu'un avion entre sur la piste alors qu'un autre en décolle. L'incurSION de piste a été à l'origine de plusieurs incidents et accidents aériens graves dans le passé. En particulier, elle est impliquée dans le plus important et grave accident de l'histoire de l'aviation, survenu à Tenerife en 1977 [175]. Étant donné son degré de criticité élevé, les stimulations vibratoires ont été envisagées dans le but de créer une modalité relativement perturbatrice. L'incurSION sur piste est une situation prioritaire par rapport à la plupart des tâches de contrôle : si elle est sur le point de se produire, les contrôleurs doivent stopper toutes leurs tâches courantes afin de gérer rapidement et efficacement la situation, afin d'éviter à tout prix un accident. Nous désignons la modalité correspondante par RWY (pour "*IncurSION de piste*", ou *Runway IncurSION*) et l'évènement lui étant associé par RWY. Elle est basée sur la perception cutanée et l'audition.

Les stimuli vibratoires impliquées dans la modalité RWY sont présentés aux contrôleurs par

le biais de la chaise “haptique” (voir Figure 6.3). Lorsqu’un événement de type RWY est détecté par le système, le siège de la chaise vibre de manière continue à l’aide d’un signal sinusoïdal à 50 Hz. Nous avons choisi le siège de la chaise pour cette modalité dans l’idée de conserver l’association que nous avons faite au cours de la conception de la modalité Audio Focus, à savoir faire correspondre l’assise au sol (l’événement se produit toujours sur la piste), mais également car la troisième modalité présentée ici utilise le dossier de la chaise (cf. section suivante). Dès que la chaise commence à vibrer, le contrôleur doit gérer la situation. Pour ce faire, il ordonne à l’avion au sol en situation d’incursion de piste de s’arrêter sans délais, donne un ordre de remise des gaz à l’avion prêt à atterrir, tout en gérant la situation actuelle en vol (c’est-à-dire les autres avions se trouvant dans le tour de piste de l’aéroport). Cette modalité demande également aux contrôleurs d’informer le système qu’ils prennent effectivement en compte la situation en validant un bouton situé devant eux sur l’IHM radar (voir Figure 8.4). Ce bouton apparaît lorsque l’incursion de piste est détectée par le système. Simultanément, la modalité SSA est déclenchée en direction du seuil de piste pour attirer spatialement l’attention du contrôleur.

Appels radio en environnement de tour de contrôle déportée multiple

Le concept de tour de contrôle multiple génère un grand nombre de situations n’ayant pas encore de solution en raison de la nouveauté du concept et du manque d’études dont il fait l’objet. L’un des problèmes les plus fréquents dans ce contexte est la communication avec plusieurs pilotes simultanément. En effet, chaque aéroport utilise sa propre fréquence radio, que les pilotes utilisent pour communiquer avec la tour. Dans le cas des tours de contrôle multiples, les contrôleurs doivent par conséquent gérer autant de radiocommunications qu’ils contrôlent d’aéroports. Ils peuvent ainsi tout à fait recevoir un message radio d’un aéroport *A* (primaire) alors qu’ils sont déjà en communication avec un pilote d’un aéroport *B* (secondaire), cette situation correspondant par ailleurs à la plus simple des situations de radiocommunication en environnement de contrôle déporté multiple. Étant donné que la gestion des radiofréquences est un sujet important dans le contrôle aérien, nous avons choisi cette situation comme troisième cas d’utilisation. Nous l’appelons CSA (pour “*Appel du Second aéroport*”, ou *Call from Secondary Airport*), et désignons l’événement lui étant associé par CALL.

Un type particulier de stimulation vibratoire a été conçu et évalué pour cette modalité. Son principe consiste à diffuser un pattern vibratoire spécifique et reconnaissable à travers le dossier de la chaise “haptique” lorsqu’un message est reçu par le contrôleur. Nous avons différencié ce pattern vibratoire de celui utilisé pour la modalité RWY : il est situé dans le dossier de la chaise (les pilotes en communications se trouvent souvent dans les airs) et n’est pas continu (voir Figure 8.5) comme celui utilisé pour l’incursion de piste. Ce pattern consiste en des vibrations intermittentes jouées en boucle à partir du moment où une communication radio est déclenchée depuis le second aéroport, alors qu’une première était déjà en cours depuis l’aéroport principal, ceci jusqu’à ce qu’une des deux communications ne se termine. Le signal utilisé est composé de sinusoïdes de 10 ms de période, cadencées à 55 Hz et séparées par 10 ms de silence. Ainsi, dans l’esprit des utilisateurs, le second aéroport doit être associé à ce pattern vibratoire. Nous avons utilisé les paramètres des tactons [25] afin de le formaliser (voir Table 8.2).

8.1 Introduction

Like for the previous part, which is reporting on the whole scientific approach that took place to develop the Audio Focus interaction modality, we report in this part on the whole design process, the experimental approach as well as the results obtained and their discussion in relation to 3 other specific use cases. This chapter focuses on explanations of the envisaged design approach (Section 8.2), in which we explain how end-users were involved in the design process and the identification of specific use cases of particular interest for ATCos. These selected use cases are presented in detail in Section 8.3. The next Section 8.4 presents how the different modalities which are responding to the previously identified use cases have been designed, in regard to state-of-the-art principles based on hearing, touch and proprioception. The chapter concludes with a summary in short synthesis 8.4.3.

8.2 Design approach

The design approach we have considered for the development of the 3 modalities presented hereafter is, like for the design phase of the Audio Focus interaction modality (Chapter 5, p. 93), a user-centered approach. During this second experimental campaign, we applied the same process as for the first experimental campaign, however, considering that we are here interested in 3 different modalities, we have parallelized it up. Thus, the identification of use cases was once again done with the help of the same professionals in the field, as well as the design of the interactive principles envisaged to meet this need. To go further, the identification of these use cases could not have been done without the involvement of ATCos. They are the only ones with sufficient expertise to project themselves into known or plausible situations for which technological assistance could be significant: the contribution of new concepts within a professional environment cannot be made without the intervention of professionals in this field.

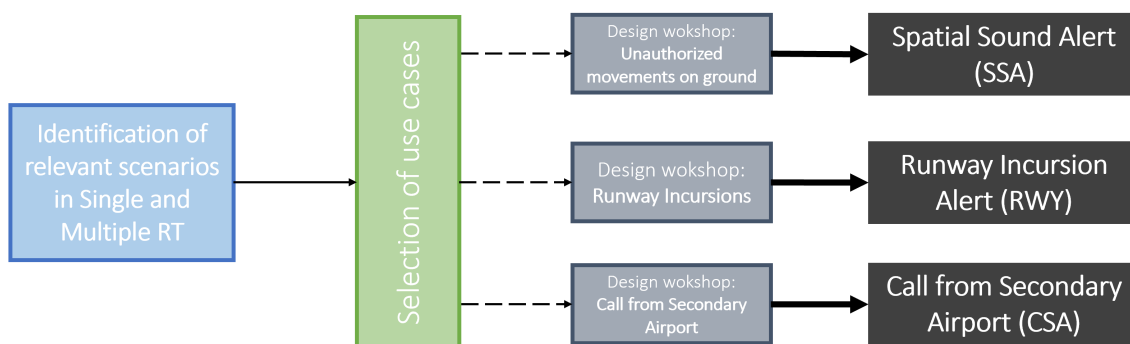


Figure 8.1 — The design process we follow for the design of the 3 additional modalities.

Through several meetings, we discussed potentially dangerous situations that could occur in the exercise of approach ATC. This phase (Figure 8.1, in clear blue) was used for the first time and to gain a better understanding of the ATCos' task as the meetings progressed. This phase took place over 3 meetings of about 1 hour each. We then selected some situations that we had discussed in the first phase (in green). This selection was guided by the realism of the discussed

situation, which only the professionals in the field involved in the design process were able to assess, and the technical feasibility of a potential reproduction of such a situation, for the needs of the experience and according to the project constraints, which only we could judge. This phase was the longest and took place over several weeks through multiple meetings. Finally, the last phase (in grey blue) consisted in considering each of the potentially risky situations previously selected to imagine adapted HCI solutions that could improve users' performance, as well as their immersion. For this phase, 2 other ATCos joined us occasionally, both of them sensitized to the field of Human Factors for ATC. It took place in parallel with the second phase: as soon as we had identified a situation of interest for professionals in the field, we focused on providing an appropriate HCI solution. Eventually at the end of this third phase, the 3 use cases and their related HCI solutions (i.e., modalities) were fully identified and theoretically designed with different expertise.

8.3 Identification of relevant ATC situations

In this section, we summarize the work done during phases 1 and 2 of the design process. Each of the subsections presents a use case of major interest to the ATCos involved, in terms of their feasibility and project constraints. We thus present 3 well-known situations: the first concerns the occurrence of unforeseen events because they are not previously approved by ATCo (Section [8.3.1](#)), the second concerns one of the most critical situations in aeronautics, namely runway incursion (Section [8.3.2](#)), and the third and last, the well-known problem of radio frequency management in the case of Multiple RCT environment (Section [8.3.3](#)).

8.3.1 Location of abnormal events

One situation to which ATCos referred several times in our interviews was that of abnormal events. This term refers to events that are abnormal from the ATCo point of view, typically unforeseen events. Most of the time, these are actions performed by pilots without prior authorization. In general, they may start their engines, perform engine tests or even move on the ground without previously having ask for permission. Under normal circumstances, ATCos respond to pilot requests and authorize or deny them the clearance to perform this type of action. This situation is the first use case considered for our study, for technical and project-related constraint it is reduced as “*Unauthorized movement on ground*”, and shortened using the term SPATIAL.

8.3.2 Runway incursion awareness

One of the most dangerous situations that an ATCo may have to deal with is the runway incursion. This situation occurs when an aircraft enters the runway while a second one is on short final and is about to land (see Figure [8.2](#)). Conversely, it may also occurs when an aircraft enters the runway while another is taking off. The runway incursion was the cause of several air incidents and crashes in the past. In particular, it is partly involved in the most significant crash in the history of aviation, that occurred in Tenerife in 1977 [\[175\]](#).¹ Many airports are already equipped with

¹The Tenerife air crash occurred on 27 March 1977 between two Boeing 747s, one belonging to the airline KLM and the other to Pan Am. They collided on the runway at Los Rodeos airport on the island of Tenerife in the Canary

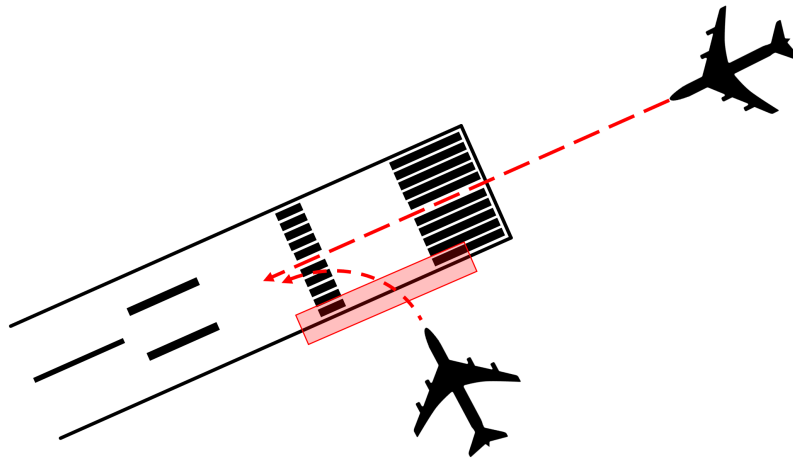


Figure 8.2 — A schematic view on a situation of type runway incursion. There, the aircraft on the right is about to land in the next seconds while another one, at the bottom, is about to cross the line between taxi routes and the runway (holding point).

systems to prevent such situations. However, this is very often not the case at small airports and aerodromes, to which our study is related. The runway incursion is then the second use case we considered for the present study. It is commonly referred as “*Runway incursion*” and abbreviate it using the term RWY.

8.3.3 Multiple radio calls in case of multiple RCT

The concept of Multiple RCT generates a large number of situations that do not yet have solutions because of the novelty of the concept and the lack of studies referred to it. One of the most common identified issue is the communication with several pilots simultaneously. Each airport uses its own radio frequency, and pilots use it for communicating with related ATC. However, in the case of Multiple RCT, ATCos must manage as many radio-communications as they are controlling airports (i.e., radio frequencies). Therefore, they may receive a radio message from an airport *A* while they are already in communication with a pilot from an airport *B*. As radio frequency management is an important topic in ATC, we have chosen this situation, which is specific to multiple RCT, for a third and final use case. We commonly refer to it as “*Call from the secondary airport*”, and abbreviate it as CALL.

8.4 HCI techniques for specific ATC use cases

HCI solutions have been designed to enable ATCos to react more effectively to the issues that are potentially generated by each of the use cases presented in the previous section. In this section we introduce these modalities, based on the state of the art previously presented.

Islands, and resulted in the death of 583 people. While the airport was surrounded with thick fog that provided no visibility for either pilots or ATCo, the pilot of the KLM 747-200 began a take-off procedure and then struck the Pan Am 747-100, which was at the same time climbing the same runway. This accident is commonly referred to as the “crash of the century” and is the most deadly accident in the history of commercial aviation.

8.4.1 Spatial alert as earcon to provide location of abnormal events

With the goal to warn ATCos from unauthorized events that could be dangerous for the smooth running of operations, we have designed the Spatial Sound Alert modality (SSA) modality. Its principle is based on an audible warning that is spatially triggered towards the azimuth to which the abnormal event occurred (e.g., unauthorized movement on ground, SPATIAL). In this way, SSA is designed to attract ATCos' attention to a specific area within the display showing the controlled airport vicinity. Like Audio Focus modality, SSA modality asks the users' to make movement with their head. However, it is not required to move the head to make the modality working because it is automatically raised. When users' head is aligned with the azimuth of the SPATIAL event, the sound alert is stopped (Figure 8.3). This principle has already been studied in the past and generally allows a finer and faster reaction, especially when the visual sense is unavailable [157]. It is also close to the concept presented in [41, 40], which was introduced in the state-of-the-art part of this manuscript (see Section 3.5.2). Table 8.1 relates how earcons design guidelines [26] were used to build this modality.

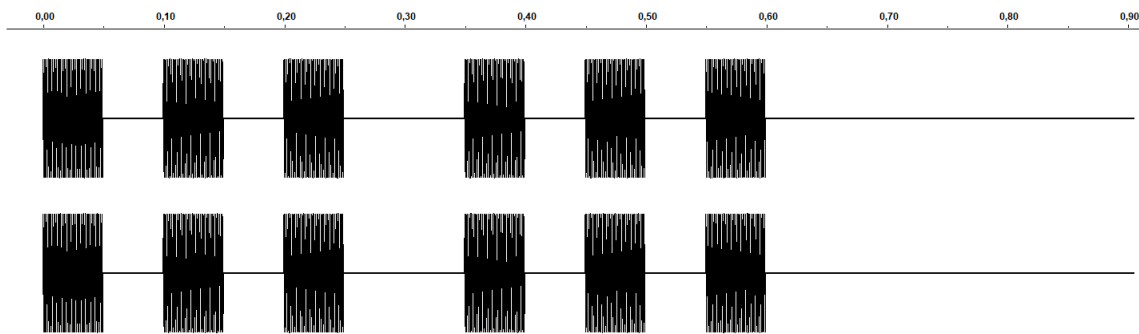
SSA modality was made of 3 A musical tones (880 Hz) with a duration of 50 ms each, in the form of sinusoid waves separated with 50 ms silences (Figure 8.3). This sequence was repeated two times with 100 ms of silence to make an earcon of type *compound* (see Section 3.3.3). This earcon was looped until the ATCos' head was aligned with the associated SPATIAL event azimuth. During the experimental phase, this modality was used in two specific situations: when an aircraft was taxiing without prior authorization (SPATIAL event), but it was also coupled with RWY event to amplify the capacity of disruption of a situation of type *runway incursion*. We applied the earcon design guidelines [26]:

8.4.2 Vibrotactile feedback to alert from critical events

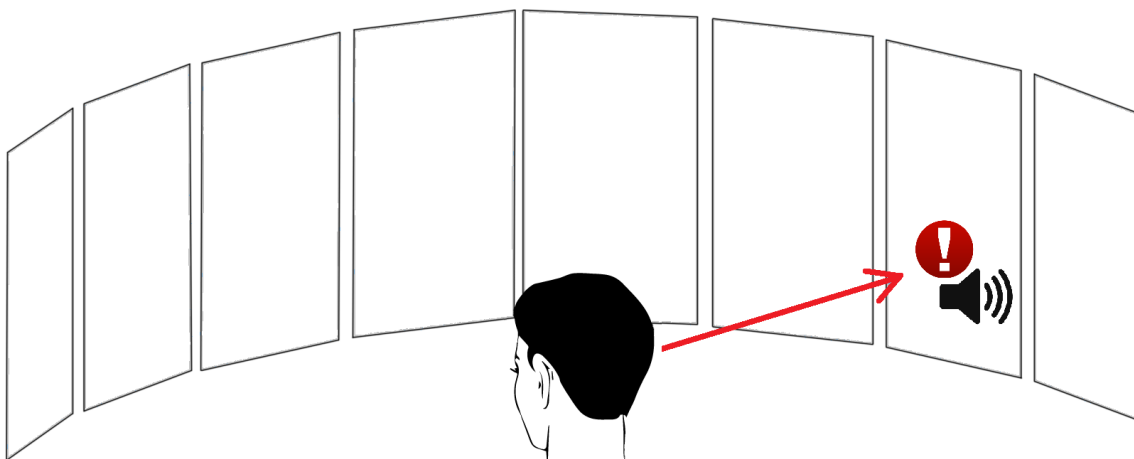
Because the runway incursion is associated with the highest degree of criticality, the related feedback should be highly disruptive. When anticipating this kind of situation, ATCos shut down all the tasks they are currently performing in order to manage quickly and efficiently the runway incursion situation and avoid an accident at all costs. Runway incursion is a priority situation for most control tasks. We simply designate the related modality with Runway Incursion modality (RWY). It relies on both haptic and audio channels.

The tactile feedback involved in the RWY modality have been presented to ATCos through the so-called "haptic" chair (as for the next modality, see Figure 6.3). When an event of type RWY is detected by the system, the seat of the chair began to vibrate continuously using a sinusoidal signal at 50 Hz. We chose the seat of the chair for this modality for two reasons. The first one is the idea to keep the association we made for the design of Audio Focus interaction modality, which is to associate the seat of the chair with the ground (of course, the aircraft that is crossing the holding point is located on the ground). The second reason is linked with the design of the third modality (i.e., CSA), that uses the back of the chair; we did not want to spread different vibrotactile patterns on the same body location (it would be quite difficult to distinguish between the two, obviously).

From the moment the chair starts to vibrate, the ATCos must manage the situation. To do so, they order the aircraft entering the runway to stop suddenly and without delay, give a go-around



(a) The wave used for the earcon related to SSA modality. The sound signal is made of two sets of three 800 Hz sinusoid waves of 10 ms separated from each other with a 10 ms silence block. The two blocks are separated with a 100 ms silence block. The scale on top is in milliseconds.



(b) When an event of type SPATIAL occurs, a spatial sound alert is raised toward its azimuth. If the users' head is not already aligned with this azimuth, the sound will be played until the users move their head towards. The sound is then stopped.

Figure 8.3 — Design of SSA modality. (a) The waveforms used for the sound signal and (b) its interaction principle.

order to the aircraft ready to land, while managing the current in-flight situation (i.e., other aircraft that may be in the airport environment). This modality also asks the ATCos to inform the system that they are effectively taking the situation into account by validating a button located in front of them on the radar HMI (see Figure 8.4). This button appears when the runway incursion is detected. At the same time, as said previously SSA modality is also activated and then spread a sound alert when the runway incursion is detected, towards its azimuth.

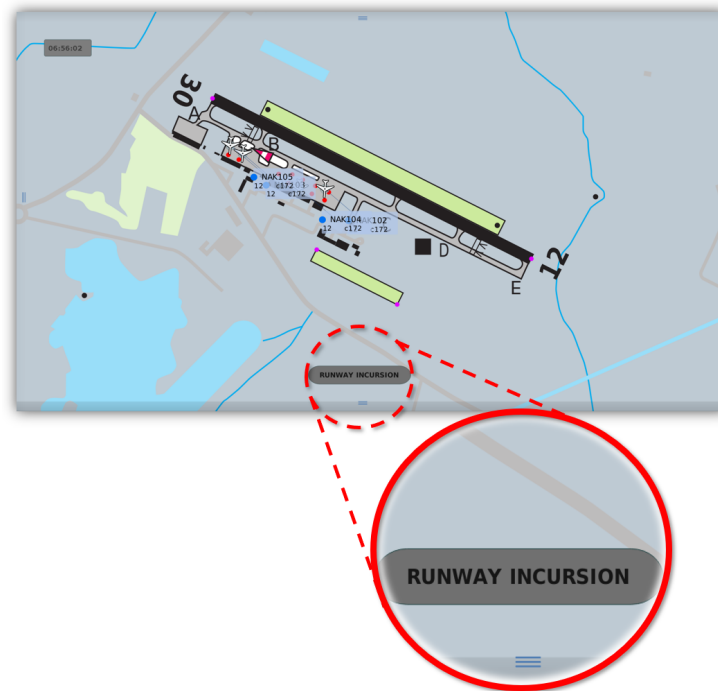


Figure 8.4 — The runway incursion button that was displayed on ground radar HMI to tell the system that the RWY event is handle. This button is part of the RWY modality.

8.4.3 Other vibrotactile feedback to distinguish between radio calls

6 In order to avoid the confusing situation that could arise when multiple and simultaneous conversations occur from two separate airports in a Multiple RCT environment, a particular type of vibrotactile feedback was designed and evaluated. We call this modality Call from Secondary Airport modality (CSA), it is associated to event of type CALL. Its principle consists in spreading a specific and recognizable vibratory pattern through the haptic chair backrest when a message is received by the ATCo from the second controlled airport. We differentiated this vibratory pattern from the one used for RWY modality: it is located in the back of the chair, while for RWY modality the seat vibrates. In addition, it is not continuous (Figure 8.5) like the one used the runway incursion, and consists of intermittent vibrations that are looped as long as a pilot located on a second airport is speaking while another one from the main airport is already in communication with the ATCo. More precisely, it is made of 55 Hz sinusoid signals of 10 ms each followed by 10 ms of silence. In this way, the second airport is associated to this vibratory pattern. We used the guidelines for the design of tactons [25] to formalize it (see Table 8.2).

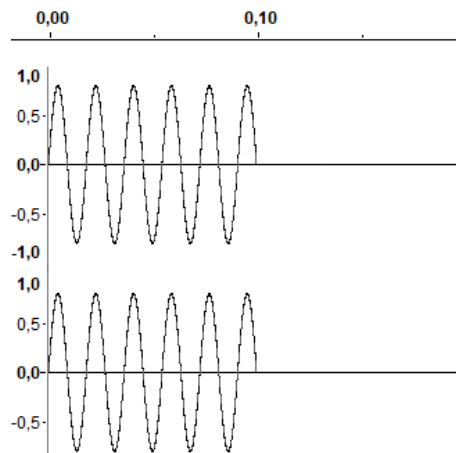


Figure 8.5 — The wave used to design a tacton involved in the CSA modality. The sound signal is made of a 55 Hz sinusoid wave àf 10 ms and 10 ms of silence. The scale on top is in milliseconds.

Wrap-up message

In this section we have presented the design process we have followed with ATC professionals, all ATCos. With their help, we were able to identify 3 problematic situations, being of high interest for ATCos. These situations are: when actions are performed by pilots without prior authorization from the tower (*SPATIAL*), runway incursions (*RWY*), which is one of the most critical situations in ATC, and the management of radio exchanges in a Multiple RCT environment (*CALL*).

For each of these situations, and always with the help of the ATCos involved in the project, we have detailed the design of 3 new modalities. These are based on the principles presented in the section related to the state-of-the-art, and allow us to respond to the problems generated by the 3 previously isolated situations. These three modalities are called:

- SSA, which allows to respond to events of type *SPATIAL* and is also coupled to the second modality, namely
- *RWY*, responding to the situation of type *RWY*, and
- *CSA*, which makes it possible to respond to the problem generated by situations of type *CALL*.

SSA is based on the earcons concept and implements kinesthetic interaction elements similar to those used for Audio Focus interaction modality. Wanting the *RWY* modality to be disruptive, it is based on continuous vibrotactile feedback. *CSA* is also based on vibrotactile feedback, more precisely formalized using the design space proposed for the concept tactons.

As in the previous section, the next chapter presents the entire experimental process used to quantify the contribution of these different modalities.

Table 8.1 — Second experimental campaign: How earcons design space [26] was considered for the design of SSA modality (see reference Table 3.1 and Figure 8.3 for an illustration).

Parameter	How it is applied
E_1 (timber)	<i>The timber is made of a sine wave at 880 Hz, which is not too complex, easy to understand and to distinguish in the RCT auditory environment.</i>
E_2 (register)	<i>This parameter is important here since the concern earcon is an auditory alert; its semantic therefore implies urgency to tell the ATCos' that a potentially dangerous situation is occurring at a specific location.</i>
E_3 (pitch)	<i>We used an A musical tone because this is a frequency that users are generally accustomed to hear.</i>
E_4 (rhythm, duration, tempo)	<i>We used one earcon with a rhythmic based on 50 ms blocks of A musical tone and silences: A — silence — A — silence — A — silence — silence, repeated twice. No other earcon was used so we had no need to use this parameter as a mean of differentiation.</i>
E_5 (intensity)	<i>SSA was audible enough to catch ATCos' attention, while being not too much disruptive as not to be too uncomfortable.</i>
E_6 (spatial location)	<i>By nature, SSA was spatial; however, the envisaged scenarios did not trigger 2 alerts simultaneously.</i>

Table 8.2 — Second experimental campaign: How tactons design space [25] was considered for the design of CSA modality (see reference Table 4.1 and Figure 8.5 for an illustration).

Parameter	How it is applied
T_1 (frequency)	<i>We used only one signal with a frequency of 55 Hz because of the pseudo infra-bass nature of the tactile transducers.</i>
T_2 (amplitude)	<i>The amplitude of the signal was configured to be as perceivable as possible, and it was also adjusted from a participant to another by directly varying the intensity level of the transducers.</i>
T_3 (waveform)	<i>Considering the low number of signal types which are perceivable by the skin, and wishing to propose a smooth signal to the participants, we chose sinusoid waveform for the tacton.</i>
T_4 (duration)	<i>One element of the tacton only last 10 ms; they are looped until the pilot from the second airport stop talking.</i>
T_5 (rhythm)	<i>The alternance of 10 ms of sinusoid signal and 10 ms of silence creates the rhythmic dimension of the tacton.</i>
T_6 (body location)	<i>Since the ATCos use their fingers for other tasks, the abdomen was not possible with the equipment we had, and the thighs were already considered for RWY modality, we retained the back to play the vibrations—which is one of the most receptive part of the body for cutaneous stimulation.</i>
T_6 (spatiotemporal pattern)	<i>The tacton is played only on a single location on the body (between the two shoulder blades) and does not have more than one spatial dimension.</i>

Chapter 9

Second experimental campaign: Evaluation of multimodal augmentations in ecological RCT environment

Résumé (Français)

Seconde campagne expérimentale (Conception d'interactions multimodales en environnement écologique)

Participants et tâche expérimentale

Seize contrôleurs aériens de nationalité Française ont participé à l'expérience (8 femmes et 8 hommes). L'âge moyen était de 39,4 ans ($SD = 7$). Comme pour la première campagne expérimentale, aucun d'entre eux n'a signalé de problème auditif. Après leur avoir expliqué de manière détaillée l'étude à laquelle ils prenaient part, ils ont été invités à prendre place. Pour cette seconde campagne, la configuration du banc expérimental a été améliorée de manière à se rapprocher le plus possible de la configuration d'une véritable salle de contrôle déportée. Grâce à l'IHM radar sol, les participants pouvaient donner des ordres simples aux pilotes (voir Figure 9.1). Un écran supplémentaire a été utilisé afin de fournir une vue minimaliste sur l'aéroport secondaire. Dans le but de rendre la plate-forme plus réaliste et vivante, deux positions de pseudo-pilotes ont été ajoutées dans une autre pièce pour disposer d'un moyen de pilotage en temps réel des avions impliqués dans les scénarios, ainsi que pour offrir aux participants (contrôleurs aérien) et aux pilotes la possibilité d'établir des communications radio réelles (voir Figure 9.1). Les scénarios expérimentaux étaient à la fois plausibles aux yeux des contrôleurs et comparables entre eux de manière à pouvoir effectuer des analyses statistiques. Ils ont été conçus pour se rapprocher le plus possible de la tâche de travail réelle des contrôleurs aérien dans un véritable aéroport. Habituellement, les participants devaient faire face à des événements communs, comme par exemple demander aux pilotes de démarrer les moteurs d'un avion, de se rendre au point d'attente, de décoller, d'atterrir ou de faire des tours de piste. Tous les scénarios ont été rédigés avec l'aide d'experts et ont été réalisés dans des conditions de mauvaise visibilité dans lesquelles aucun avion n'était visible.

Hypothèses

Globalement, un des buts du projet MOTO est d'améliorer l'immersion ainsi que le sentiment de présence en environnement de tour de contrôle déportée au moyen d'augmentations basées sur des canaux sensoriels sous-exploités tels que l'ouïe et le toucher, et ce afin d'améliorer le niveau général de sécurité. Dans cette perspective, les hypothèses de travail pour cette seconde étude peuvent être formulées comme suit :

- *Lors d'un événement de type SPATIAL, les temps de réaction et la charge de mentale perçue diminuent...*
 - *H_{2.1}: SSA en environnement simple → ...lorsque la modalité SSA est activée en environnement de tour de contrôlé déportée simple;*
 - *H_{2.1}: SSA en environnement multiple → ...lorsque la modalité SSA est activée en environnement de tour de contrôlé déportée multiple;*
- *Lors d'un événement de type RWY, les temps de réaction et la charge de mentale perçue diminuent...*
 - *H_{2.3}: RWY en environnement simple → ...lorsque la modalité RWY est activée en environnement de tour de contrôlé déportée simple;*
 - *H_{2.4}: RWY en environnement multiple → ...lorsque la modalité RWY est activée en environnement de tour de contrôlé déportée multiple;*
- *Lors d'un événement de type CALL, les temps de réaction et la charge de mentale perçue diminuent...*
 - *H_{2.5}: CSA en environnement simple → ...lorsque la modalité CSA est activée en environnement de tour de contrôlé déportée simple;*
 - *H_{2.6}: CSA en environnement multiple → ...lorsque la modalité CSA est activée en environnement de tour de contrôlé déportée multiple;*

Scénarios et scripts

De manière à proposer aux participants des scénarios à la fois réalistes, différents, mais comparables entre eux, 4 scripts distincts ont été imaginés avec l'aide des contrôleurs impliqués dans le processus de conception : Single 1 et 2, et Multiple 1 et 2. Ces quatre scripts ont servi à élaborer huit scénarios différents qui ont été présentés aux participants dans le but d'évaluer les différentes conditions : deux en environnement de type tour de contrôle déportée simple sans augmentation (*SRT* 1 et 2, pour *Single Remote Tower*), deux autres en environnement de type tour de contrôle déportée simple avec augmentations (*SART* 1 et 2, pour *Single Augmented Remote Tower*), puis de la même manière, *MRT* 1 et 2 en environnement de type tour de contrôle déportée multiple sans augmentation et *MART* 1 et 2 en environnement de type tour de contrôle déportée multiple avec

augmentations. Chacun de ces scénarios comportait le même nombre d'évènements. Les participants ont été divisés en deux groupes. La Figure 9.2 ci-après montre comment cette répartition a été gérée.

Conditions expérimentales

Ces scénarios devaient être à la fois plausibles du point de vue des contrôleurs impliqués dans l'expérience (i.e., les participants) et comparables entre eux pour permettre l'analyse des différentes modalités une fois l'expérience terminée. Nous voulions que les conditions expérimentales soient aussi proches que possible d'un environnement de tour de contrôle déportée réel, tel que celui que l'on trouve déjà dans certains endroits du monde (voir Chapitre 1, p. 13). La volonté de fournir aux participants des scénarios de contrôle différents et plausibles va dans ce sens. D'un point de vue général, les contrôleurs ont dû faire face à des situations communes auxquelles ils sont habitués. Deux variables indépendantes ont alors été considérées :

- le *Contexte*, évoluant dans le domaine de valeurs [*Simple, Multiple*];
- et l'*Augmentation*, pour l'activation ou non des modalités testées ([*Oui, Non*]).

Les scripts *Simple* et *Multiple* ont été rendus similaires en utilisant une complexité de trafic comparable et les mêmes événements à déclencher, en nombre et en types (c'est-à-dire un minimum d'un événement de chaque type dans chaque script, voir Annexe 10.8). Du brouillard a été ajouté aux visuels pour rendre les conditions de visibilité très médiocres tout au long de l'expérience (voir Figure 9.3) ; aucun avion n'était visible pendant toute la durée de l'expérience, à l'exception des avions qui étaient sur le parking jusqu'au point d'attente, juste avant de pénétrer sur la piste.

Protocole expérimental

Le protocole expérimental a été conçu dans le but de collecter des données comportementales, subjectives et neurophysiologiques par rapport aux modalités testées (SSA, RW et CSA). De la même manière que pour la première campagne expérimentale, cette seconde campagne a été divisée en deux parties principales. La première correspond à la phase de formation, au cours de laquelle une présentation générale de l'expérience a été faite aux participants, avant de leur présenter un scénario d'apprentissage de 30 min pour les familiariser avec la plateforme expérimentale, les technologies employées ainsi que les différentes modalités d'interaction. La seconde partie, cœur de l'expérience, durait environ 1h30. Après l'accueil des participants, l'équipement électroencéphalographique était ensuite installé et configuré. La phase expérimentale principale commençait peu après par un briefing du premier scénario expérimental, dans lequel la configuration aérienne (nombre, position, direction et phase de vol des différents avions actuels et futurs impliqués dans le scénario) était décrite avec précision aux participants avant de commencer. L'expérience se terminait lorsque les quatre scénarios avaient été présentés au participant (voir Figure 9.5).

Mesures

Les mesures effectuées lors de cette expérience furent nombreuses et de différentes natures. Nous avons en effet pu collecter des données :

- *Comportementales* : pour chacun des événements de chaque scénario, des *temps de réaction* ont été collectés. Un compteur était paramétré pour commencer un décompte à partir du moment où un événement (e.g., SPATIAL, RWY ou CALL) était déclenché par la configuration de contrôle du scénario courant (orchestrée par les pseudo-pilotes), et était arrêté lorsque l'événement en question était détecté, voire géré par le participant. Nous avons ainsi pu collecter des temps de réaction pour chacun des événements déclenchés, correspondant à chacune des modalités testées (respectivement SSA, RWY et CSA);
- *Subjectives*, via :
 - un questionnaire de type *NASA-TLX*, de manière à disposer de valeurs subjectives par rapport à la charge mentale, aux demandes physiques et temporelles perçues par les participants, ainsi que la performance, l'effort et la frustration ressentie, pour chacune des modalités testées;
 - une *entrevue dirigée* à la fin de l'expérience, pour collecter des données subjectives générales en termes de contribution à la précision, la conscience de la situation, le sentiment de présence et la charge cognitive perçue pour chacune des modalités;
 - un questionnaire de type *SASHA*, pour collecter des données subjectives par rapport à la charge mentale, à la conscience de la situation, au travail en équipe et à la confiance dans le système lorsque celui-ci est automatisé, au travers de 6 questions;
 - les notations d'un *expert du domaine* (*SME*, pour *Subject Matter Expert*), lui-même ex-contrôleur d'approche aux Aéroports de Paris, pour collecter des données subjectives de performance générale et de charge mentale de chacun des participants, pendant chaque scénario mais également après l'expérience;
- *Neurophysiologique*, par l'intermédiaire d'un indice de charge mentale objective calculé à partir des relevés EEG de chacun des participants, de la même manière que pour la première campagne expérimentale.

Résultats

Sur la base de nos résultats précédents (voir Chapitre 6, p. 107), la modalité Audio Focus a également été considérée pour cette expérience, sans pour autant être évaluée une seconde fois. Elle a simplement été utilisée par les participants au cours de l'expérience lorsque les trois autres modalités étaient activées dans le but d'augmenter le sentiment d'immersion. Les principaux résultats obtenus sont résumés dans le tableau 9.1 ci-après.

Table 9.1 — Seconde campagne expérimentale : Synthèse des résultats obtenus lors de la seconde campagne expérimentale concernant les modalités SSA, RWY et CSA, respectivement liées aux évènements de type SPATIAL, RWY et CALL.

Variable	Synthèse du résultat
<i>Temps de réaction</i>	Les participants ont été significativement plus rapides pour gérer un événement de type SPATIAL lorsque la modalité SSA était activée en environnement de tour de contrôle déportée simple.
<i>Performance SME</i>	En environnement de tour de contrôle déportée multiple, les scores attribués par le SME montrent une baisse significative de la performance générale.
<i>Charge mentale (SME)</i>	Lorsqu'aucune modalité n'est activée, les scores attribués par le SME montrent une baisse significative de la charge mentale. En environnement de tour de contrôle déportée multiple, comparé à l'environnement simple, les scores attribués par le SME montrent une augmentation significative de la charge mentale.
<i>Charge mentale (NASA-TLX)</i>	Lorsqu'aucune modalité n'est activée, le facteur <i>demande mentale</i> du NASA TLX suggère une baisse significative de la charge mentale.
<i>Charge mentale (EEG)</i>	En environnement de tour de contrôle déportée simple, lorsqu'aucune modalité n'est activée, l'indice EEG montre une baisse significative de la charge mentale.
<i>Résultats descriptifs</i>	La modalité RWY a été jugée par les participants comme étant, en moyenne, la modalité la plus utilisable, la plus précise et celle favorisant le mieux la conscience de la situation. La modalité Audio Focus (prise en compte uniquement dans les questions ouvertes dans les questionnaires de fin), a été jugée par les participants comme étant, en moyenne, la modalité favorisant le mieux l'immersion et la baisse de charge mentale (ressentie).

9.1 Multimodal augmentations in ecological RCT environment

Audio Focus interaction modality, presented in the previous part of this manuscript, addressed the specific use case of non-visibility in a context of remote control. However, 3 other use cases were isolated during the design process, and 3 other modalities were created to address them, namely SSA, RWY et CSA. In this chapter, we present the entire experimental process that allowed us to quantify the contribution of these 3 additional modalities, in terms of behavioral (reaction times), subjective (NASA-TLX and SASHA questionnaires, Subject Matter Expert ratings, perceived accuracy, usefulness, SA, and cognitive workload), as well as neurophysiological measures (EEG). We have also improved the experimental platform to approach the most ecological environment possible, so as to be able to offer an almost real experimental task to the participants, and therefore to take them not too far from their everyday working task.

This chapter is divided into 3 main sections. In Section 9.2 we present the method used. The group of participants (Section 9.2.1), the experimental task (Section 9.2.2), the working hypotheses (Section 9.2.3), the different scenarios which were designed to propose an ecological experimental environment (Section 9.2.4), the experimental conditions (Section 9.2.5), the entire experimental platform (Section 9.2.6) are presented in detail, before ending with a precise explanation of the envisaged experimental protocol (Section 9.2.7). Next, Section 9.3 presents the different data we collected and how we statistically processed them. Thus, we present the type of behavioral (Section 9.3.1), subjective (Sections 9.3.2 and 9.3.3), and neurophysiological measures (Section 9.3.4), as well as their statistical treatment. Finally, Section 9.4 presents all the results obtained from several angles. An analysis of performance, both behavioral and subjective, is provided in Section 9.4.1. The perceived mental load is developed in Section 9.4.2. The post-experimental questionnaires are presented in Section 9.4.5, before ending with a presentation of the neurophysiological results in Section 9.4.3. Finally, the chapter is concluded with a synthesis in Section 9.4.5.

9.2 Method

In this section we discuss the experimental aspects of this second experimental campaign. After describing the group of participants involved, we formalize the hypotheses which we tested with a specific experimental protocol. Then we give more detail about the experiment itself by presenting the different working scenarios and experimental conditions. The setup is described before ending with a detailed description of the behavioral and subjective values which were recorded during the experiment.

9.2.1 Participants

Sixteen professional ATCOs participated in the experiment. Genres were equally distributed: 8 women and 8 men, all of French nationality. The average age was 39.4 years ($SD = 7$). All of them reported no hearing problems, which is consistent with their profession that requires them to be tested regularly, as that was also the case for the first experimental campaign. After providing

them with a detailed explanation of the study and experimentation, all of them gave their consent to participate in accordance with the revised Declaration of Helsinki¹.

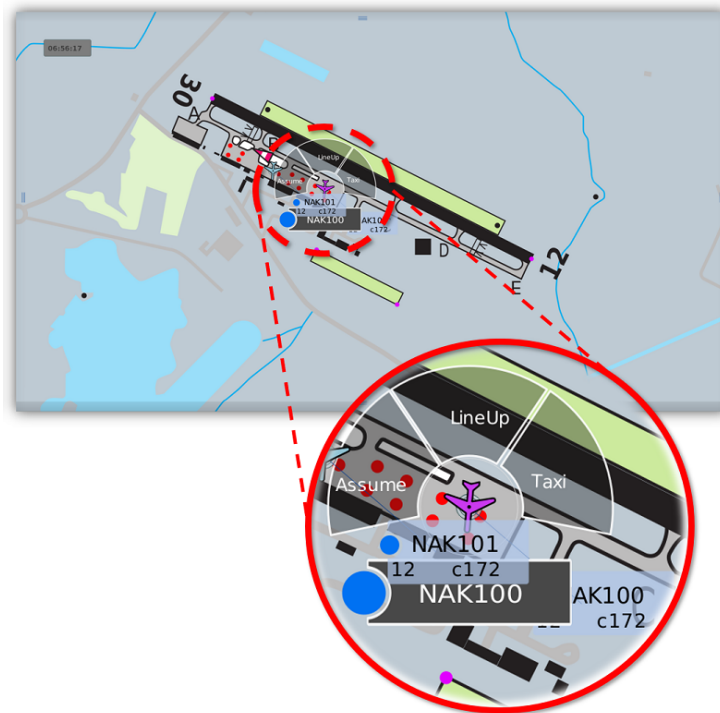
9.2.2 Experimental task overview

Participants were asked to seat on the haptic chair in front of the panoramic display wall, as for the first experimental campaign. However for this second experimental campaign, the setup was upgraded to be more comparable to the ones we can find in an operational RCT facility room (see Section 9.2.6). Using ground radar HMI, they could give simple orders to (pseudo-) pilots through a radially arranged interface that was displayed when they clicked on an airplane (see Figure 9.1). Another screen was used to have a minimalist view on the secondary airport on their right, in the case of Multiple RCT context. With the goal to make the platform more realistic and alive, two pseudo-pilots positions have been added in another room to pilot the aircraft involved in real time during the experimental scenarios, and to speak with the participant in aeronautical language (see Figure 9.1), following the scenario scripts (see Section 9.2.4). These scripts had to be both plausible for the ATCos and comparable to each other to further be able to conduct statistical analyses. They were designed to get as close as possible to the actual working conditions in a real airport. Typically, participants had to deal with common events such as pilots asking to start aircraft engines, order to reach the holding point, taking off, landings or circling the runway. They also had to manage abnormal situations such as unauthorized events on the parking area or runway incursions. All the scenarios have been written with the help of ATM experts and were conducted in poor visibility conditions where no aircraft were visible, to avoid the use of visual sensory channel, as it was made for the testing of Audio Focus interaction modality.

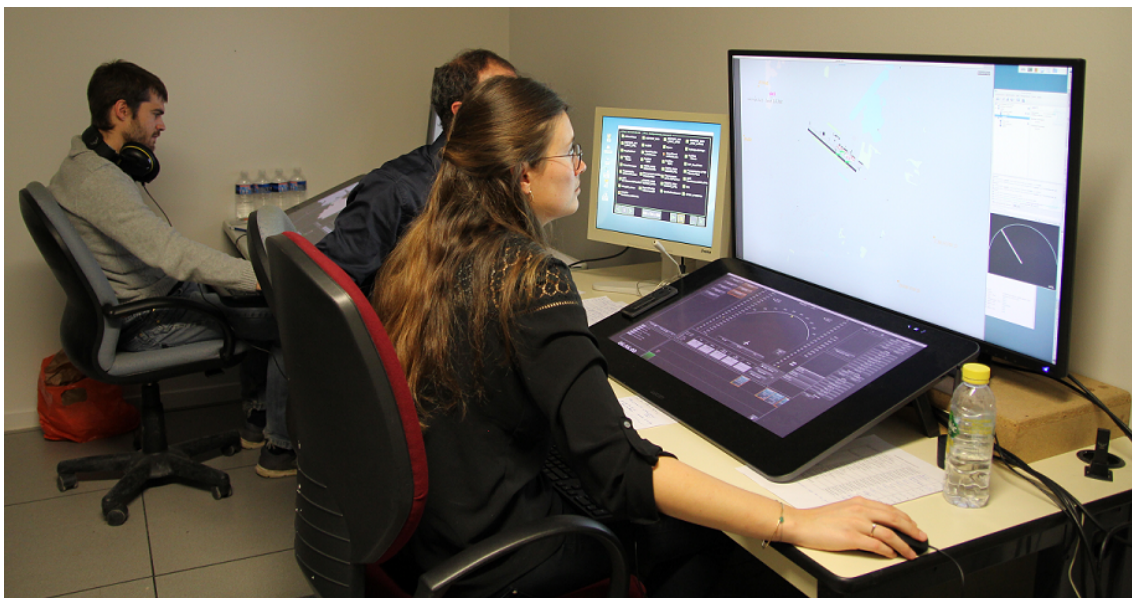
9.2.3 Hypothesis

As a result of our workshops with expert ATCos during the design process to well design scenarios and solutions, we have isolated some situations that seemed to require new tools in order for them to provide ATS more comfortably. In this way, we have designed specific interaction and feedback modalities, which we have therefore tested through the experimental campaign reported in this chapter. More precisely, the previous terms “*more comfortably*” are used here to describe a more flexible working task, giving better results in terms of cognitive workload, reaction times, or more generally trying to give ATCos a general impression of ease compared to their current working conditions. Therefore, our work aims to improve immersion and sense of presence through interactions based on under-exploited sensory channels (e.g., hearing, touch), in order to improve safety in RCT environment. In this perspective, the working hypothesis can be formulated as following:

¹See: <https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/>



(a) Using this radial HMI displayed on the ground radar when the users clicked on an airplane, they could give simple orders to pseudo-pilots during the experiment (such as clearances for taxiing and take-off).



(b) The two pseudo-pilots positions used during the second experimental campaign.

Figure 9.1 — Details on experimental task. (a) Element of the ground radar interface for ATCos to give orders to pilots. (b) The two pseudo-pilots positions.

Experimental hypotheses for SSA, RWY and CSA modalities

- *While addressing an event of type unauthorized movement on ground (SPATIAL event), reaction times and perceived workload are decreased...*
 - $H_{2.1}$: SSA in Single RCT context \rightarrow ...when SSA modality is activated under Single RCT context;
 - $H_{2.2}$: SSA in Multiple RCT context \rightarrow ...when SSA modality is activated under Multiple RCT context;
- *While addressing an event of type runway incursion (RWY event), reaction times and perceived workload are decreased...*
 - $H_{2.3}$: RWY in Single RCT context \rightarrow ...when RWY modality is activated under Single RCT context;
 - $H_{2.4}$: RWY in Multiple RCT context \rightarrow ...when RWY modality is activated under Multiple RCT context;
- *While addressing an event of type call from secondary airport (CALL event), reaction times and perceived workload are decreased...*
 - $H_{2.5}$: CSA in Single RCT context \rightarrow ...when CSA modality is activated under Single RCT context;
 - $H_{2.6}$: CSA in Multiple RCT context \rightarrow ...when CSA modality is activated under Multiple RCT context.

9.2.4 Scenarios

Four different configurations were used to assess the different remote ATC conditions:

- Single RCT context without any augmentation, or *SRT* (Single Remote Tower);
- Single RCT context with augmentations, or *SART* (Single Augmented Remote Tower);
- Multiple RCT context without augmentation, or *MRT* (Multiple Remote Tower);
- Multiple RCT context with augmentations, or *MART* (Multiple Augmented Remote Tower).

Each of these configurations had to be presented once to each participant. In order to decrease the learning effect, 4 different scripts were designed for a total of 8 scenarios. In this way, 4 different scenarios using different scripts could be used for each participant. The scripts were designated by *Single 1*, *Single 2*, *Multiple 1*, and *Multiple 2*. They were all different scripts while including equivalent operational events, with the aim to decrease potential learning effects. The same amount of event of each type was raised for each scenario. More precisely, considering that SPATIAL event corresponds to *Unauthorized movement on ground*, RWY event to *Runway incursion*, and CALL event to *Call from secondary airport* (see Appendix [10.8](#)):

- SRTs and SARTs scenarios contained a minimum of:
 - 1 scripted SPATIAL event,
 - 2 scripted RWY events each,

- MRTs and MARTs contained a minimum of:
 - 1 scripted SPATIAL events,
 - 2 scripted RWY events,
 - 2 scripted CALL event each.

It should be noted here that, since this is a “living” simulation, it was difficult to properly control the repetitions of the cases we wanted to evaluate. However, pseudo-pilots were encouraged to raise more events during the experiment. Therefore, we always had more events per participant than the scripted ones. For the analysis, we always considered the same number of events to compare the scenarios.

The 4 augmentation modalities, namely SSA, RWY, CSA and additionally, AF, were activated only during augmented scenarios (i.e., SARTs and MARTs). *Single 1* was used for SRT1 and SART1, *Single 2* for SRT2 and SART2, *Multiple 1* for MRT1 and MART1, and *Multiple 2* for MRT2 and MART2 (see Figure 9.2). AF modality was not investigated a second time during this experimental campaign, and was always activated during augmented scenarios, regarding the results we had during the first experimental campaign. However, the vibrotactile feedback used to discriminate aircraft on the vertical axis were disabled for this experimental campaign because we saw during the pre-tests that too much vibrotactile feedback were provided to the participants (i.e., considering RWY and CSA) that may led to some discomfort. During the passes, 4 different scenarios of each configuration were randomly presented to each participant, whose were divided into two groups:

- G_1 was composed of SRT1, SART2, MRT1 and MART2 scenarios;

- G_2 was composed of SRT2, SART1, MRT2 and MART1 scenarios.

Two other scripts were designed for ordering operations at the second airport (included in *Multiple* scenarios). All these scripts were followed by the pseudo-pilots during the core experimental phase to create the different situations in which they could raise the events we wanted to test. During augmented scenarios, each event was linked to the appropriate augmentation modality. For the scenarios without augmentations, these events were raised in the same way than for augmented scenarios, to make the comparison between these two types of scenario feasible. The only difference was that for scenarios without augmentations, the augmentation modalities were not activated. With this experimental design, each participant had to pass through 4 distinct scripts (see Figure 9.2 for a direct understanding).

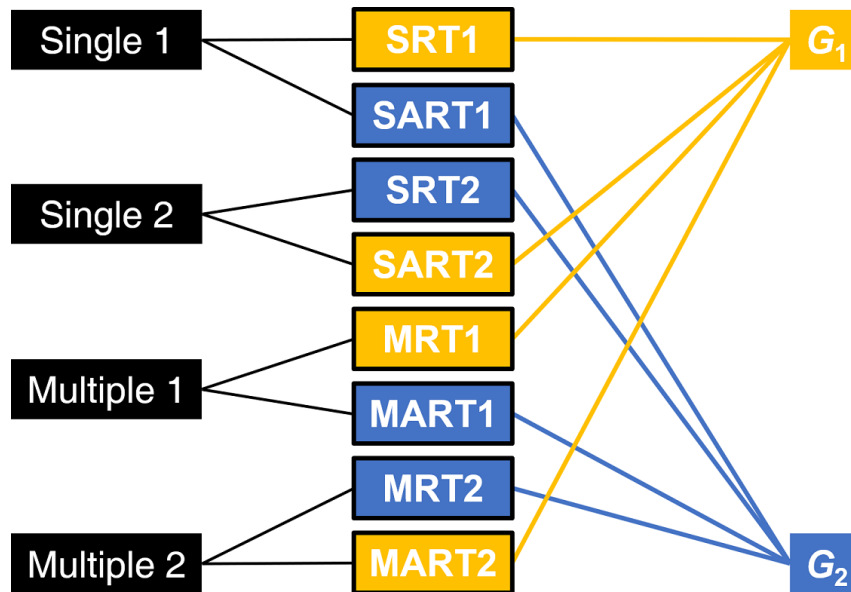


Figure 9.2 — Correspondence between scripts, scenarios and groups of participants (respectively) during the second experimental campaign.

9.2.5 Experimental conditions

These scripts had to be designed to be both plausible from the point of view of the volunteers involved in the experiment, and comparable to each other to allow us to analyse the different modalities once the experiment was finished. We wanted the experimental conditions to be as close as possible to a real RCT environment, such the one we can already find in specific places in the world (see Chapter 1, p. 13). The wish to provide different and plausible ATC scenarios to the participants goes into that direction.

From a general point of view, ATCos have had to deal with common situations to which they are accustomed. All these scenarios were written in collaboration with an ATC expert, former ATCo, in low visibility conditions, and no aircraft were visible. Within such a design, 2 exogenous parameters have been considered as independent variables:

- *Context*, that evolves in the set of values [Single, Multiple];
- *Augmented*, which could be [No, Yes].

Singles and Multiples scripts were made similar using comparable traffic complexity and events to raise (i.e., a minimum of one event of each type within each scripts, see Appendix 10.8). Fog was added to the visuals to make poor visibility conditions all along the experiment (see Figure 9.3); no aircraft were visible throughout the entire experiment, apart from aircraft which were on the parking area until the holding point, just before entering the runway.²

²The visuals used for this second experimental campaign were not made using Flight Gear open flight simulator as for the first experimental campaign, but with an home-made solution called RealTower. This tool allowed us to provide photorealistic visuals to participants and to add 3D models and visual effects. This solution was developed using Unity IDE.

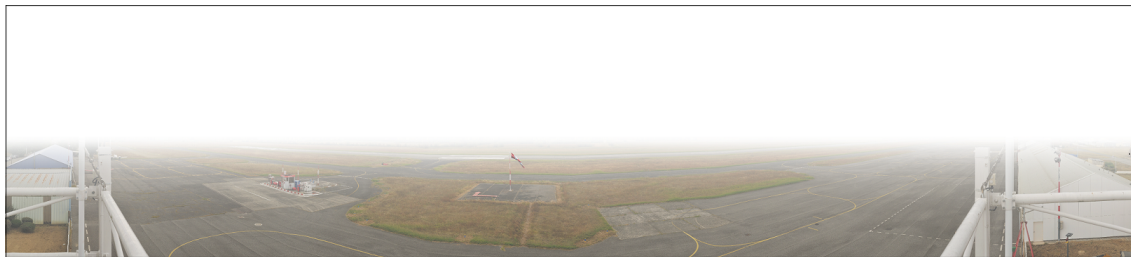


Figure 9.3 — A screen capture of the visual conditions displayed during the experiment. The fog was configured to completely cover the Muret airport runway and its environment. No aircraft were visible when they were not on the parking areas.

9.2.6 Apparatus

The experimental apparatus was quite substantial, while aiming at providing realistic working conditions to participants (see Figure 9.4). We use exactly the same experimental platform as the one used during the first experimental campaign. However, we added some features.

Two locations were needed: the ATCo position, i.e., where the participants took place, and two pseudo-pilots positions located in another room, to provide a way to control the airplanes in real time during the experiment. As for the first experimental campaign, the apparatus was composed of 8 UHD 40 screens, arranged vertically and semi-circularly to provide panoramic view on the controlled airport³. Main airport visuals were made using RealTower ad-hoc solution providing a photorealistic view on the controlled airport using photo mapping coupled with 3D animated airplane models. The secondary airport⁴ (only for MRTs and MARTs scenarios) was displayed using a 40 inches Eizo Flexscan screen. Secondary airport visuals were made using Flight Gear open flight simulator software. Radio communication with pseudo-pilots were carried out using Griffin PowerMate buttons used as push-to-talk actuators coupled with two microphones (one for each controlled airport). Participants were asked to seat on the haptic chair, located in front of the table on which the two radars were placed (see Figure 6.3). Finally, the different software modules were written in C# language under Microsoft .Net 4.6 framework. Network communications were developed using ad-hoc Ivy bus software, as for the first experimental campaign.

A pseudo-pilot position was composed of a Iiyama 40 inches Prolite X4071UHSU screen, a Wacom 24 inches tablet for the ground radar, and the same push-to-talk buttons used for participants' position, coupled with a Corsair H2100 headset, for radio communications. The only difference between the two pseudo-pilots positions was a supplementary screen for one of them in order to monitor the overall exercise.

9.2.7 Experimental protocol

The protocol (see Figure 9.5) was designed to acquire behavioral, subjective and neurophysiological measurements in relation to the feedback and interaction modalities tested, namely SSA, RWY and CSA. It was designed to include well-defined situations which were encapsulated in the sce-

³Muret aerodrome, near Toulouse, France.

⁴Lasbordes aerodrome, also near Toulouse.

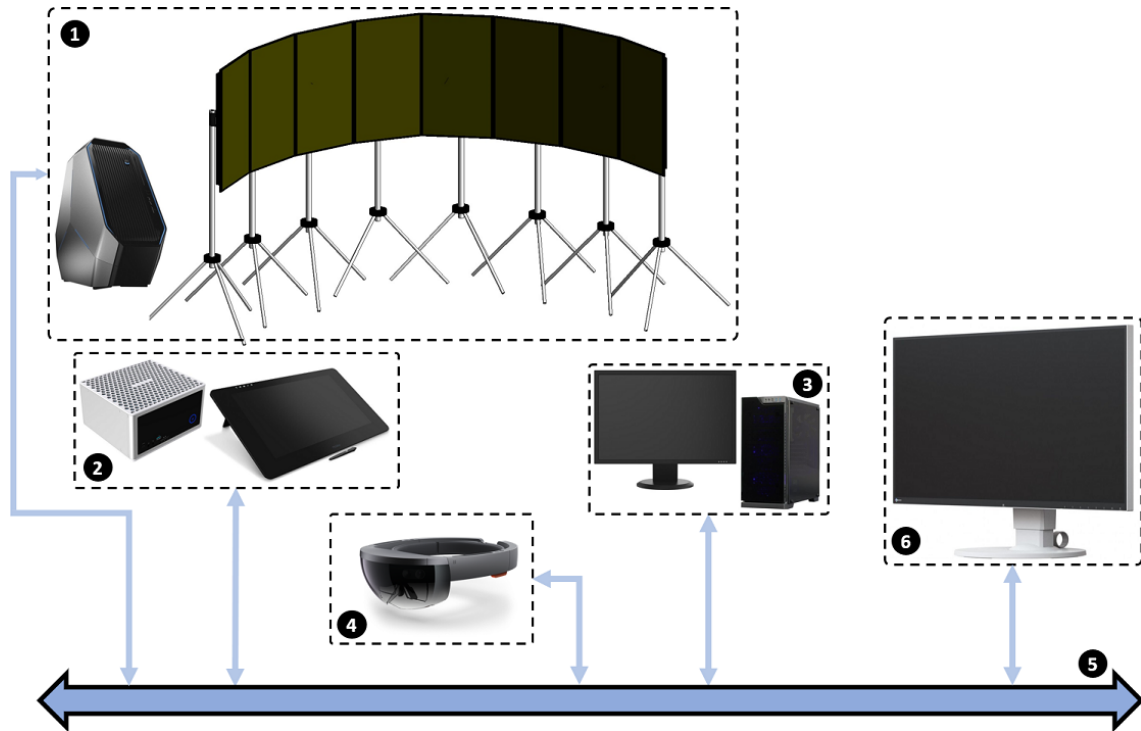


Figure 9.4 — A photo of the setup used for the experiment. The main panoramic view displays Muret airport vicinity. See Figure 6.3 for a detailed explanation. 6 is the screen used for providing a simple view on Lasbordes secondary airport, located on the right.

narios mentioned after (see Section 9.2.4). As for the first experimental campaign, the second one was divided into 2 main parts. The first was the training phase, in which a general presentation of the experiment was made to the participants, as well as a 30 minute training scenario to make them familiar with the experimental platform, employed technologies and interaction modalities. Once this training phase was completed, the participants had the opportunity to continue to familiarize with the test bench if they wished to.

Otherwise, they were considered as ready and would be invited back for the second part of the experiment. This latter lasted about 1 hour and 30 minutes, and began with a welcome of the participants and a reminder of the experiment and technologies used. Then, EEG equipment was installed and configured. A calibration phase was necessary to have minimum and maximum workload value for each participant. To do so, two calibration scenarios have been designed: *Calib-EASY*, involving 2 aircraft, and *Calib-HARD*, involving 8 aircraft. The main experimental phase itself began shortly after with a briefing of the experimental scenario, in which the air configuration (number, position, direction, and flight phase of current and future aircraft to come) was accurately described to the participant before starting each of the 4 scenarios. Then the scenario began. At the end of the four scenarios, participants were asked to complete a post-run questionnaire. The experiment ended once all the scenarios had been presented.

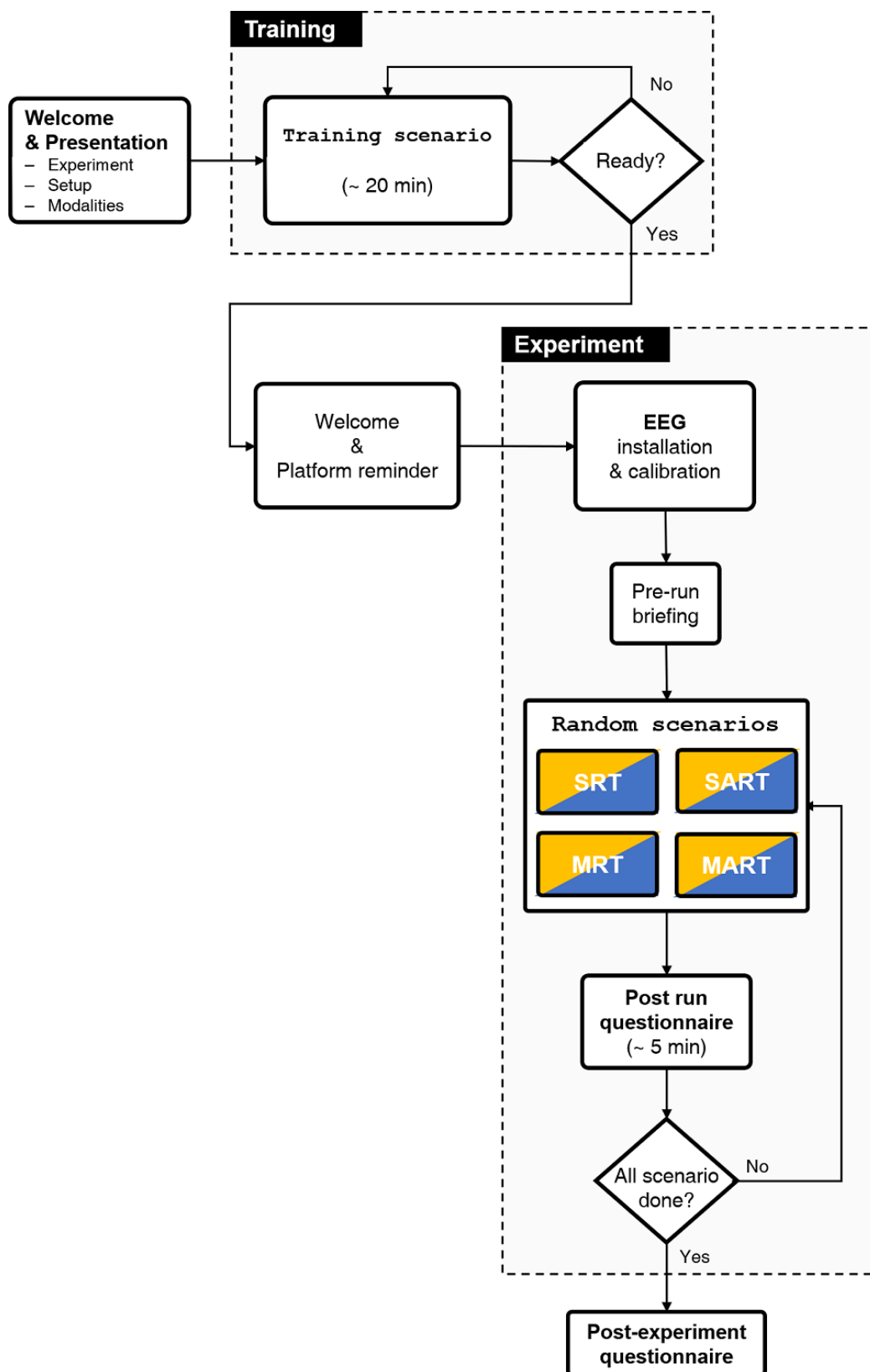


Figure 9.5 — Detailed diagram for the protocol which was designed to evaluate SSA, RWY and CSA modalities during the second experiment (if required, see Section 9.2.4 and 9.2.7 for a detailed explanation).

9.3 Measurements and data analysis

In this section we relate the different types of measurements (e.g., behavioral measurements, subjective and neurophysiological) we choose for quantifying the contribution of the 3 new modalities.

9.3.1 Behavioral data

Due to the complexity of the experimental setup, involving several people and software to work together, and the heterogeneous nature of the different modalities tested, only reaction times were acquired during the experiment as objective behavioral measurements. A timer was triggered when an event linked to a specific modality occurred until the moment when the situation was managed by the participant, in augmented scenarios or not. The Table 9.2 below resume the exact moments the timers were started and stopped for each modality.

Table 9.2 — Second experimental campaign: Start and stop times for the acquisition of reaction times during the second experimental campaign.

Modality	Start time	Stop time
SSA	When an aircraft moves on the ground without clearance.	When the participants' head is aligned with the azimuth of the event.
RWY	When an aircraft crosses the holding point while another one was on the final leg, ready to land in the next couple of seconds (this was detected by the system).	When the participants press a button on the radar interface in front of them to tell the system that they manage the situation.
CSA	When a message comes from the secondary airport.	When the participants answered to this message by clicking on the secondary airport radio communication button end starting to speak to the pilot.

Since all the scripts contained at minimum one of each event, average values were computed from each type of event, for Single and Multiple contexts, and when the augmentations were activated or not. In total, there were 4 mean values per participant for Spatial Sound Alert and Runway Incursion modalities (i.e., for SRT, SART, MRT and MART scenarios). Call from Secondary Airport modality could be tested only in Multiple RCT context. Therefore, for this particular event, 2 average values were computed for each participant (i.e., for MRT and MART scenarios).

Two-way ANOVAs with repeated measures ($CI = .95$) 2×2 (Context [Single, Multiple] \times Augmented [Yes, No]) were conducted on Reaction times measurements concerning SSA and RWY modalities. CSA modality could be activated only under Multiple conditions, so considering

this modality, a one-way ANOVA with repeated measures ($CI = .95$) (Augmented [Yes, No]) was conducted on Reaction times measurements.

9.3.2 Subjective data

9.3.2.1 NASA Task Load Index

After each of the 4 scenarios they had to pass, participants were asked to fill a post-run questionnaire (see Figure 9.6). The first part was the NASA-TLX [76, 75], which was used to estimate the cognitive workload (see Chapter 6, Section 6.3.2, p. 120). It uses six 100-points range sub-scales to assess mental workload. The *Performance* part of the NASA-TLX questionnaire is called P_{TLX} . A weighting procedure was then used to combine the 6 individual dimension ratings into a global score. This procedure required a paired comparison task to be performed prior to the workload assessments. Paired comparisons require the participant to choose which dimension is more relevant to workload across all pairs of the 6 dimensions. The number of times a dimension is chosen as more relevant is the weighting of that dimension scale for a given task for that participant. Finally, the global workload score W_{TLX} from 0 to 100 was obtained for each rated task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by 15 (the total number of paired comparisons).

9.3.3 Subject Matter Expert ratings

In addition, direct and non-intrusive evaluations were carried out by a Subject-Matter Expert (SME) during the experimental session. This SME was a professional ATCo with more than 25 years' experience who contributed to the experiment. A dedicated post-run questionnaire was designed to collect his insights about the participants' performance, including aspects related to experienced workload, stress, satisfaction/frustration and faced difficulties. More precisely, he filled the following points. For each of these, the SME provided a rate from 0 (Very Low) to 10 (Very High):

- *Overall performance during the run*, called $P_{SME_{post}}$;
- *Overall workload during the run*, called W_{SME} ;
- *Overall Situation Awareness during the run* SA_{SME} .

In addition, the SME filled simultaneously the performance reached by the participants, by using a tablet with a simple software showing a slider that he could move from 0 to 10, in order to make a real-time rating (see Figure 9.6). We called P_{SME} this subjective performance measurement.

9.3.3.1 Other subjective measurements

At the end of the experiment the participants were asked to be part of a guided interview. In this questionnaire they were asked to rate each of the augmented solutions they tested in terms of contribution to the *Usefulness*, *Accuracy*, *SA*, *sense of presence* and *Cognitive workload*. The



(a) A participant filling the post-run questionnaire.



(b) A participant answering during the second part of the guided interview.



(c) A photo of the SME using the performance measurement tool during the experiment.

Figure 9.6 — The different subjective measurements acquired during the second experimental campaign: (a) NASA-TLX measurements, (b) guided interview, and (c) SME ratings.

sense of presence and immersion was measured using a single question where the participant had to rate their perceived feeling of immersion in the RCT environment using a scale from 1 (not at all) to 10 (absolutely): *To what extent did you feel immersed in a real Control Tower environment while performing the present scenarios?*

The questionnaire that was used to measure SA is called SASHA, which is part of the SHAPE questionnaires (Solutions for Human Automation Partnerships in European ATM). SHAPE questionnaires were developed by EUROCONTROL to assess the effect of automation on controller workload, SA, teamwork and trust in the system. SASHA was originally developed to assess the effect of automation on controller situation awareness [88]. In the questionnaire, items were formulated which addressed the 3 different aspects of SA: *information extraction*, *integration* and *anticipation*. SASHA comprises 6 items, which are not assigned to individual scales. Reaction to the items were given on a 7-point Likert scale [3] ranging from “never” to “always” (see Appendix 10.8:

- “I had to search for an item of information”;
- “I was able to plan and organize my work as I wanted”;
- “I was ahead of the traffic”;
- “I was surprised by an event I did not expect”;
- “I started to focus on a single problem or a specific area of the sector”;
- “There was a risk of forgetting something important (like transferring an aircraft on time or communicating a change to an adjacent sector)”.

The participants were asked to rate their overall SA during the run, in a dedicated scale from 0 (Very poor) to 10 (Very high), and to report any specific reasons that contributed for the SA rating given by them. In the second part of this guided interview, participants were finally asked to rate the *Suitability* of each modality (including AF) in the different operational contexts. These questionnaires can be read in Appendix 10.8.

Situation Awareness was evaluated by using in-run and post-run questionnaires. Both data coming from SASHA questionnaire (SA_{SASHA}), participants ratings (SA_{given}), and SME ratings (SA_{SME}) were considered. Two-way ANOVAs with repeated measures ($CI = .95$) 2×2 (Context [Single, Multiple] \times Augmented [Yes, No]) were performed for each variable (SA_{SASHA} , SA_{given} and SA_{SME}).

9.3.4 Neurophysiological data

As for the first experimental campaign, neurophysiological measurements have been undertaken to assess participants’ mental states during the experimental phase, with the goal to achieve measures with more resolution with respect to subjective measures detailed above, and providing additional information with respect to performance-based behavioral measures (such as reaction times, in this case) [8, 18]. To simplify the investigations, neurophysiological measurements were considered only for Single scenarios (SRTs and SARTs). As explained in Section 9.2.7, 2 calibrations

scenarios were designed to calibrate the EEG signal for each participants. In these scenarios, pseudo-pilots were always asked to perform the same actions for each aircraft involved: request for engine start, then a request for reaching the holding, and a request for taking-off clearance. However the rate and number of these requests were low in *Calib-EASY* (only 3 actions required from pseudo-pilots), but very high in *Calib-HARD* (17 actions required from pseudo-pilots), in the same amount of time (3 min). As such, almost all participants reported that *Calib-EASY* scenario was really easy, while they were at their extreme limit for the *Calib-HARD* scenario. These statements have been objectively validated by the EEG measurements.

For each participant, different physiological measurements have been recorded. More specifically, EEG was recorded to compute the workload index W_{EEG} , and GSR (tonic components) was used to quantify the level of arousal. To give more details on the acquisition of EEG data, we report here that the researchers in charge of these measurements explained that “*for each ATCo attending the protocol, EEG and ECG recordings were carried out using 13 Ag/AgCl passive wet electrodes (EEG, in particular FPz, AFz, AF3, AF4, Fz, F3, F4, Pz, P3, P4, POz, PO3, PO4) referenced to both the earlobes and grounded to the left mastoid, according to the 10-20 standard [92] + 1 bipolar channel (ECG) placed on the chest of the [participant]. The device adopted in the experiment was the BEMicro (from EBNeuro Company) and the sampling rate was set to 256Hz. In addition, galvanic skin reaction activity was monitored by means of the NeXus-10MKII system (MindMedia BV, Netherlands) and its dedicated software BioTrace+ with a sampling rate of 64Hz. Biosignals have been synchronized with all the events coming from the simulation platform thanks to a specific device (Trigger Station, BrainTrends srl). In particular, at the beginning of each experimental condition, the simulation platform sent a specific trigger to both the two biosignal amplifiers, in order to be able to synchronize all the information offline*”.

In addition, they reported that “*once all the biosensors were placed and before starting with the experimental scenarios execution, each [participant] was asked to perform short recordings, to be used as calibration for neurophysiological indexes computation. In particular, 1 minute with eyes closed, to compute the Individual Alpha Frequency [96] [...], and two recordings of 3 minutes each, of [Calib-EASY] and [Calib-HARD] [...] scenarios, to be used as a calibration for the algorithm used to evaluate the EEG-based workload index [8] [...]. After these calibration recordings, the [...] experimental scenarios were run in a random order and the EEG, ECG and GSR signals have been recorded simultaneously during their execution*”.

Two-sided Wilcoxon signed rank tests were performed for the statistical comparison. Data came from multiple observations for each participant, however we were not able to robustly assess that the distribution of our measurements were Gaussian. Therefore, paired non-parametric tests have been used.

9.4 Experimental results

In this section we report the results of the various statistical analyses previously presented. Regarding the different performance measurements:

- perceived performance from NASA-TLX questionnaire P_{TLX} ,
- subjective performance from SME ratings $P_{SME_{post}}$,
- subjective real-time performance from SME ratings P_{SME} ,

and workload measurements:

- perceived workload from NASA-TLX questionnaire W_{TLX} ,
- subjective workload from SME ratings W_{SME} ,
- neurophysiological workload computed from EEG signals W_{EEG} ,

we have organized the results along 4 sections: Section 9.4.1 relates the performance results through a behavioral point of view (Section 9.4.1.1) and a subjective point of view (Section 9.4.1.1), Section 9.4.2 presents the results concerning the perceived workload, Section 9.4.3 the results concerning the workload computed from neurophysiological data and finally, Section 9.4.5 presents the results concerning the post-experiment interview.

9.4.1 Performance analysis

9.4.1.1 Behavioral results

Statistical analyses are reported in Table 9.3 below. Concerning SSA modality, the analysis revealed a main effect of Augmented factor and a Context \times Augmented interaction. However, no results were found concerning the two other modalities (see Figures 9.7 and 9.7).

Table 9.3 — Second experimental campaign: Results from 2×2 ANOVA (CI = .95) with repeated measures (Context [Single, Multiple] \times Augmented [No, Yes]) conducted on Reaction time measurements (SSA modality).

Variable	ddl	F	p	η_p^2
Context	1,14	.85	.37	.06
Augmented	1,14	7.27	< .05	.34
Context \times Augmented	1,14	6.92	< .05	.33

Post-hoc analysis was conducted using Tukey's HSD tests when a main effect or an interaction between two factors was revealed. At first sight, it seems that SSA modality allowed ATCos to resolve unauthorized movements on ground more quickly: under augmented scenarios, they were faster to manage these specific situations ($M = 18455.54$; $SD = 4217.22$) than when SSA modality was not activated ($M = 41438.31$; $SD = 7733.1$). More precisely, post-hoc analysis conducted after Context \times Augmented interaction revealed that when they were under Single RCT environment, ATCos were significantly faster to resolve these kind of events using SSA modality ($M = 9820.64$; $SD = 1276.44$) than without ($M = 58103.2$; $SD = 14813.18$, see Figure 9.7).

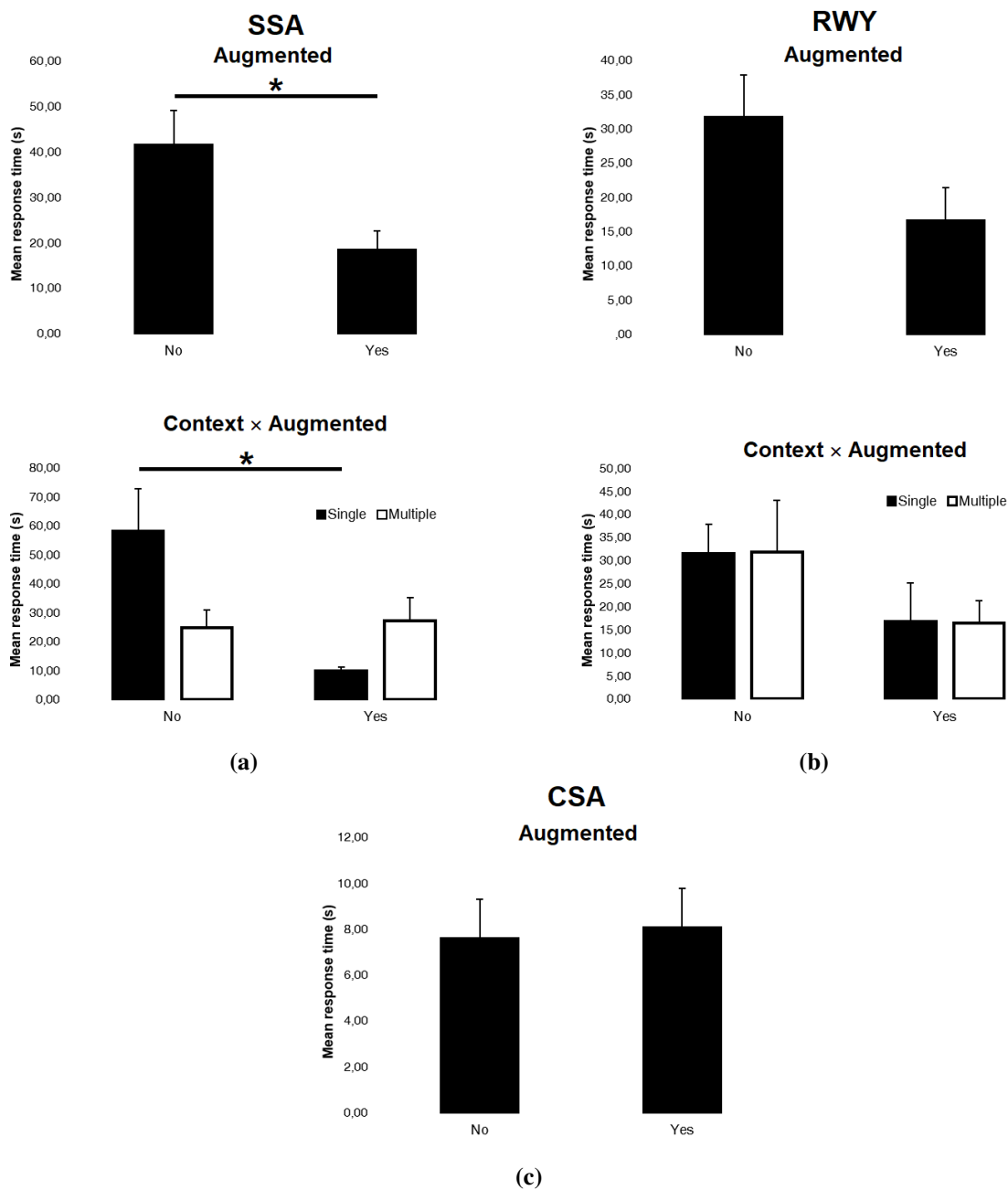


Figure 9.7 — Second experimental campaign — Results from 2×2 ANOVA ($CI = .95$) with repeated measures (Context [Single, Multiple] \times Augmented [No, Yes]) conducted on Reaction time measurements for (a) SSA and (b) RWY modalities — (c) Results from ANOVA ($CI = .95$) with repeated measures (Augmented [No, Yes]) conducted on Reaction time measurements for CSA modality. Error bars are standard errors.

9.4.1.2 Subjective results

This behavioral measurements analysis of performance was then supported by the subjective measurements analysis of performance, which was conducted to confront:

- (i) the perceived performance P_{TLX} by each participant using NASA-TLX factors,
- (ii) to the SME performance ratings, both during each scenario P_{SME} , and after each scenario $P_{SME_{post}}$.

Two-way 2×2 ANOVA (CI = .95) with repeated measures (Context [Single, Multiple] \times Augmented [No, Yes]) on the normalized performance values has been conducted using Arcsin transform [181] for each of these variables. P_{TLX} (resp. $P_{SME_{post}}$) did not show any significant difference [$F = .00034$; $p = .98$] (resp. [$F = 1.53$; $p = .02$]). However, a significant trend on P_{SME} highlighted a decrease of performances when no modality was activated [$F = 5.98$; $p < .05$]. Duncan post-hoc analysis showed that augmentation modalities induced a significant decrement in overall performance ($p < .05$) under Multiple airport contexts (i.e., MARTs scenarios). Besides, no result was highlight under Single airport contexts (i.e., SARTs scenarios).

To conclude on performance results, despite finding no result concerning RWY and CSA modality, SSA modality seems to enhance the operators' performance by reducing the reaction times when ATCos had only one airport to manage. In addition, no significant result were found concerning subjective measurements. In particular, the only observed result is that SME highlighted a significant decrease in performance under multiple airport contexts when augmentation modalities were activated.

9.4.2 Perceived workload analysis

NASA-TLX subjective workload scores W_{TLX} and post-run SME ratings W_{SME} were considered for the evaluation of perceived workload. As done before, two-way ANOVAs 2×2 (CI = .95) with repeated measures (Context [Single, Multiple] \times Augmented [No, Yes]) on the normalized workload values with the same Arcsin transformation [181] were performed. A significant main effect was highlighted for each factor. It means that for each factor, the same results were found both for W_{TLX} and W_{SME} : the perceived workload was lower when the interaction modalities were disabled ($p < .05$). Moreover, in multiple airport context (i.e., MRTs and MARTs scenarios), the perceived workload was higher with respect to single airport contexts (i.e., SRTs and SARTs scenarios) ($p < .05$). However this second result was not replicated with NASA-TLX values in which no significant difference were found regarding Multiple and Single airport contexts. No other significant main effect was found.

To sum up, subjective workload measurements showed that i) Multiple airport context (i.e., MRTs and MARTs scenarios) induced higher workload with respect to the single airport context (i.e., SRTs and SARTs scenarios), and ii) the augmentation modalities generate a lower perceived workload when augmentations were disabled. Finally, no significant difference have been highlighted elsewhere.

9.4.3 Neurophysiological analysis

After having computed W_{EEG} index for each SRT and SART scenario and for each participants, statistical analysis showed a significant increase of the W_{EEG} index during SART scenarios execution (SRT = .004; $SD = .09$; $p < .05$) compared to the SRTs scenario (SART = .1; $SD = .15$). These results are show on Figure [9.8](#).

9.4.4 Situational Awareness

Results did not show any significant trend across different conditions and modalities, from both subject (SA_{given} : $F = 1.64$; $p = .22$) and SME (SA_{SME} : $F = 0.27$; $p = .61$) sides. A similar result has been achieved by considering the SASHA questionnaire felt by each subject at the end of each condition, in particular, the ANOVA did not show any significant trend (SA_{SASHA} : $F = 1.18$; $p = .293$).

9.4.5 Post-experiment interview analysis

In the post-experiment questionnaire, participant ATCos were asked to rate the 4 augmentation modalities (including AF) along a scale going from 1 (strongly disagree) to 5 (strongly agree), regarding the 5 following assertions:

- Perceived contribution to usefulness: “*The current augmentation modality is a useful aid for RCT operations*”,
- Perceived contribution to accuracy: “*The current augmentation modality is accurate enough to support you during the RCT operations*”,
- Perceived contribution to situation awareness improvement: “*The current augmentation modality improves your situation awareness in RCT operations*”,
- Perceived contribution to the sense of immersion: “*The current augmentation modality improves your sense of immersion in RCT operations*”,
- Perceived contribution to workload reduction: “*The current augmentation modality does not have a negative impact on your workload in RCT operations*”.

By reading these post-experiment interview descriptive results, we can tell that from a general point of view the different augmentation modalities have been well received by the participant ATCos. In particular, the one that was felt the more useful, accurate and providing the better support to SA improvement was the RWY modality (see Figure [9.9](#)). Besides, from a descriptive point of view, on average the different interaction modalities received a score higher than 2.5 over 5 but we cannot strictly tell that they were perceived to greatly contribute to the sense of immersion and to workload reduction. Nevertheless, AF modality received the highest scores concerning these two points, with respectively 3.5 and 3.6 on average.

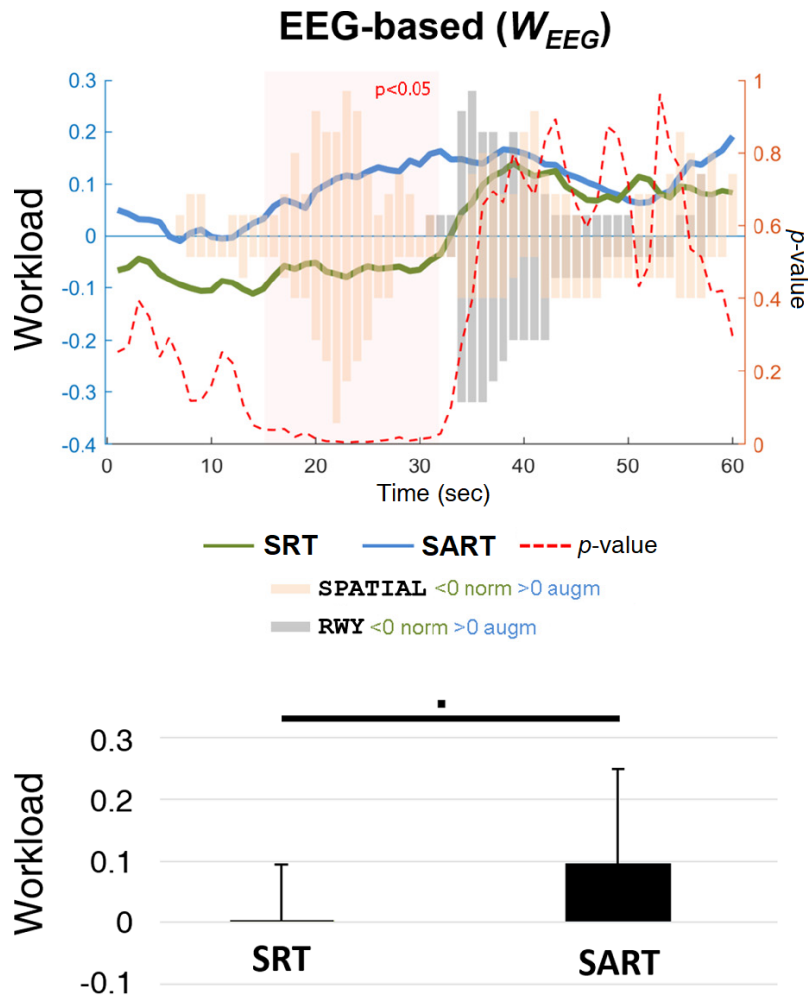


Figure 9.8 — W_{EEG} score exhibited a significant increase in the experienced workload during SART scenarios compared to SRTs scenarios, which is consistent with the results provided by the SME workload assessment. On top, a grand average of W_{EEG} workload index is reported, over all the participants, for each of the Single experimental scenarios. Orange and grey rectangles are respectively examples of events of type SPATIAL and RWY, occurring on average for all the participants. At the bottom, significant p value have been marked using a black line. Values showed are means and standard deviations.

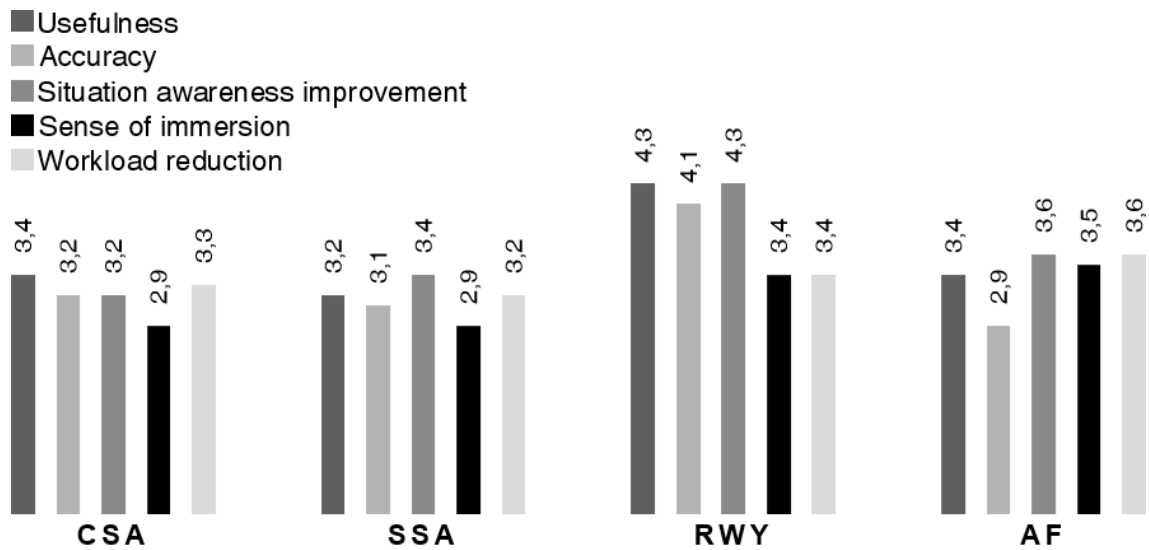


Figure 9.9 — Subjective results from the post-experiment questionnaire.

Wrap-up message

In this chapter we presented the updated ecological platform, experimental protocol, data analysis choices and results concerning the 3 remaining modalities based on touch, hearing and kinesthetics which were designed to answer the 3 remaining use cases previously identified (see Chapter 8, p. 145). Associations between these modalities and these use cases were:

- SSA, associated with unauthorized movement on ground, i.e., SPATIAL events;
- SSA, associated with runway incursions, i.e., RWY events;
- CSA, associated with call from secondary airport, i.e., CALL events.

Based on our previous results (see Chapter 6, p. 107), Audio Focus modality was also considered, however we did not evaluate it a second time. Sixteen professional ATCos have been invited to pass through our experimental protocol, mainly based on 4 different but comparable scenarios, with augmentations (i.e., AF, SSA, RWY and CSA activated) or without:

- Single Remote Tower context without augmentation: *SRT*;
- Single Remote Tower context with augmentations: *SART*;
- Multiple Remote Tower context without augmentation: *MRT*;
- Multiple Remote Tower context with augmentations: *MART*.

In addition, we have formally stated our working hypotheses, in the same way as for the first experimental campaign (see $H_{2.1} \dots H_{2.6}$, Section 9.2.3). Behavioral (Reaction times), subjective (NASA-TLX and SASHA questionnaires, SME ratings, and perceived accuracy, usefulness, SA, and cognitive workload), and neurophysiological data (mainly EEG) have been collected to try to assess the contribution of these 3 modalities (i.e., SSA, CSA and RWY) for Single and Multiple RCT environment. The main results are summarized in the Table 9.4 below.

Table 9.4 — Second experimental campaign: Summary of the results obtained during the second experimental campaign regarding SSA, RWY and CSA modalities. .

Category/variable	Synthesized result
<i>Reaction times</i>	SSA modality decreases reaction times to solve an event of type SPATIAL in Single RCT environment.
P_{SME}	SME ratings highlighted a decrease of performance when no augmentation is activated in Multiple RCT environment.
W_{SME}	SME ratings highlighted a decrease of workload when no augmentation is activated. SME ratings highlighted an increase of workload in Multiple RCT environment compared to Single RCT environment.
W_{TLX}	NASA-TLX factors highlighted a decrease of workload when no augmentation is activated.
W_{EEG}	EEG computed index highlighted a decrease of workload when augmentations are activated in Single RCT environment.
<i>Descriptive results</i>	RWY modality has been viewed by the participants as the most useful, accurate and the one best supporting SA. AF modality has been viewed by the participants as the one providing the best immersion and workload reduction.

Chapter 10

Second experimental campaign: Discussion

Résumé (Français)

Seconde campagne expérimentale (Discussion)

Les résultats obtenus concernant la performance montrent que les modalités d'interaction testées ont été bien reçues par les utilisateurs finaux (contrôleurs aérien volontaires pour participer à l'expérience). Conçues pour répondre à des situations problématiques, voire dangereuses, identifiées lors de la phase de conception sous la forme de trois cas d'utilisation distincts, nous avons montré expérimentalement que certaines de ces modalités d'interaction semblent être adaptées aux besoins des contrôleurs, dans le sens où elles engendrent de meilleures performances en termes de temps nécessaire pour gérer la situation correspondante. C'est clairement le cas de l'Alerte Sonore Spatialisée (modalité SSA), permettant aux contrôleurs d'être plus réactifs lorsqu'un événement non autorisé se produit au sol. Dans ce cas particulier, les participants ont réagi trois fois plus vite lorsque la modalité SSA était activée en environnement de tour de contrôle déportée simple. Par ailleurs, aucun résultat n'a été obtenu en environnement de tour de contrôle déportée multiple, ce qui signifie également qu'il n'y a pas eu de diminution significative de la performance.

Cependant, nous notons que les scores de performance attribués par l'expert (SME) sont significativement plus faibles lorsque les modalités d'augmentation étaient activées en environnement de tour de contrôle déportée multiple. D'autre part, l'analyse de performance perçue globale n'a mis en évidence aucune différence significative lorsque les augmentations étaient activées ou non, en environnements de tour de contrôle déportée simples ou multiples. Ces résultats sont assez difficiles à interpréter. D'une part, nous disposons d'une mesure montrant que la performance globale diminue lorsqu'il n'y a pas d'augmentation dans le cas très particulier des tours de contrôle déportées multiples. En revanche, les analyses statistiques ne montrent pas d'autre différence, que ce soit par rapport aux données subjectives que par rapport aux scores attribués par le SME lorsque les augmentations étaient activées en environnement de tour de contrôle déportée simple. La seule interprétation que nous pourrions proposer est que l'environnement de tour de contrôle déportée multiple est encore nouveau pour les contrôleurs : l'ajout de nouvelles modalités ne dégrade pas significativement leurs performances globales dans un environnement connu (tour de

contrôle déportée simple), mais cela peut les perturber lorsque leur environnement de travail devient partiellement inconnu et plus complexe (ou du moins dans lequel il n'ont que très peu voire aucun repère(s)), comme cela est le cas en un environnement de tour de contrôle déportée multiple

Par ailleurs, une tendance a pu être observée concernant une augmentation potentielle de la charge mentale perçue lorsque les modalités d'interaction étaient activées. Cependant, l'analyse des mesures subjectives de la charge mentale (SME et NASA TLX) montre que cette donnée (subjective, perçue) est inférieure lorsque les augmentations étaient désactivées. Les mesures faites par le SME montrent également que la charge mentale perçue est supérieure en environnement de tour de contrôle déportée multiple. Ces résultats, qui étaient attendus en raison du degré de complexité relativement élevé du protocole expérimental, peuvent également s'expliquer par un manque d'expérience des contrôleurs. Ils n'ont en effet pas l'habitude de travailler dans un environnement dans lequel l'audition et le toucher sont presque autant stimulés que la vision. De plus, des études montrent que ce type de concept d'interaction peut améliorer significativement les performances (les résultats relatifs à la modalité Audio Fous en témoignent). Il peut donc y avoir un problème de familiarisation. Quoiqu'il en soit, cela reste hypothétique à ce stade de l'étude. De plus, les chercheurs chargés de l'analyse des mesures neurophysiologiques ont conclu que :

- (i) Les modalités d'interaction proposées ont induit une hausse locale de la performance des participants lorsque les modalités d'interaction étaient activées. La performance globale ne semblent quant à elle pas être affectées par ces modalités d'interaction.
- (ii) Les modalités d'interaction ont induit une hausse de la charge mentale des participants. Quoiqu'il en soit, cette augmentation reste acceptable car elle n'a pas eu d'impact négatif sur les performances (qui ont en fait été améliorées) et ne doit être considérée que comme une conséquence de l'engagement plus important des contrôleurs lorsque les modalités étaient activées. Cet effet comportemental est en phase avec les résultats physiologiques obtenus en termes d'éveil (point suivant).
- (iii) Les modalités d'interaction ont induit une hausse du niveau d'éveil (*awareness*) des participant, ce qui représente un point positif pour l'évaluation globale de la performance.

Ils ont finalement conclu que “[...] *bien que le niveau de performance global des [participants] ne semble pas avoir changé de manière significative, les modalités d'interaction proposées ont induit une augmentation locale des performances. Quoiqu'il en soit, les événements liés aux scénarios [augmentés], en particulier ceux liés à la modalité SSA, ont induit une charge mentale plus importante, identifiée à l'aide de l'indice W_{EEG} . Même si cette augmentation de la charge mentale n'a pas entraîné de diminution des performances, il convient de souligner les effets négatifs possibles d'une charge mentale supérieure à la normale, comme par exemple la fatigue [sur la tâche de l'opérateur].*”

Avant de conclure, nous synthétisons dans le Tableau 10.1 ci-après les hypothèses que nous avons formulées dans le chapitre 9 précédent (p. 157).

Cette expérience nous a permis de valider partiellement nos hypothèses. En particulier, l'un des principaux résultats est que certaines modalités peuvent améliorer les performances, du moins

Table 10.1 — Deuxième campagne expérimentale : Synthèse des résultats par rapport aux hypothèses formulées au Chapitre 9

Hypothèse	Statut	Conclusion
$H_{2,1}$	Validée	Les temps de réaction ont été plus courts lorsque les participants ont utilisé la modalité SSA en environnement de tour de contrôle déportée simple. L'indice W_{EEG} indique également une diminution de la charge mentale lorsque les modalités d'interaction étaient activées en environnement de tour de contrôle déportée simple. Toutefois, comme la moyenne a été calculée pour toutes les modalités, nous ne pouvons pas réellement déterminer quelle modalité contribue spécifiquement à cette réduction de la charge de travail.
$H_{2,2}$	Partiellement validée	La modalité SSA n'a pas amélioré les temps de réaction en environnement de tour de contrôle déportée multiple. Cependant, même commentaire que pour $H_{2,1}$ concernant l'indice W_{EEG} .
$H_{2,3}$	Partiellement validée	La modalité RWY n'a pas amélioré les temps de réaction en environnement de tour de contrôle déportée simple. Même commentaire que pour $H_{2,1,2,2}$ concernant l'indice W_{EEG} .
$H_{2,4}$	Partiellement validée	La modalité RWY n'a pas amélioré les temps de réaction en environnement de tour de contrôle déportée multiple. Même commentaire que pour $H_{2,1,2,2}$ concernant l'indice W_{EEG} .
$H_{2,5}$	Partiellement validée	La modalité CSA n'a pas amélioré les temps de réaction en environnement de tour de contrôle déportée simple. Même commentaire que pour $H_{2,1,2,4}$ concernant l'indice W_{EEG} .
$H_{2,6}$	Partiellement validée	La modalité CSA n'a pas amélioré les temps de réaction en environnement de tour de contrôle déportée multiple. Même commentaire que pour $H_{2,1,2,5}$ concernant l'indice W_{EEG} .

en termes de réactivité. C'est clairement le cas pour la modalité SSA. En outre, le message sous-jacent est que l'introduction de nouvelles modalités d'interaction, basées sur l'audition et le toucher, ne diminue pas nécessairement la charge mentale des contrôleurs. Elle n'est pas significativement améliorée en environnement de tour de contrôle déportée simple, mais nous n'avons cependant pas été en mesure de démontrer une diminution, ne restant dès lors qu'au stade de l'hypothèse. Quoiqu'il en soit, il semble que ces types de modalités d'interaction n'affectent pas les performances, et que le présent travail a démontré l'adéquation des mesures neurophysiologiques lors de la phase de validation des nouvelles solutions d'interaction en environnement opérationnel de contrôle à distance. D'autres études sont cependant nécessaires pour étudier plus précisément les impacts potentiels de l'introduction de nouveaux concepts d'interaction dans le contexte des tours de contrôle déportées.

10.1 Introduction

The work reported in these two previous chapters aimed at properly designing adapted multimodal interaction and feedback and at studying their contribution in an ecological RCT environment. One of the aims of this work was also to design novel forms of interaction for the context of Single RCT, but also for the one of Multiple RCT, which is still poorly studied in HCI literature. Knowing that the most commonly used sense among ATCos is vision [179], we have focused our work on interaction and feedback modalities that used the senses of passive perception, hearing, and proprioception, in order to prevent potential cognitive overload related to the visual channel, like for the study of Audio Focus interaction related in the previous part, and to provide solutions for specific situations in Remote Tower environment.

This last chapter discuss the results obtained after the second experimental campaign (see Sections 9.4 and 9.4.5). The Section 10.2 discuss the results related to the performance, both objective (Reaction times) and subjective (NASA-TLX and other questionnaires). The Section 10.3 discuss the workload results from a subjective point of view (perceived workload), while next Section 10.4 do the same regarding neurophysiological results. Other subjective results are discussed in Section 10.5. A word is provided regarding the limitations of this study in Section 10.6, before giving an overall synthesis of the results in Section 10.7, and a conclusion about this second experimental campaign in 10.8.

10.2 Performance

The results obtained regarding the performance show that the interaction and feedback modalities are well received by the end users. Designed to respond to specific problematic or even dangerous situations, identified during the design phase as 3 distinct use cases, we have experimentally shown that some of these interaction modalities seems to be adapted to ATCos need, in the sense that they allow better performance regarding the quantity of time needed to manage the corresponding situation. This is clearly the case for the SSA modality, which allows ATCos to be more responsive when an unauthorized event occurs on ground, e.g., when pilots begin to move their aircraft to taxi without prior authorization. In this particular case, participants were 3 times faster to react when the SSA modality was audible, in Single RCT environment. No results were obtained into Multiple RCT environment, which means that there was no significant decrease in performance due to the use of the other different modalities introduced.

However, we note that the performance scores assigned by the SME were clearly reduced when the augmentation modalities were activated in Multiple RCT environment (i.e., this subjective performance measurement was worst compared to Single RCT environment). On the other hand, the analysis of the participants' perception of their own overall performance didn't highlight any significant difference when the augmentations were activated or not, in Single or Multiple RCT environments. These results aside are quite difficult to interpret. On one hand, we have a measure showing that overall performance decreases when there is no augmentations in the very particular and quite novel case of Multiple RCT. On the other hand, the statistical analyses show no other difference, either from the subjective point of view of the participants, than from the SME point

of view when the augmentations are activated in Simple RCT environment. The only possible interpretation we see in these results is that the Multiple RCT environment is still new for ATCos: the addition of new modalities, moreover multimodal, does not significantly degrade their overall performance in a known environment (Single RCT), but this may disturb them when their working environment becomes partly unknown and more complex (or at least for which they don't have a lot of expertise), as it is the case in Multiple RCT environment where a second airport is integrated.

10.3 Perceived workload

A trend could be observed regarding a potential increase in perceived workload when the interaction and feedback modalities were activated. However the main messages resulting from the analyses of the subjective workload measurements are that both SME and NASA-TLX measurements show that perceived workload was lower when augmentations were disabled, and that SME measurements show that perceived workload was higher in Multiple RCT environment with respect to Single RCT one. These results, which were anticipated due to the complexity of the setup, can also be explained by a lack of experience from the ATCos. They are not accustomed to work in an environment in which their senses of touch and hearing are almost as stressed as their sense of vision. In addition, studies show that, taken separately, this type of interaction concept (at least based on sound) could significantly improve performance (the results related to the Audio Focus interaction modality reflect this). There may therefore be a familiarization issue. Anyway, this remains hypothetical at this stage of the study.

10.4 Results from neurophysiological measurements and performance analyses

The goal of the neurophysiological measurements was to demonstrate that the combination of information coming from neurophysiological measures, together with behavioral ones, could provide a more comprehensive assessment in the design process of new technologies and solutions in operational environments with respect to the evaluation of only behavioral and/or subjective measures. In this, considering these measurements could not be enough informative to investigate the contribution of new HCI concepts. Actually, it could even be strong to consider only one of these measures: results show that W_{TLX} and W_{SME} highlight a decrease of perceived workload when augmentations are disabled, while W_{EEG} objective workload index highlight a decrease in workload when augmentations are activated in Single RCT environment. It is obvious here that these two measures lead to contradictory results.

Anyhow, this information alone could not be enough if we do not look at the same time at the performance achieved by the participants. In this regard, the overall performance filled by the SME did not reveal any difference when augmentations were activated compared to when they were not, however reaction times measurements significantly decreased during the experimental events (RWY and SPATIAL) if the related augmented solutions were activated.

Eventually, researchers in charge of the neurophysiological measurements analyses concluded that:

- (i) *“The proposed augmented solutions induced a local increase in [participant] performance on those operational events where the augmentation was applied. The overall performance seems to not be affected by the proposed solutions.”*
- (ii) *“The augmented solutions induced an increase in the workload experienced by the participants. Anyhow, this increase is still acceptable and also fruitful, since it did not negatively impact the performance (that indeed locally increased) and has to be intended only as a consequence of the higher engagement of the [ATCos]. This behavioral effect is totally in line with physiological results obtained in terms of arousal (following point).”*
- (iii) *“Augmented solutions induced an increase in the arousal level of the [participant], representing a positive aspect for the overall performance assessment. This result corroborates the workload highlights and supports the overall interpretation of the propaedeutic effect of augmented solutions in enhancing the [ATCo] engagement, and consequently its performance. [...]”*

They finally concluded that *“[...] although the overall performance level of the [participants] seemed to not change significantly, the augmented solutions induced a local increase in performance. Anyhow, the related events during [augmented] scenarios, in particular during the event related to the [SSA modality], induced a higher experienced workload, identified by the W_{EEG} index. This effect could be due to the information that the ATCos would not be aware of, if they would not have been alerted. Even if in this specific experiment workload increasing did not induce a decrement in performance, it has to be stressed possible negative effects of increased workload over time, such as fatigue. [...] training of operators could have a significant effect on the experienced workload.”*

10.5 Other subjective results

The descriptive results informally inform us about how the different modalities of interaction and feedback were perceived by the participants during the experimental campaign. With regard to these results, we note that all the modalities were positively received in terms of their *usefulness*, *accuracy*, *contribution to SA improvement*, *sense of immersion* and *workload reduction*, since they all received average scores greater than 2.5 over 5. Several trends can be identified, but these cannot be considered as outcomes given their descriptive nature.

From the point of view of *usability*, the RWY modality received the highest average score (4.3 over 5). On average, participants considered this modality to be the most useful, surely because it is a very clear, and therefore potentially effective, way of alerting to a situation at risk which is almost feared by ATCos. Still in terms of usability, the AF and CSA modality also received good average scores (3.4 over 5), which on one hand corroborates the results obtained in the previous section, and on another hand confirms the need for an effective means of mentally separating the different radio frequencies in the case of a Multiple RCT environment.

In terms of *accuracy*, it is always the RWY modality that has received the highest scores. Participants considered this modality to be a clear and precise means of alerting to a runway incursion situation. We can also note that the AF modality received the lowest average score (2.9 over 5). Drowned in other modalities and feedback based on the same sensory channels (i.e., hearing and touch), it would seem that this modality, unlike the results obtained during the first experimental campaign, is losing its readability. Especially since vibrotactile feedback was disabled for this second experimental campaign. The vertical selection was therefore not present and AF was therefore only hearing-based. The configuration in which it was tested during the first experimental campaign (e.g., a laboratory task in an undisturbed environment requiring participants to focus only on aircraft location) seems to be the best known way to use it.

Finally, regarding the *contribution to SA improvement*, RWY modality again received the highest average score (4.3 over 5). However, AF modality is in second place with an average score of 3.6. These two modalities were perceived by participants as the most effective contributors to their situation awareness regarding the control situation, compared to the 2 other modalities. AF was perceived by the participants as the interaction modality that best allowed them to immerse themselves in the control situation, and the one that provided them with the least mental workload.

10.6 Limitations

The main limitation of this experimental campaign paradoxically lies in one of its strengths, namely the wish to make the experience as ecological as possible. This has involved relatively complex constraints to implement. Hence, it was necessary to make the experience alive so that it would be plausible for the participants, who are professionals. For example, pseudo-pilots had to be used to fly the controlled aircraft in real time and provide participants with real radio communications. The aeronautical language had to be considered for the dialogs, and the platform had to be considerably improved by adding all the necessary material to create this realistic environment (especially from a network point of view). Such a configuration clearly the experimental process more complex. In addition, this forced us to write, with the help of professionals, plausible control scenarios. Although realistic, they were perceived by participants as a little too simplistic, but still sufficient. These scenarios also made the experimental process more complex, particularly from a combinatorial point of view.

Testing modalities based simultaneously on different sensory channels proved to be relatively confusing for participants. They are not used to perceive the necessary information for their job other than through sight. The multiplicity of new interactions and feedback introduced during this experimental campaign added a little more to this impression. It seems that there was too much newness for the participants all at once. However, the training phase has reduced this effect, and although these modalities were new to them (especially in a realistic environment), they have nevertheless succeeded, at least partially, in using them. It even helps them sometimes. We have thus succeeded in quantifying, at least partially, the contribution of these interactive concepts.

10.7 Synthesis

We resume here in Table 10.2 the hypotheses that we had announced in the previous chapter 9 (Section 9.2.3, p. 163), and conclude with regard to each of them in the light of the objective results (i.e., *reaction times* and W_{EEG} index) obtained and previously discussed. The conclusions which are formulated in this table should be moderate in regard with the subjective results, telling that workload was lower when no augmentation were activated (W_{SME} and W_{TLX}) and that workload was higher in Multiple RCT environment (W_{SME}).

Table 10.2 — Second experimental campaign: Synthesis of the results regarding the hypothesis formulated in Chapter 9.

Hypothesis	Status	Conclusion
$H_{2,1}$	Validated	Reaction times were significantly shorter when using SA modality in Single RCT environment. W_{EEG} also show a decrease of workload when augmentations are activated in Single RCT environment. However, as it was averaged regarding all the modalities, we cannot tel which modality contributes to these workload reduction.
$H_{2,2}$	Partially validated	SSA modality has not improved reaction times in Multiple RCT environment. However, same comment as for $H_{2,1}$ concerning W_{EEG} index.
$H_{2,3}$	Partially validated	RWY modality has not improved reaction times in Single RCT environment. Same comment as for $H_{2,1,2,2}$ concerning W_{EEG} index.
$H_{2,4}$	Partially validated	RWY modality has not improved reaction times in Multiple RCT environment. Same comment as for $H_{2,1,...,2,3}$ concerning W_{EEG} index.
$H_{2,5}$	Partially validated	CSA modality has not improved reaction times in Single RCT environment. Same comment as for $H_{2,1,..,2,4}$ concerning W_{EEG} index.
$H_{2,6}$	Partially validated	CSA modality has not improved reaction times in Multiple RCT environment. Same comment as for $H_{2,1,...,2,5}$ concerning W_{EEG} index.

10.8 Conclusion

Our exchanges with ATCos during the design process allowed us to observe that they do not consciously integrate the information that can reach them through sensory channels other than sight.

Control towers are generally sound-proofed, making them almost airtight to the outer world (especially to the airport, actually). However, some elements can still be perceived from a control room (e.g., aircraft engines, wind, rain). The state-of-the-art of HCI in RCT environment shows that hearing remains the most considered sensory channel for designing new interaction modalities. The addition of new interaction modalities based on sensory channels other than vision would give RCT environments more immersion and sense of presence, making communication with the system more natural and embodied. However, these new forms of interaction require time to adapt, and their learning needs to be further explored. Even if participants have been thoroughly trained before participating in the experiment, this type of interaction modality is still too far from their normal working environment, which can be disturbing. Incidentally, once the experience was over, some participants expressly told us that the vibrotactile feedback felt through the haptic chair could become embarrassing at times. In any case, regarding the non-significant results, we can assume that the interaction modalities and feedback tested did not increase the cognitive workload of the participants. Moreover, performance has been improved in the case of SSA modality in Single RCT environment.

Even if the use of the two sensory channels of hearing and touch could be seen as natural, their novelty and the information provided consequently could need a longer appropriation time, implying specific care in learning sessions. In other words, although ATCos have been well trained to use in a correct way the proposed HCI solutions before the experiment and they could be useful in specific operational situations (i.e., RWY and SSA modalities), they could become too much distracting, inducing in some way an increase in experienced workload. However, novelty and lack of familiarity can be modulated by learning, which could mitigate this effect in the long term by decreasing workload, and eventually related induced fatigue effects or potential lacks in performance.

Therefore, this experiment allowed us to partially validate our hypotheses. In particular, one of the main results is that some modalities can improve performance, at least in terms of user responsiveness. This was clearly the case for SSA modality. Besides, the underlying message is that the introduction of new interaction modalities, based on the senses of hearing and touch, does not necessarily improve ATCos' workload. It is not significantly increased in Single RCT environment, however we have not been able to demonstrate an hypothetical decrease. Anyway, it seems that these types of interaction modalities do not affect performance, and that the present work demonstrated the suitability of neurophysiological measures during the validation phase of new HCI solutions in operational environments, in particular for remote tower operations. Further studies are needed to investigate more precisely the positive and negative impacts of the introduction of novel interaction concepts in Single and Multiple RCT context.

Conclusion & Perspectives

Résumé (Français)

Conclusion & perspectives

Synthèse générale

Cette thèse de doctorat s'est inscrite dans le cadre d'un financement européen H2020, plus précisément dans le cadre du projet SESAR MOTO. Ce projet a été présenté dans le Chapitre (p. [1](#)); son but premier était d'améliorer l'immersion, le sens de présence et la cognition incarnée dans les tours de contrôle déportées, au moyen d'augmentations technologiques se basant sur des concepts issus du domaine de l'Interaction Homme-Machine. Un intérêt particulier a été porté sur les sens autres que la vision, de manière à décharger le canal visuel des contrôleurs aériens (une grande partie des informations leur étant utiles leur parviennent par ce canal sensoriel qui se retrouve souvent surchargé). Dans un premier temps, le projet a consisté à se rapprocher au maximum des conditions réelles d'une tour de contrôle, en effectuant des enregistrements audio et vibratoires dans une tour de contrôle physique de manière à être en mesure de les reproduire au sein d'un premier prototype expérimental déporté. Cette tâche a été menée par les équipes italiennes de l'Université de la Sapienza à Rome. Dans un second temps, et partant de cette base de travail tendant à être suffisamment réaliste, le projet a consisté à concevoir avec l'aide des utilisateurs finaux (i.e., de contrôleurs aériens), développer et tester des concepts d'interaction novateurs, dans le but d'améliorer leur utilisabilité, le sentiment d'immersion globale au sein de cet environnement, ainsi que la performance des contrôleurs aériens d'approche et d'aéroport. Cette tâche a été du ressort des équipes de l'ENAC, à Toulouse. Ce travail de doctorat s'inscrit dans cette seconde problématique. Une attention particulière a également été portée à l'impact engendré sur la charge mentale des nouvelles modalités d'interaction ainsi créées.

Dans le but de concevoir des interactions adaptées au travail des contrôleurs, une étude de la littérature s'est imposée. De manière à être en mesure de comprendre leurs tâches de travail quotidiennes, nous nous sommes penchés de plus près sur le domaine du contrôle aérien et du concept de tour de contrôle déportée (Chapitre [1](#), p. [13](#)). Nous avons vu que le concept de contrôle déporté était un cas particulier de contrôle aérien. Comme son nom l'indique, il s'agit de contrôler un terrain distant depuis une position plus facile à gérer du point de vue des ressources humaines, techniques et par conséquent budgétaires. Nous appelons ce genre d'environnement digitalisé *tour de contrôle déportée*. La seule différence majeure existant entre une tour de contrôle déportée et une tour de contrôle dite "*physique*" réside dans le fait qu'au sein des premières, le contrôle se

fait de manière distante; l'environnement visuel de la tour de contrôle physique est filmé à l'aide de caméras implantées sur le terrain distant pour être diffusé dans la tour de contrôle déportée à l'aide d'écrans. Il existe plusieurs types de tours de contrôle déportées, à savoir celles de type *hybrides*, dans lesquelles sont uniquement regroupés les instruments nécessaires au contrôle à distance, celles de type *simples* (ou “*single airport*”), au sein desquelles un contrôleur aérien ne peut contrôler qu'un seul terrain distant à un instant t , celles de type *multiples* (ou “*multiple airport*”), au sein desquelles un contrôleur aérien peut contrôler plusieurs terrains distants simultanément, et celles de type *augmentées*, représentant un des trois concepts précédents dans lequel des augmentations IHM ont été ajoutés sous la forme de modalités interactives.

La présente recherche s'inscrit dans le cadre des tours de contrôle augmentées simples et multiples, sachant que ces dernières restent encore à ce jour un domaine relativement peu étudié comparé aux tours de contrôle déportées simples. Nous avons pu constater que, mise à part la vue, les canaux sensoriels des contrôleurs aériens ne sont pas toujours très sollicités. La vision est impliquée dans la majeure partie de leurs tâches quotidiennes, mis à part celles liées aux communications radio. Au sein d'une tour de contrôle physique, leurs sens haptiques (i.e., toucher et proprioception) leur permettent, par exemple, de ressentir le vent (et donc sa force via les oscillations de la tour), et leur ouïe peut leur permettre dans certains cas très précis de prendre conscience d'un moteur qui démarre. Cependant, ces sens sont rarement mis à contribution dans le but de leur faire parvenir de l'information. La problématique de cette thèse réside précisément dans cette constatation.

De manière à proposer des concepts IHM adaptés, il convient d'étudier dans un premier temps la perception humaine non-visuelle. Ce fut l'objet de la Partie I de ce manuscrit et des Chapitres 2, 3 et 4. Dans le premier d'entre eux (p. 35), nous proposons un aperçu du fonctionnement de la perception humaine. Celle-ci peut se diviser en trois parties: l'*extéroception*, qui concerne la perception externe au corps, l'*intéroception*, qui concerne la perception interne, et la *proprioception*, qui concerne la perception de notre corps et de ses mouvements dans l'espace environnement. L'extéroception et la proprioception sont les deux types de perception considérées dans ce travail de doctorat. L'ouïe fait partie de la première de ces deux formes de perception. Nous percevons notre environnement au travers de sons via cinq paramètres distincts: le *pitch* (ou fréquence perçue), le *timbre*, la *puissance* sonore (ou niveau sonore, ou amplitude), le *temps* et la position *spatiale*. Notre perception sonore nous permet dans de nombreuses occasions de percevoir des éléments de notre environnement n'étant pas accessibles à la vision seule. C'est principalement en cela que l'ouïe est un candidat intéressant pour notre étude. Nous avons également abordé le *système somatosensoriel*, système perceptuel complexe du corps humain, permettant l'acquisition d'informations sous la forme de *sensations* au travers de récepteurs sensoriels, ainsi que leur acheminement jusqu'au cerveaux où elles sont ensuite traitées pour créer notre *perception*. Le sens haptique, incluant entre autre notre perception *passive* ou *cutanée*, est géré par ce système somatosensoriel. La perception kinesthésique, quant à elle, est le produit de notre perception haptique et de notre proprioception.

L'ouïe est largement prise en compte dans le domaine de l'Interaction Homme-Machine (Fr., see HMI) (IHM) (voir Chapitre 3, p. 3). En plus des raisons évoquées ci-avant, elle favorise l'immersion. Ce dernier point est d'un intérêt tout particulier pour notre problématique d'augmentation

en environnement de contrôle déporté. Grâce aux paramètres spécifiés précédemment, elle fait le fruit d'une discipline à part entière appelée *sonification*. Il s'agit de l'ensemble des procédés permettant de transformer une information d'entrée en des sons en sortie. Plusieurs concepts liés à la sonification existent. Notamment, l'*audification* est le simple fait de jouer un son. Par ailleurs, les *earcons* permettent d'associer un son synthétique à un évènement. Dans ces cas, nous parlons d'IHM sonores dites *structurées*. Associés à notre bonne capacité à distinguer des sons sur l'axe horizontal, nous avons retenus ces différents concepts dans la conception de nouvelles modalités d'interaction. Celles-ci ont été inspirées de travaux précédents visant à élaborer des interfaces basées sur le son interactif.

Le Chapitre 4 (p. 4) quant à lui, nous a permis d'explorer les technologies haptiques existantes pour l'interaction. Ces technologies haptiques s'avèrent être un sujet largement étudiés, et ce depuis des décennies. Elles peuvent être catégorisées en deux grandes familles: les interfaces *haptiques*, qui se basent sur le sens du toucher, et les interfaces *kinesthésiques*, qui se basent sur la proprioception. Tout comme l'ouïe, elles favorisent l'immersion via la stimulation sensorielle. En fait, il serait dommage d'omettre ces canaux sensoriels dans la conception d'interfaces, tant ils nous fournissent tous les jours une grande quantité d'information relative à notre environnement. Nous avons présenté certains concepts tels que les *tactons*, qui permettent de donner un cadre formel à la conception de stimulations vibrotactiles. Nous avons également abordé le domaine de l'*interaction kinesthésique*, dont l'espace de conception permet de considérer certaines dimensions utiles dans la conception d'interfaces basées sur le mouvement du corps. Cette revue de la littérature nous a ensuite permis de concevoir des modalités d'interaction basées à la fois sur l'ouïe, la perception passive (ou cutanée) ainsi que la proprioception.

La première de ces modalités a été nommée *Audio Focus*. Sa conception, tout comme son évaluation ainsi que la discussion des résultats engendrés ont fait l'objet de la Partie III (p. 93). Avec l'aide de professionnels du domaine – tous contrôleurs aériens – nous avons isolé le cas d'utilisation de la perte de visibilité sur l'environnement de l'aéroport contrôlé. Celui-ci s'est avéré être d'un intérêt particulier pour les contrôleurs aériens, désireux de disposer d'une représentation mentale aussi fine que possible de la situation aérienne courante, toujours dans le but d'améliorer le niveau de sécurité global. Dans le Chapitre 5 (p. 93), nous expliquons comment la modalité *Audio Focus* a été conçue en se basant sur certains points de notre perception et des concepts d'interaction présentés ci-avant. Ainsi, nous avons par exemple tenu compte de l'*Angle Audible Minimum* (voir MAA) et de notre bonne résolution sonore horizontale, du concept d'*audification* et des stimulations vibrotactiles pour améliorer la localisation des avions lorsque la visibilité sur le terrain contrôlé est très mauvaise. Une campagne expérimentale a ensuite été menée (Chapitre 6, p. 107), dans laquelle nous avons convié vingt-deux contrôleurs aériens à prendre part à une tâche expérimentale visant à quantifier la contribution de cette nouvelle modalité. Des données subjectives, comportementales et neurophysiologiques ont été collectées. Les résultats sont discutés dans le Chapitre 7 (p. 135) et synthétisés dans la section suivante.

Durant la seconde campagne expérimentale, relatée dans la Partie III (p. 145), nous avons tenté de nous rapprocher le plus possible de l'environnement réel des tours de contrôle physiques afin d'améliorer l'expérience en tour de contrôle augmentée, simples mais également multiples. De la même manière que pour la première campagne expérimentale, nous avons isolés des situa-

tions ayant un intérêt pour les contrôleurs aériens, également avec leur participation. Ainsi, nous avons conçu des concepts interactifs servant à améliorer leur performance générale dans le cas de mouvements non autorisés sur le terrain contrôlé, des situations d'incursion de piste, et des communications radio multiples et simultanées dans le cas des tours de contrôle multiples (Chapitre 8, p. 145). Les trois modalités interactives associées respectivement à ces trois cas d'utilisation ont été activées au non au sein de scénarios de contrôle réalistes mettant en scène les événements associés (Chapitre 9, p. 157). Via un protocole expérimental ainsi qu'une plateforme plus complexe que celle ayant été considérée lors de la première expérience, nécessitant notamment l'intervention de pseudo-pilotes pour contrôler des avions en temps réel et ainsi pouvoir dialoguer de manière réaliste avec les participants, dans le but de proposer un environnement plus écologique, nous avons collecté des données subjectives, comportementales et physiologiques auprès de seize contrôleurs aériens. Ces résultats ont été discutés dans le Chapitre 10 (p. 183) et sont également synthétisés dans la section suivante de ce chapitre de conclusion.

Contributions scientifiques

Ces différentes études nous ont permis de valider, totalement ou partiellement, la plupart des hypothèses que nous avons formulé dans les Chapitre 6 (Section 6.2.4, p. 113) et 9 (Section 9.2.3, p. 163). Plus précisément, la première campagne expérimentale nous permet d'affirmer que (voir le Chapitre 6, Figures 6.7, 6.8, 6.9 et 6.11, ainsi que le Chapitre 7, Figure 7.1 et Table 7.1):

- (i) Le *concept* Audio Focus contribue à améliorer la représentation mentale d'un environnement immersif: il permet aux utilisateurs d'être plus précis dans la localisation perçue des éléments sonores spatialisés et d'en être les acteurs [140];
- (ii) Prévue pour une tâche de recherche spatiale, la contribution (i) se fait de manière moins fatigante qu'avec seulement du son spatialisé [140].

La contribution (i) se fait au détriment de temps de réaction significativement plus longs. Ce point est cependant négligeable si l'on considère la résolution temporelle généralement assez lente des tâches de travail des contrôleurs aériens. La contribution (ii), quant à elle, est un résultat subjectif.

Par ailleurs, la seconde campagne expérimentale nous a permis de constater que (voir le Chapitre 9, Figures 9.7, 9.8, 9.9 et Table 10.2):

- (iii) Les alertes sonores spatialisées (modalité SSA) permettent d'attirer l'attention de l'utilisateur plus rapidement vers un endroit précis de l'environnement dans lequel il est immergé [137];
- (iv) Les mesures subjectives communes telles que des questionnaires ouverts ou type NASA Task Load Index, ou bien des notations de performance de la part d'experts, peuvent être contredites par des mesures objectives plus invasives comme, dans notre cas, des mesures électroencéphalographiques [10, 9].

La contribution (iii) ne concerne pour l'instant que le cas d'un environnement immersif de tour de contrôle déportée et ne s'applique par conséquent qu'au contrôle aérien. D'autres études

sont nécessaires pour comprendre si elle peut s'appliquer à d'autres domaines, telles qu'à des applications en réalité virtuelle, par exemple. La contribution (iv) nécessite elle aussi des analyses plus poussées car elle a été découverte dans le contexte très particulier de la tâche de contrôle aérien augmentée par des modalités interactives. Il conviendrait d'étudier cet effet dans d'autres domaines; potentiellement, cela pourrait mener à des analyses objectives plus précises et réalistes de la charge mentale des utilisateurs.

Conclusion

Le travail relaté dans ce manuscrit a permis une réflexion sur les futurs outils du contrôle aérien sous l'angle de l'*Interaction Homme-Machine*. Ainsi, des concepts appartenant à la littérature de l'interaction ont été considérés, tels que les *earcons*, les *tactons* ou encore l'*interaction kinesthésique*. À la connaissance de l'auteur, ces genres de concepts très spécifiques sont rarement considérés dans la conception d'outils technologiques dans le cadre du contrôle aérien. Ils permettent toutefois d'avoir un regard différent sur la conception d'interaction en favorisant la conception participative et centrée utilisateur au travers d'espaces de conception éprouvés. Les résultats issus de la première campagne expérimentale suggèrent que le concept interactif Audio Focus s'avère être un bon candidat afin d'améliorer la détection des avions en situation de non-visibilité (Section 10.8). Bien reçue par les participants, il pourrait faire l'objet d'une fonctionnalité activable à la demande dans des solutions de contrôle déporté. Par ailleurs, la seconde campagne expérimentale, bien qu'ayant engendrée des résultats moins nets, s'avère être une première étude concluante quant à l'apport d'augmentations multimodales en environnement écologique de contrôle déporté. Par ailleurs, l'alerte sonore spatialisée a montré son efficacité, en termes de temps de réaction, afin d'alerter les contrôleurs aériens d'un événement relevant d'une situation potentiellement dangereuse.

Le cadre dans lequel s'est inscrit ce travail (financement européen H2020 SESAR), a favorisé les collaborations avec des équipes de recherche spécialisées dans les domaines des neurosciences et des facteurs humains. Cette collaboration a permis de considérer des méthodes non conventionnelles, et encore peu présentes dans le domaine de l'IHM, pour le test de concepts interactifs. Ainsi, les mesures neurophysiologiques ont été considérées comme des contre-mesures objectives aux habituelles mesures objectives comportementales (précision, temps de réaction) et subjectives (utilisabilité, fatigue, performance, charge mentale perçue, questionnaires ouverts ou de type NASA TLX, par exemple). Cependant, les expériences étaient intrinsèquement différentes : la première fut très ciblée et pouvait être qualifiée de tâche "de laboratoire", alors que la seconde se voulait se rapprocher d'un environnement écologique et fut plus souple et permissive d'un point de vue méthodologique (et donc plus complexe à gérer, également). C'est pourquoi d'autres études doivent être menées pour comprendre plus en détail l'impact de stimuli et d'interaction multimodales sur la tâche de travail des contrôleurs aériens et leur niveau de charge mentale. Un tel résultat doit être investigué plus en détails et peut être, également, dans d'autres domaines, dans le but d'obtenir des mesures objectives précises de charge mentale lorsqu'il s'agit d'évaluer de nouveaux concepts interactifs.

D'un point de vue strictement *interaction*, la modalité Audio Focus, quant à elle, a été perçue par les participants comme un outils envisageable dans leur environnement de travail. Les résultats

comportementaux obtenus dans le but de quantifier sa contribution sont très prometteurs. Il semblerait que le concept mis en œuvre, considérant l'ouïe et le toucher comme "mediums" principaux pour la recherche d'éléments spatialisés, de la même manière que la vue peut l'être en situation de visibilité, soit efficace. Il est donc nécessaire d'explorer plus en détails ce concept, de l'affiner d'un point de vue technique, conceptuel, et de le rendre plus adapté à un contexte écologique. Par ailleurs, les modalités SSA et RWY ont également reçu un bon accueil de la part des participants. La modalité SSA a reçu des résultats comportementaux (temps de réaction) concluants. Celle-ci a en effet permis aux contrôleurs de localiser plus efficacement des zones dans lesquelles une situation à risque était en train de se dérouler. Cela leur a ensuite permis de gérer ces situations plus rapidement. Sans une telle alerte, les contrôleurs peuvent passer à côté de ce genre d'évènements s'ils ne se trouvent pas dans leur champs de vision. Enfin, la dimension fortement intrusive de la modalité RWY, bien qu'ayant parfois été perçue comme gênante par certains participants, s'avère être, de leur propre propos, un moyen efficace pour signaler les incursions de piste. Les différents concepts interactifs conçus, mis en œuvre et évalués durant ces trois années ont par conséquent permis de proposer des outils novateurs et efficaces.

Perspectives

Au regard de la question générale sous-jacente au projet MOTO, à savoir *est-ce que l'ajout de nouvelles modalités interactives peut contribuer positivement à la performance et à l'immersion des contrôleurs aériens en contexte de tour de contrôle déportée ?*, ces résultats sont globalement encourageants. Comme évoqués précédemment, d'autres études sont nécessaires pour évaluer plus finement ces contributions, que ce soit dans le domaine du contrôle aérien, mais également dans d'autres domaines pouvant potentiellement bénéficier de l'apport de ce genre de concepts interactifs.

Il en est ainsi de la réalité virtuelle. Domaine actuellement en grande effervescence, la VR constitue un des principaux moyens mettant en œuvre des environnements réellement immersifs. Dans de tels environnements, où très souvent le son spatialisé joue un rôle prépondérant dans le degré d'immersion des utilisateurs, un concept tel qu'Audio Focus pourrait tout à fait contribuer à une représentation plus fine de l'espace. Les jeux vidéo pourraient en bénéficier, mais également des applications d'analyse immersive d'informations ou bien la réhabilitation d'handicaps, par exemple. La modalité SSA pourrait également apporter sa contribution pour focaliser l'utilisateur vers un point d'intérêt localisé. Les autres réalités, notamment les réalités augmentées et mixtes, représentent par ailleurs un terrain d'expérimentation prometteur pour un tel concept. Dans la vraie vie, certaines situations pourraient être facilitées par le fait d'amplifier les signaux sonores provenant d'une direction choisie par l'utilisateur (par l'intermédiaire des mouvements de la tête comme c'est le cas dans la version actuelle du concept, mais pourquoi pas, également, par l'intermédiaire du regard via le principe de la désignation oculaire). Ainsi, si on fait la forte hypothèse que les positions des sources sonores dans l'espace sont connues du système, la parole de certains locuteurs choisis pourraient être amplifiée lorsque l'utilisateur évolue dans un espace sonore surchargé (comme une terrasse de bistrot, un restaurant ou un centre commercial, par exemple). D'autres environnements immersifs pourraient être impactés de la même manière,

comme par exemple les environnements de type CAVE, constitués d'une multitude d'écrans englobant l'utilisateur. Le pilotage aéronautique pourrait aussi bénéficier d'Audio Focus, car les pilotes regardent fréquemment par les vitres de leur avion à la recherche d'autres avions dans leur environnement proche, pouvant être situés sur le même niveau de vol ou ceux situés en dessous et au-dessus. Associé à une carte de l'environnement connue du système, qui contiendrait la position des avions actuellement en vol, Audio Focus pourrait leur permettre un contact visuel plus rapide (les avions sont souvent difficiles à trouver dans ce contexte).

Cependant des améliorations sont à prévoir; Audio Focus n'est encore qu'au stade de prototype. Comme expliqué précédemment, la tâche expérimentale qui a permis de quantifier sa contribution était une tâche de laboratoire. Le concept doit être testé en environnement écologique. La désignation oculaire, également évoquée ci-avant, doit être envisagée et comparée au concept actuel qui considère les mouvements de la tête comme seule méthode de pointage. Le traitement du son doit également être amélioré : actuellement l'amplification des sources sonores se fait de manière linéaire. Il serait intéressant de tester le concept avec des méthodes de rendu plus évoluées. Le concept doit également être éprouvé en tant que fonctionnalité activable à la demande de ses utilisateurs. Son apport au contrôle aérien ne pourra se quantifier réellement qu'au travers d'une mise en situation en conditions réelles.

Par ailleurs, avant de conclure ce manuscrit, il est intéressant d'aborder les autres concepts qui n'ont pas été complètement étudiés durant ces trois années, notamment par manque de temps. Une technologie que nous avons sérieusement envisagée dans le but d'apporter de nouvelles modalités d'interaction au contexte des tours de contrôle déportées a été la stimulation cutanée dans les airs par ultrasons. Cette technique, évoquée dans le Chapitre 4 (voir Section 4.4.1, p. 82), permet de faire ressentir à l'utilisateur des sensations cutanées dans les airs par l'intermédiaire d'un procédé basé sur la différence de pression créée par le croisements d'ultrasons. Des formes, des directions peuvent ainsi être ressenties de manière haptique dans les airs, sans pour autant effectivement voir la forme en question. Cette technologie pourrait servir pour, par exemple, transmettre la force et la direction du vent aux contrôleurs aériens. Durant ce travail de doctorat, le périphérique a été utilisé et des prototypes ont été développés mais aucune expérience n'a réellement vu le jour afin de quantifier un apport potentiel. Par ailleurs, une telle technologie pourrait servir à faire ressentir les objets avec lesquels le contrôleur est en contact dans le cas des environnements de contrôle en réalité virtuelle.

Thesis summary

This Ph.D. thesis was part of a European H2020 funding, more precisely within the framework of the Single European Sky ATM Research (SESAR) project called MOTO. This project was introduced in the Chapter (p. [11](#)), explaining that its primary goal was to improve immersion, sense of presence and embodied cognition in remote control towers, through technological increases based on concepts coming from the field of Human Computer Interaction (HCI). Particular attention has been paid to the non-visionary senses, so as to relieve the visual channel of air traffic controllers, as most useful information reaches them through this channel. Initially, the project consisted in getting as close as possible to the real conditions of a physical control tower, by making audio and vibration recordings in a physical control tower in order to be able to reproduce them within a first experimental prototype. This task was carried out by the Italian teams at the University of La Sapienza in Rome. In a second step, and starting from this work base tending to be sufficiently realistic, the project then consisted in designing, with the help of end-users (i.e., air traffic controllers), developing and testing innovative interaction concepts, with the aim of improving their usability, the overall feeling of immersion, as well as the performance of approach and airport air traffic controllers. This task was the responsibility of ENAC teams in Toulouse. This doctoral work is part of this second theme. Particular attention was also paid to the impact of new interaction modalities on the mental workload.

This manuscript attempted to transcribe the different stages of this work. The design interactions adapted to the work of controllers primary required a literature review. In order to be able to understand their daily working tasks, we looked more closely at the field of air traffic control and the remote control tower concept (Chapter [1](#), p. [13](#)). We have seen that the concept of remote control is a special case of air traffic control. As its name suggests, it is about controlling an airport from a distant location that is easier to manage from the point of view of human, technical and therefore budgetary resources. We call this kind of digitalized environment *remote control tower*. The only major difference between a remote control tower and a *physical* control tower known is that using the former, airport control task is carried out remotely. The visual environment of the physical control tower is filmed using cameras located in the remote field to be broadcast in the remote control tower using screens. There are several types of remote control towers, namely those of the *hybrid* type, in which only the instruments necessary and vital for remote control are grouped together, those of the *single* type, in which an air traffic controller is able to control only one remote site at a time, those of type *multiple*, in which an air traffic controller is able to control several remote sites simultaneously, and those of type *augmented*, representing one of the three previous concepts in which HCI augmentations have been added as interactive modalities.

This research is contributing to the framework of single and multiple augmented remote control towers, knowing that the latter still remain to this day a relatively little studied field compared to single remote control towers. We have observed that the sensory channels of air traffic controllers other than sight are not always very considered. Their vision is used for almost all tasks except for radio communications. In a physical control tower, their haptic senses (i.e., touch and proprioception) allow them to feel the wind (and therefore its strength through the oscillations of the tower), for example, and their hearing can allow them in some very specific cases to be-

come aware of a starting engine. However, these sensory channels are rarely involved in providing information to air traffic controllers. This thesis lies precisely in this observation.

In order to propose adapted HCI concepts, it is necessary to first study non-visual human perception. This was the topic of Part I (p. 35) of this manuscript and the Chapters 2, 3 and 4. In the first one (p. 35), we provided an overview of how human perception works. This can be distinguished in three parts: *extéroception*, which concerns perception external to the body, *intéroception*, which concerns internal perception, and *proprioception*, which is the perception of our body and its movements in the surrounding space. Exteroception and proprioception are the two types of perception which are considered in this manuscript. Hearing is part of this first form of perception. We perceive our sound environment through five main parameters: *pitch* (or perceived frequency), *timbre*, *loudness* (sound level, or amplitude), *time* and *spatial* position. Our sound perception allows us on many occasions to perceive elements of our environment that are not accessible to vision. This is the main reason why hearing is an interesting candidate for our study. We also discussed the *somatosensory system*, a complex perceptual system of the human body, allowing the acquisition of information in the form of *sensations* through sensory receptors, as well as their routing to the brain where they are then processed to create our *perceptions*. The haptic sense, including among other things our *passive* or *cutaneous* perception, is managed by the somatosensory system. Kinesthetic perception is the product of our haptic perception and proprioception.

Hearing is widely studied in HCI (see Chapter 3, p. 3). In addition to the reasons mentioned above, it promotes immersion. This last point is of particular interest for our concern of HCI augmentations in remote control environment. Thanks to the aforementioned parameters, it is a discipline in its own right called *sonification*, which is the set of processes used to transform input information into output sounds. Several concepts related to sonification exist. In particular, *audification* is the simple act of playing a sound. In addition, *earcons* allow to associate a synthetic sound to an event. In these cases, we are talking about *structured* sound interfaces. Combined with our ability to distinguish sounds on the horizontal axis, we have retained these different concepts in the design of new interaction modalities. These were inspired by previous work to design interfaces based on interactive sound.

The Chapter 4 (p. 4), on the other hand, allowed us to explore existing haptic technologies for interaction. These haptic technologies have proven to be a widely studied subject for decades. They can be categorized into two main families: *haptic* interfaces, which are based on the sense of touch, and *kinesthetic* interfaces, which are based on proprioception. Like hearing, they promote immersion through sensory stimulation. In fact, it would be a shame to omit these sensory channels in the design of interfaces, as they provide us with a large amount of information about our environment every day. We presented some concepts such as *tactons*, which provide a formal framework for the design of vibrotactile stimuli. We also addressed the field of kinesthetic interaction, whose design space allows us to consider certain useful dimensions in the design of interfaces based on body movement. This literature review then allowed us to design interaction modalities based on hearing, passive (or skin) perception and proprioception.

The first of these modalities was named *Audio Focus*. Its design, as well as its evaluation and the discussion of the achieved results, is the topic of Part II (p. 93). With the help of professionals

from the field— being all air traffic controllers—we isolated the use case of loss of visibility on the airport vicinity. This has proved to be of particular interest to air traffic controllers, who wish to have a detailed mental representation of the current air configuration, always with the aim of improving the overall level of safety. In the Chapter 5 (p. 93), we explain how the Audio Focus modality was designed, based on specific parameters of our perception and the interaction concepts presented above. For example, we have taken into account the Minimum Audible Angle and our good horizontal sound resolution, the concept of audification and vibrotactile feedback to improve aircraft location when visibility is very poor. An experimental campaign was then carried out (Chapter 6, p. 107), in which we invited twenty-two air traffic controllers to take part in an experimental task aimed at quantifying the contribution of this new modality. Subjective, behavioral and neurophysiological data were collected. The results are discussed in the Chapter 7 (p. 135) and synthesized in the next section.

During the second experimental campaign, reported in Part III (p. 145), we have tried to get closer to the real environment of physical control towers to improve remote control tower experience (single and multiple). As in the first experimental campaign, we have isolated situations of interest to air traffic controllers (use cases), also with their participation. More precisely, we have developed interactive concepts to improve their overall performance in the case of unauthorized movements in the controlled field, runway incursion situations, and multiple simultaneous radio communications in the case of multiple airports (Chapter 8, p. 145). The three interactive modalities associated respectively with these three use cases were activated within realistic control scenarios involving the associated events (Chapter 9, p. 157). Using a more complex experimental protocol and platform, requiring in particular the use of pseudo-pilots to control aircraft in real time and thus be able to have a realistic dialog with participants, we proposed a more ecological environment to collect subjective, behavioral and physiological data from 16 air traffic controllers. These results were discussed in the Chapter 10 (p. 183) and are also summarized in the next section of this concluding chapter.

Scientific contributions

These studies allowed to validate totally or at least partially the hypotheses that have been formulated in Chapters 6 (Section 6.2.4, p. 113) and 9 (Section 9.2.3, p. 163). More specifically:

- (i) The *concept* of Audio Focus helps to improve the mental representation of an immersive environment: it allows users to be more precise in the perceived location of spatial sound elements and to be the actors of this representation [140];
- (ii) Anticipated for a spatial research task, contribution (i) generated less fatigue than with only spatial sound [140].

Contribution (i) was made at the expense of significantly longer reaction times. However, this point is negligible considering the generally slow temporal resolution of air traffic controllers' working tasks. Contribution (ii) is a subjective result.

The second experimental campaign allowed us to assess that (see Chapter 9, Figures 9.7, 9.8, 9.9 and Table 10.2):

- (iii) Spatial Sound Alert (SSA modality) allow the user's attention to be drawn more quickly to a specific location in the environment in which they are immersed [137];
- (iv) Common subjective measures such as open questionnaires or NASA Task Load Index, or performance ratings from experts, can be contradicted by more invasive objective measures such as, in our case, electroencephalographic measurements [10, 9].

Contribution (iii) currently only concerns the case of an immersive environment of remote control tower and therefore only applies to air traffic control. Further studies are needed to see if it can be applied to other fields, such as virtual reality. The contribution (iv) also requires further analysis because it was discovered in the very specific context of air traffic control task augmented with interactive modalities. This effect should be studied in other areas; potentially, this could lead to more precise and realistic objective analyses of users' mental workload.

Conclusion

This doctoral work allowed a reflection on the future tools of air traffic control from a *Human-Computer Interaction* angle. Hence, concepts from the literature of interaction have been considered, such as *earcons*, *tactons* and *kinesthetic interaction*. To the author's knowledge, these kinds of very specific concepts are rarely considered in the design of technological enhancements for air traffic control. However, they allow a different perspective on interaction design by promoting participatory and user-centered design through proven design spaces. The results of the first experimental campaign suggest that the Audio Focus interactive concept is proving to be a good candidate to improve the detection of aircraft in non-visibility situations (Section 10.8). Well received by the participants, it could take the form of a functionality that can be activated on demand in remote control solutions. In addition, the second experimental campaign, although producing less clear-cut results, is proving to be a first conclusive study on the contribution of multimodal augmentations in ecological remote control environment. Moreover, spatial sound alert has proven its effectiveness, in terms of reaction time, in alerting air traffic controllers to a potentially dangerous situation.

The framework within which this work was carried out (European funding H2020 SESAR) has encouraged collaboration with research teams specializing in neuroscience and human factors. This collaboration made it possible to consider unconventional methods, still not very present in the field of interaction, for the testing of interactive concepts. Therefore, neurophysiological measurements were considered as objective countermeasures to the usual objective behavioral (accuracy, reaction time) and subjective (usability, fatigue, performance, perceived mental load, open or NASA Task Load Index type questionnaires, for example) measures. We were able to highlight that, in some cases, subjective measures may be in contradiction with some objective measures. This was the case in the second experimental campaign with the mental load measurements: the

subjective measurements from the questionnaires and the expert's evaluation highlighted significantly lower mental workload values when the interaction modalities were disabled, while the objective EEG measurement highlighted significantly lower mental workload values when the interaction modalities were enabled in the context of single remote control towers (single airport). This result is in contradiction with the mental workload measurement highlighted during the first experimental campaign to evaluate the Audio Focus interaction modality, which, from a general point of view on the whole of this doctoral work, makes the interpretation of these results rather limited. However, the experiments were intrinsically different: the first was very focused and could be described as a laboratory task, while the second wanted to be closer to an ecological environment and was more flexible and permissive from a methodological point of view (and therefore more complex to manage, too). Therefore, further studies are needed to understand in more detail the impact of multimodal stimuli and interaction on the working task of air traffic controllers and their mental workload level. Such a result needs to be investigated in more detail and perhaps in other areas, in order to obtain precise objective measures of mental workload when evaluating new interactive concepts.

From a strictly *interaction* point of view, the Audio Focus modality was perceived by participants as a possible tool in their working environment. The behavioral results obtained in order to quantify its contribution are very promising. It would seem that the concept, considering hearing and touch as the main mediums for the search for spatial elements, in the same way that sight can be in a visible situation, is effective. It is therefore necessary to explore this concept in more detail, to refine it from a technical and conceptual point of view, and to make it more suitable for an ecological context. In addition, the SSA and RWY modalities were also well received by participants, as well as, for the SSA modality, conclusive behavioral results (reaction time). This has enabled controllers to more effectively locate areas where a critical situation was taking place. This then allowed them to manage these situations more quickly. Without such an alert, controllers can miss this type of event if they are not in their field of view. Finally, the highly intrusive dimension of the RWY modality, although perceived as annoying by some participants, is proving to be, of their own accord, an effective means to report runway incursions. The various interactive concepts which were designed, implemented and evaluated during these three years have therefore made it possible to propose innovative and effective tools.

Perspectives

Considering the general underlying question of the MOTO project, namely whether the addition of new interactive modalities can positively contribute to the performance and immersion of air traffic controllers in the context of a remote control tower, these results are generally encouraging. As mentioned above, further studies are needed to assess these contributions more accurately, both in the field of air traffic control and in other areas that could potentially benefit from the contribution of these types of interactive concepts.

This is the case with virtual reality. In great effervescence, this field is one of the main means of working in truly immersive environments. In such environments, where, very often, spatial sound plays a major role in the level of immersion, a concept such as Audio Focus could well

contribute to a fine representation of space. Video games could benefit, but also applications for immersive information analysis or disability rehabilitation, for example. The SSA modality could also contribute to focus the user towards a localized point of interest. Other realities, particularly augmented and mixed realities, also represent a promising field of experimentation for such a concept. In real life, some situations could be facilitated by amplifying the sound signals from a direction chosen by the user (through head movements as in the current version of the concept, but why not, also, through the gaze through the principle of gaze designation). Thus, if we make the strong assumption that the positions of sound sources in space are known to the system, the speech of some selected speakers could be amplified when the user is in a noisy space (such as a restaurant or a shopping mall, for example). Other immersive environments could be impacted in the same way, such as CAVE-type installations. Aeronautics could also benefit from Audio Focus, as pilots frequently look through the windows of their aircraft in search of other aircraft in their immediate environment, which may be located on the same flight level or those just below and above. Combined with a map of the system's known environment, which will contain the position of aircraft currently in flight, Audio Focus could allow them to have faster visual contact (aircraft are often difficult to find in this context).

However, improvements are to be expected; Audio Focus is still only at the prototype stage. As explained above, the experimental task that quantified its contribution was a laboratory task. The concept must be tested in an ecological context. Gaze designation, also mentioned above, should be considered and compared with the current concept that considers head movements as the only pointing method. Sound processing also needs to be improved: currently, sound sources are amplified in a linear way. It would be wise to test the concept considering more advanced rendering methods. The concept must also be proven as a feature that can be activated at the request of its users. Its contribution to air traffic control can only be really quantified through a real-life simulation.

In addition, before concluding this manuscript, it is interesting to consider the other concepts that have not been fully studied during these three years, particularly due to lack of time. One technology that we had seriously considered in order to bring new interaction concepts to the context of remote control tower was mid-air haptics using ultrasounds. This technique, mentioned in the Chapter 4 (see Section 4.4.1, p. 82), allows users to feel tactile-kinesthetic sensations in the air through a process based on the pressure difference created by the crossing of ultrasound. Shapes and directions can therefore be felt "*haptically*" in the air, without actually seeing the shape. This technology could be used to transmit wind strength and direction to air traffic controllers, for example (that was our first idea). During this doctoral work, the device was used and prototypes were developed but no experiment have been carried out to quantify a potential contribution.

List of publications

- International conferences:
 - **Maxime Reynal**, Jean-Paul Imbert, Pietro Aricó, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, and Fabio Babiloni. *Investigating Multimodal Augmentations Contribution to Remote Control Tower Contexts for Air Traffic Management*. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP, ISBN 978-989-758-354-4, pages 50–61. DOI: 10.5220/0007400300500061. INSTICC, ScitePress, 2019 [137]
 - Pietro Aricó, **Maxime Reynal**, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Enea Pavone, Alexandru C. Telea, and Fabio Babiloni. *Human–Machine Interaction Assessment by Neurophysiological Measures: A Study on Professional Air Traffic Controllers*. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pages 4619–4622, July 2018 [10]
- Journals:
 - **Maxime Reynal**, Pietro Aricó, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, and Fabio Babiloni. *Involving hearing, haptics and kinesthetics into non-visual interaction concepts for an augmented remote tower environment*. In Ana Paula Cláudio, KadiBouatouch, ManuelaChessa, AlexisPaljic, AndreasKerren, ChristopheHurter, AlainTremeau, and Giovanni Maria Farinella, editors, Computer Vision, Imaging and Computer Graphics Theory and Applications, pages 73–100, Cham, 2020. Springer International Publishing [138]
 - **Maxime Reynal**, Jean-Paul Imbert, Pietro Aricó, Gianluca Borghini, Christophe Hurter. *Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility*. International Journal of Human–Computer Studies, 129:116 – 128, 2019 [140]
 - Pietro Aricó, **Maxime Reynal**, Gianluca Di Flumeri, Gianluca Borghini, Nicolina Sciaraffa, Jean-Paul Imbert, Christophe Hurter, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Alexandru C. Telea, and Fabio Babiloni. *How neurophysiological measures can be used to enhance the evaluation of remote tower solutions*. Frontiers in Human Neuroscience, 13:303, 2019 [9]
- Before the Ph.D.:
 - **Maxime Reynal**, Frank Rister, Sébastien Scannella, Charles D. Wickens, and Frédéric Dehais (2017). *Investigating Pilot’s Decision Making when Facing an Unstabilized Approach: An Eye-Tracking Study*. 19th International Symposium on Aviation Psychology, 335-340 [141]
 - **Maxime Reynal**, Yvonne Colineaux, André Vernay, and Frédéric Dehais. *Pilot Flying Vs. Pilot Monitoring During the Approach Phase: An Eye-Tracking Study*. In Proceedings of the International Conference on Human-Computer Interaction in Aerospace, HCI-Aero ’16, pages 7:1–7:7, New York, NY, USA, 2016. ACM [139]

Bibliography

- [1] *The Sonification Handbook*. Logos Publishing House. [52](#), [58](#)
- [2] Merwan Achibet, Adrien Girard, Anthony Talvas, Maud Marchal, and Anatole Lécuyer. Elastic-arm: Human-scale passive haptic feedback for augmenting interaction and perception in virtual environments. In *2015 IEEE Virtual Reality (VR)*, pages 63–68, March 2015. [81](#)
- [3] Gerald Albaum. The likert scale revisited. *Market Research Society. Journal.*, 39(2):1–21, 1997. [174](#)
- [4] Robert L. Alexander, Jason A. Gilbert, Enrico Landi, Mary Simoni, Thomas H. Zurbuchen, and D. Aaron Roberts. Audification as a diagnostic tool for exploratory heliospheric data analysis. International Community for Auditory Display, 2011. [61](#)
- [5] J Andary, K Halterman, D Hewitt, and P Sabelhaus. The flight telerobotic servicer (fts) nasa’s first operational robotic system. 1990. [vii](#), [77](#)
- [6] Guillaume Andéol, Clara Suied, Sébastien Scannella, and Frédéric Dehais. 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, 2017. [95](#), [101](#)
- [7] Guillaume Andéol, Sophie Savel, and Anne Guillaume. Perceptual factors contribute more than acoustical factors to sound localization abilities with virtual sources. *Frontiers in Neuroscience*, 8:451, 2015. [37](#), [43](#), [44](#)
- [8] Pietro Aricò, Gianluca Borghini, Gianluca Di Flumeri, Alfredo Colosimo, Stefano Bonelli, Alessia Golfetti, Simone Pozzi, Jean-Paul Imbert, Géraud Granger, Raïlane Benhacene, and Fabio Babiloni. Adaptive automation triggered by eeg-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, 10:539, 2016. [174](#), [175](#)
- [9] Pietro Aricò, Maxime Reynal, Gianluca Di Flumeri, Gianluca Borghini, Nicolina Sciaraffa, Jean-Paul Imbert, Christophe Hurter, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Alexandru C. Telea, and Fabio Babiloni. How neurophysiological measures can be used to enhance the evaluation of remote tower solutions. *Frontiers in Human Neuroscience*, 13:303, 2019. [196](#), [203](#), [206](#)
- [10] Pietro Aricò, Maxime Reynal, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Enea Pavone, Alexandru C. Telea, and Fabio Babiloni. Human-machine interaction assessment by neurophysiological measures: A study on professional air traffic controllers. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4619–4622, July 2018. [123](#), [196](#), [203](#), [206](#)

- [11] Pietro Aricò, Gianluca Borghini, Gianluca Di Flumeri, Alfredo Colosimo, Simone Pozzi, and Fabio Babiloni. A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. In *Progress in brain research*, volume 228, pages 295–328. Elsevier, 2016. [123](#)
- [12] Mark Edward Ballora. Sonification, science and popular music: In search of the 'wow'. *Organised Sound*, 19(1):30–40, 2014. [59](#)
- [13] Stephen Barrass. Sonification design patterns. Georgia Institute of Technology, 2003. [55](#), [66](#)
- [14] Stephen Barrass and Björn Zehner. Responsive sonification of well-logs. In *in Proc.2000 of Int. Conf. On Auditory Display*, pages 72–80, 2000. [66](#)
- [15] Brain Basics. Music is good for your brain, but don't blast it. [41](#)
- [16] Olivier Bau, Ivan Poupyrev, Ali Israr, and Chris Harrison. Teslatouch: Electro-vibration for touch surfaces. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 283–292, New York, NY, USA, 2010. ACM. [83](#)
- [17] Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. Normaltouch and texturetouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 717–728, New York, NY, USA, 2016. ACM. [vii](#), [82](#), [83](#)
- [18] Benjamin Blankertz, Laura Acqualagna, Sven Dähne, Stefan Haufe, Matthias Schultze-Kraft, Irene Sturm, Marija Ušćumlic, Markus A. Wenzel, Gabriel Curio, and Klaus-Robert Müller. The berlin brain-computer interface: Progress beyond communication and control. *Frontiers in Neuroscience*, 10:530, 2016. [174](#)
- [19] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. Earcons and icons: Their structure and common design principles (abstract only). *SIGCHI Bull.*, 21(1):123–124, August 1989. [ix](#), [53](#), [62](#), [63](#)
- [20] Richard A. Bolt. Gaze-orchestrated dynamic windows. *SIGGRAPH Comput. Graph.*, 15(3):109–119, August 1981. [vii](#), [55](#), [66](#), [67](#)
- [21] Till Bovermann, Thomas Hermann, and Helge Ritter. Tangible data scanning sonification model. Georgia Institute of Technology, 2006. [65](#)
- [22] R. Bowen Loftin. Multisensory perception: beyond the visual in visualization. *Computing in Science Engineering*, 5(4):56–58, July 2003. [100](#)
- [23] Svein Braathen. Air transport services in remote regions. 2011. [1](#), [6](#)
- [24] Eoin Brazil and Mikael Fernström. *The Sonification Handbook*, chapter Auditory Icons, pages 325–338. Logos Publishing House. [62](#)
- [25] Stephen Brewster and Lorna M. Brown. Tactons: Structured tactile messages for non-visual information display. In *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28*, AUIC '04, pages 15–23, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc. [ix](#), [x](#), [72](#), [84](#), [85](#), [96](#), [100](#), [104](#), [105](#), [147](#), [153](#), [156](#)
- [26] Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. Experimentally derived guidelines for the creation of earcons, 1995. [ix](#), [63](#), [151](#), [155](#)
- [27] Stephen Anthony Brewster. Providing a structured method for integrating non-speech audio into human-computer interfaces. Technical report, 1994. [ix](#), [54](#), [62](#), [63](#)

- [28] Frederick P. Jr. Brooks, Ming Ouh-Young, James J. Batter, and P. Jerome Kilpatrick. Project grope – haptic displays for scientific visualization. *SIGGRAPH Comput. Graph.*, 24(4):177–185, September 1990. [vii](#), [79](#)
- [29] D. S. Brungart and B. D. Simpson. Auditory localization of nearby sources in a virtual audio display. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 107–110, Oct 2001. [68](#)
- [30] Marcellin Buisson, Alexandre Bustico, Stephane Chatty, Francois-Regis Colin, Yannick Jestin, Sebastien Maury, Christophe Mertz, and Philippe Truillet. Ivy: a bus software to the development of prototype systems interactive. In *ACM International Conference Proceeding Series*, volume 32, pages 223–226, 2002. [117](#), [118](#)
- [31] Simon Carlile. *The Physical and Psychophysical Basis of Sound Localization*, pages 27–78. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996. [37](#), [42](#), [45](#)
- [32] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. Ultrahaptics: Multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 505–514, New York, NY, USA, 2013. ACM. [72](#), [83](#)
- [33] Stephane Chatty. The ivy software bus. *white paper*, [www.tls.cena.fr/products/ivy/documentation/ivy.pdf](#), 2003. [117](#), [118](#)
- [34] Corey I. Cheng and Gregory H. Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. *Journal of Audio Engineering Society*, 49(4):231–249, 2001. [68](#)
- [35] Philip R Cohen and David R McGee. Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1):41–46, 2004. [136](#), [139](#)
- [36] Maxime Cordeil, Tim Dwyer, and Christophe Hurter. Immersive solutions for future air traffic control and management. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, ISS Companion '16, pages 25–31, New York, NY, USA, 2016. ACM. [15](#), [27](#)
- [37] Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. Agency in mid-air interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2426–2439, New York, NY, USA, 2017. ACM. [72](#), [83](#)
- [38] AD (Bud) Craig. Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology*, 13(4):500 – 505, 2003. [35](#), [40](#)
- [39] James C Craig and Carl E Sherrick. Dynamic tactile displays. *Tactual perception: A sourcebook*, pages 209–233, 1982. [78](#), [85](#)
- [40] Kai Crispian, Klaus Fellbaum, Anthony Savidis, and Constantine Stephanidis. A 3d-auditory environment for hierarchical navigation in non-visual interaction. In *Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology, 1996. [vii](#), [56](#), [67](#), [68](#), [136](#), [141](#), [151](#)
- [41] Kai Crispian, Wolfgang Würz, and Gerhard Weber. Using spatial audio for the enhanced presentation of synthesised speech within screen-readers for blind computer users. In Wolfgang L. Zagler, Geoffrey Busby, and Roland R. Wagner, editors, *Computers for Handicapped Persons*, pages 144–153, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg. [vii](#), [67](#), [68](#), [136](#), [141](#), [151](#)

- [42]  Csap, Gyrgy Wersenyi, Hunor Nagy, and Tony Stockman. A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research. *Journal on Multimodal User Interfaces*, 9(4):275–286, Dec 2015. [75](#)
- [43] Thomas J. Csordas. Embodiment as a paradigm for anthropology. *Ethos*, 18(1):5–47, 1990. [87](#)
- [44] Stanislas Dehaene. The neural basis of the weber–fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4):145 – 147, 2003. [36](#), [43](#)
- [45] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9 – 21, 2004. [123](#)
- [46] Florian Dombois. Using audification in planetary seismology. Georgia Institute of Technology, 2001. [53](#), [61](#)
- [47] Florian Dombois and Gerhard Eckel. *The Sonification Handbook*, chapter Audification, pages 301–324. Logos Publishing House. [53](#), [61](#)
- [48] Paul Dourish. *Where the action is*. MIT press Cambridge, 2001. [73](#), [86](#)
- [49] Gael Dubus and Roberto Bresin. A systematic review of mapping strategies for the sonification of physical quantities. *PLOS ONE*, 8(12):1–28, 12 2013. [36](#), [42](#), [51](#), [52](#), [54](#), [57](#), [60](#), [64](#), [101](#)
- [50] Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H. Thomas. *Immersive Analytics: An Introduction*, pages 1–23. Springer International Publishing, Cham, 2018. [xi](#), [15](#), [27](#), [55](#), [67](#), [81](#)
- [51] Mike Eissele, Oliver Siemoneit, and Thomas Ertl. Transition of mixed, virtual, and augmented reality in smart production environments - an interdisciplinary view. In *2006 IEEE Conference on Robotics, Automation and Mechatronics*, pages 1–6, June 2006. [81](#)
- [52] Jan B. F. Van Erp, Hendrik A. H. C. Van Veen, Chris Jansen, and Trevor Dobbins. Waypoint navigation with a vibrotactile waist belt. *ACM Trans. Appl. Percept.*, 2(2):106–117, April 2005. [73](#), [86](#)
- [53] Nicolas Esposito and Charles Lenay. Feeltact: rich tactile feedback for mobile gaming. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, page 71. ACM, 2011. [84](#)
- [54] W Tecumseh Fitch and Gregory Kramer. Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In *Santa Fe Institute Studies in the Sciences of Complexity, Proceeding Volume*, volume 18, pages 307–307. Addison-Wesley Publishing Co, 1994. [62](#)
- [55] Maiken Hillerup Fogtmann, Jonas Fritsch, and Karen Johanne Kortbek. Kinesthetic interaction: Revealing the bodily potential in interaction design. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, OZCHI ’08, pages 89–96, New York, NY, USA, 2008. ACM. [ix](#), [73](#), [86](#), [87](#), [88](#), [100](#), [101](#), [102](#)
- [56] Keith M. Franklin and Jonathan C. Roberts. Pie chart sonification. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.*, pages 4–9, July 2003. [64](#)
- [57] Sarah F. Frisken-Gibson, Paul Bach-y Rita, Willis J. Tompkins, and John G. Webster. A 64-solenoid, four-level fingertip search display for the blind. *IEEE Transactions on Biomedical Engineering*, BME-34(12):963–965, Dec 1987. [78](#), [79](#)

- [58] B. Frohlich, S. Barrass, B. Zehner, J. Plate, and M. Gobel. Exploring geo-scientific data in virtual environments. In *Proceedings Visualization '99 (Cat. No.99CB37067)*, pages 169–173, Oct 1999. [66](#)
- [59] Norbert Fürstenau. *Virtual and Remote Control Tower: Research, Design, Development and Validation*. Springer Publishing Company, Incorporated, 1st edition, 2016. [21](#)
- [60] Steve P. Frysinger. Applied research in auditory data representation. volume 1259, 1990. [61](#)
- [61] Norbert Fürstenau, Markus A. Schmidt, Michael Rudolph, Christoph Möhlenbrink, Maik Friedrich, Anne Papenfuss, and Sven Kaltenhäuser. Steps towards the virtual tower: Remote airport traffic control center (raice). volume 1, page 14. Citeseer, 2009. [1](#), [6](#)
- [62] Erich Gamma, Richard Helm, Ralph E. Johnson, and John M. Vlissides. Design patterns: Abstraction and reuse of object-oriented design. In *Proceedings of the 7th European Conference on Object-Oriented Programming, ECOOP '93*, pages 406–431, Berlin, Heidelberg, 1993. Springer-Verlag. [55](#), [66](#)
- [63] William W. Gaver. The sonicfinder: An interface that uses auditory icons (abstract only). *SIGCHI Bull.*, 21(1):124–, August 1989. [53](#), [61](#), [62](#)
- [64] William W. Gaver and Randall B. Smith. Auditory icons in large-scale collaborative environments. In Ronald M. Baecker, Jonathan Grudin, William A.S. Buxton, and Saul Greenberg, editors, *Readings in Human-Computer Interaction*, Interactive Technologies, pages 564 – 569. Morgan Kaufmann, 1995. [62](#)
- [65] James Jerome Gibson. The senses considered as perceptual systems. 1966. [38](#), [47](#)
- [66] Adrien Girard, Maud Marchal, Florian Gosselin, Anthony Chabrier, François Louveau, and Anatole Lécuyer. Haptip: Displaying haptic shear forces at the fingertips for multi-finger interaction in virtual environments. *Frontiers in ICT*, 3:6, 2016. [82](#)
- [67] Rob Gray, Cristy Ho, and Charles Spence. A comparison of different informative vibrotactile forward collision warnings: Does the warning need to be linked to the collision event? *PLOS ONE*, 9(1):1–8, 01 2014. [86](#)
- [68] M Grohn. Sound probe: An interactive sonification tool. *Proc. ICAD 1992*, 1992. [55](#), [65](#)
- [69] Markus Guldenschuh and Alois Sontacchi. Application of transaural focused sound reproduction. In *6th Eurocontrol INO-Workshop 2009*, 2009. [15](#), [30](#)
- [70] Eric Gunther and Sile O'Modhrain. Cutaneous grooves: Composing for the sense of touch. *Journal of New Music Research*, 32(4):369–381, 2003. [85](#)
- [71] Eric Eric Louis Gunther. *Skinscape: A tool for composition in the tactile modality*. PhD thesis, Massachusetts Institute of Technology, 2001. [85](#)
- [72] M.P. Jacob Habgood, David Wilson, David Moore, and Sergio Alapont. Hci lessons from playstation vr. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '17 Extended Abstracts*, pages 125–135, New York, NY, USA, 2017. ACM. [81](#)
- [73] Nur Al-huda Hamdan, Adrian Wagner, Simon Voelker, Jürgen Steimle, and Jan Borchers. Springlets: Expressive, flexible and silent on-skin tactile interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, pages 488:1–488:14, New York, NY, USA, 2019. ACM. [84](#)

- [74] Stephen Handel. *Listening: An introduction to the perception of auditory events*. The MIT Press, 1993. [37](#), [43](#)
- [75] Sandra G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. [120](#), [172](#)
- [76] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland, 1988. [120](#), [172](#)
- [77] Thomas Hermann. Taxonomy and definitions for sonification and auditory display. In *Proc. 14th Int. Conference on Auditory Display (ICAD)*, pages 24–27, 2008. [51](#), [54](#), [58](#), [64](#), [65](#)
- [78] Thomas Hermann, Andy Hunt, and John G. Neuhoff. *The Sonification Handbook*, chapter Introduction, pages 21–26. Logos Publishing House. [vii](#), [42](#), [59](#)
- [79] Thomas Hermann and Helge Ritter. Listen to your data: Model-based sonification for data analysis. *Advances in intelligent computing and multimedia systems*, 1999. [54](#), [65](#)
- [80] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, pages 452–458, New York, NY, USA, 1994. ACM. [79](#), [80](#)
- [81] Thilo Hinterberger and Gerold Baier. Parametric orchestral sonification of eeg in real time. *IEEE MultiMedia*, 12(2):70–79, April 2005. [65](#)
- [82] Ira J Hirsh and Carl E Sherrick Jr. Perceived order in different sense modalities. *Journal of experimental psychology*, 62(5):423, 1961. [78](#)
- [83] Cristy Ho, Hong Z. Tan, and Charles Spence. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(6):397 – 412, 2005. [86](#)
- [84] Takayuki Hoshi, Masafumi Takahashi, Takayuki Iwamoto, and Hiroyuki Shinoda. Noncontact tactile display based on radiation pressure of airborne ultrasound. *IEEE Transactions on Haptics*, 3(3):155–165, July 2010. [72](#), [83](#)
- [85] Christophe Hurter, Rémi Lesbordes, Catherine Letondal, Jean-Luc Vinot, and Stéphane Conversy. Strip'tic: Exploring augmented paper strips for air traffic controllers. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 225–232, New York, NY, USA, 2012. ACM. [15](#), [30](#), [139](#)
- [86] Christophe Hurter, Nathalie Henri Riche, Steven M. Drucker, Maxime Cordeil, Richard Alligier, and Romain Vuillemot. Fiberclay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):704–714, Jan 2019. [81](#)
- [87] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. *Human–Computer Interaction*, 1(4):311–338, December 1985. [56](#), [68](#), [79](#)
- [88] Emanuelle Jeannot, Chris Kelly, and David Thompson. The development of situation awareness measures in atm systems. 2003. [174](#)
- [89] Matthew J. Jensen, A. Madison Tolbert, John R. Wagner, Fred S. Switzer, and Joshua W. Finn. A customizable automotive steering system with a haptic feedback control strategy for obstacle avoidance notification. *IEEE Transactions on Vehicular Technology*, 60(9):4208–4216, Nov 2011. [86](#)

-
- [90] Emil Jovanov, Dusan Starcevic, Kristen Wegner, Daniel Karron, and Vlada Radivojevic. Acoustic rendering as support for sustained attention during biomedical procedures. Georgia Institute of Technology, 1998. [66](#)
- [91] Valer Jurcak, Daisuke Tsuzuki, and Ipeita Dan. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611, 2007. [123](#)
- [92] Valer Jurcak, Daisuke Tsuzuki, and Ipeita Dan. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, 34(4):1600 – 1611, 2007. [175](#)
- [93] Randa Kassab and Frédéric Alexandre. Integration of exteroceptive and interoceptive information within the hippocampus: a computational study. *Frontiers in Systems Neuroscience*, 9:87, 2015. [35](#), [40](#)
- [94] Peter G.W. Keen and Michael S. Scott Morton. Decision support systems; an organizational perspective. Technical report, 1978. [55](#), [66](#)
- [95] Scott R. Klemmer, Björn Hartmann, and Leila Takayama. How bodies matter: Five themes for interaction design. In *Proceedings of the 6th Conference on Designing Interactive Systems*, DIS '06, pages 140–149, New York, NY, USA, 2006. ACM. [73](#), [86](#)
- [96] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2):169 – 195, 1999. [123](#), [175](#)
- [97] Gregory Kramer. *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Perseus Publishing, 1993. [51](#), [57](#), [60](#)
- [98] Gregory Kramer, BN Walker, Terri Bonebright, Perry Cook, J Flowers, Nadine Miner, John Neuhoff, R Bargar, S Barrass, J Berger, G Evreinov, WT Fitch, M Gröhn, S Handel, H Kaper, H Levkowitz, S Lodha, B Shinn-Cunningham, M Simoni, and S Tipei. The sonification report: Status of the field and research agenda. report prepared for the national science foundation by members of the international community for auditory display. In *International Community for Auditory Displays (ICAD)*, 1999. [37](#), [44](#), [51](#), [57](#)
- [99] John Krantz. Experiencing sensation and perception. *Upper Saddle River, NJ*, 2012. [40](#)
- [100] James R Lackner and Paul DiZio. Vestibular, proprioceptive, and haptic contributions to spatial orientation. *Annu. Rev. Psychol.*, 56:115–147, 2005. [87](#)
- [101] Catherine Letondal, Christophe Hurter, Rémi Lesbordes, Jean-Luc Vinot, and Stéphane Conversy. Flights in my hands: Coherence concerns in designing strip'tic, a tangible space for air traffic controllers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2175–2184, New York, NY, USA, 2013. ACM. [15](#), [18](#), [30](#)
- [102] Grace Li, Chen Wang, Chandani Doshi, Tania Yu, Jialin Shi, and Charlene Xia. Text to braille converter, March 22 2018. US Patent App. 15/374,398. [vii](#), [75](#), [76](#)
- [103] William Little. The nasa augmented/virtual reality lab: The state of the art at ksc. 2017. [81](#)
- [104] Allan Christian Long, Jr., James A. Landay, and Lawrence A. Rowe. Implications for a gesture design tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 40–47, New York, NY, USA, 1999. ACM. [73](#), [86](#)
- [105] Karon Maclean and Mario Enriquez. Perceptual design of haptic icons. In *In Proceedings of Euro-haptics*, pages 351–363, 2003. [84](#)

- [106] Tara M Madhyastha and Daniel A Reed. A framework for sonification design. In *Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume*, volume 18, pages 267–267. Addison-Wesley Publishing Co, 1994. [64](#)
- [107] Y. Makino, N. Asamura, and H. Shinoda. A whole palm tactile display using suction pressure. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, volume 2, pages 1524–1529 Vol.2, April 2004. [84](#)
- [108] Jessica L. Maples-Keller, Matthew Price, Sheila Rauch, Maryrose Gerardi, and Barbara O. Rothbaum. Investigating relationships between ptsd symptom clusters within virtual reality exposure therapy for oef/oif veterans. *Behavior Therapy*, 48(2):147 – 155, 2017. Special Issue: Treating Posttraumatic Stress Disorder (PTSD): Innovations and Understanding Processes of Change. [81](#)
- [109] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology*. Pearson Education, 2007. [35](#), [40](#)
- [110] Thomas H. Massie and J. K. Salisbury. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME Dynamic Systems and Control Division*, pages 295–301, 1994. [vii](#), [75](#), [80](#)
- [111] Michael W. McGreevy. Chapter 8 - virtual reality and planetary exploration. In Alan Wexelblat, editor, *Virtual Reality*, pages 163 – 197. Academic Press, 1993. [81](#)
- [112] Claudine Mélan and Edith Galy. Recall performance in air traffic controllers across the 24-hr day: influence of alertness and task demands on recall strategies. *Advances in Air Navigation Services*, pages 35–54, 2012. [100](#)
- [113] Fanxing Meng, Rob Gray, Cristy Ho, Mujthaba Ahtamad, and Charles Spence. Dynamic vibrotactile signals for forward collision avoidance warning systems. *Human factors*, 57(2):329–346, 2015. [86](#)
- [114] John C. Middlebrooks. Narrow-band sound localization related to external ear acoustics. *The Journal of the Acoustical Society of America*, 92(5):2607–2624, 1992. [37](#), [44](#)
- [115] A. W. Mills. On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, 1958. [45](#), [95](#), [99](#), [104](#)
- [116] Christoph Moehlenbrink and Anne Papenfuss. Atc-monitoring when one controller operates two airports: Research for remote tower centres. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1):76–80, 2011. [22](#)
- [117] Ashley Montagu. Touching, the human significance of the skin. *Perennial Library*, pages 98–99, 1972. [38](#), [46](#)
- [118] Brian CJ Moore. Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America*, 54(3):610–619, 1973. [43](#)
- [119] Elizabeth D. Mynatt and Gerhard Weber. Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In *Conference Companion on Human Factors in Computing Systems, CHI '94*, pages 211–, New York, NY, USA, 1994. ACM. [136](#), [141](#)
- [120] Vilas Nene. *Remote Tower Research in the United States*, pages 279–312. Springer International Publishing, Cham, 2016. [1](#), [6](#)
- [121] Keith V Nesbitt and Stephen Barrass. Evaluation of a multimodal sonification and visualisation of depth of market stock data. Georgia Institute of Technology, 2002. [61](#)

- [122] Marianna Obrist, Sriram Subramanian, Elia Gatti, Benjamin Long, and Thomas Carter. Emotions mediated through mid-air haptics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2053–2062, New York, NY, USA, 2015. ACM. [83](#)
- [123] International Civil Aviation Organization. The convention on international civil aviation. [18](#)
- [124] Sharon Oviatt. Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 576–583, New York, NY, USA, 1999. ACM. [136](#), [139](#)
- [125] Sabrina Paneels and Jonathan C. Roberts. Review of designs for haptic data visualization. *IEEE Transactions on Haptics*, 3(2):119–137, April 2010. [81](#)
- [126] Sabrina A. Panëels, Jonathan C. Roberts, and Peter J. Rodgers. Hitproto: a tool for the rapid prototyping of haptic interactions for haptic data visualization. In *2010 IEEE Haptics Symposium*, pages 261–268, March 2010. [75](#)
- [127] A. Papenfuss and M. Friedrich. Head up only — a design concept to enable multiple remote tower operations. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, pages 1–10, Sep. 2016. [22](#)
- [128] Narcís Parés, Anna Carreras, and Miquel Soler. Non-invasive attitude detection for full-body interaction in mediate, a multisensory interactive environment for children with autism. In *VMV*, pages 37–45. Citeseer, 2004. [73](#), [86](#)
- [129] Sergey V Pereverzev, A Loshak, Scott Backhaus, JC Davis, and RE Packard. Quantum oscillations between two weakly coupled reservoirs of superfluid ^3He . *Nature*, 388(6641):449, 1997. [53](#), [61](#)
- [130] S.M. Petermeijer, S. Cieler, and J.C.F. de Winter. Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis and Prevention*, 99:218 – 227, 2017. [73](#), [86](#), [95](#), [104](#)
- [131] Reinier Plomp. Rate of decay of auditory sensation. *The Journal of the Acoustical Society of America*, 36(2):277–282, 1964. [37](#), [43](#)
- [132] Arnaud Prouzeau, Maxime Cordeil, Clement Robin, Barrett Ens, Bruce H. Thomas, and Tim Dwyer. Scaptics and highlight-planes: Immersive interaction techniques for finding occluded features in 3d scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 325:1–325:12, New York, NY, USA, 2019. ACM. [82](#)
- [133] Michael Quinton, Iain McGregor, and David Benyon. Sonifying the solar system. International Community on Auditory Display, 2016. [59](#)
- [134] Anil K. Raj, Steven J. Kass, and James F. Perry. Vibrotactile displays for improving spatial awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(1):181–184, 2000. [73](#), [86](#)
- [135] Christophe Ramstein and Vincent Hayward. The pantograph: A large workspace haptic device for multimodal human computer interaction. In *Conference Companion on Human Factors in Computing Systems*, CHI '94, pages 57–58, New York, NY, USA, 1994. ACM. [79](#), [80](#), [84](#)
- [136] Laurie Hodges Reuter, Paul Tukey, Laurence T. Maloney, John R. Pani, and Stuart Smith. Human perception and visualization. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 401–406, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press. [57](#)

- [137] Maxime Reynal, Pietro Aricò, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, and Fabio Babiloni. Investigating multimodal augmentations contribution to remote control tower contexts for air traffic management. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP*, pages 50–61. INSTICC, SciTePress, 2019. [196](#), [203](#), [206](#)
- [138] Maxime Reynal, Pietro Aricò, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, and Fabio Babiloni. Involving hearing, haptics and kinesthetics into non-visual interaction concepts for an augmented remote tower environment. In Ana Paula Cláudio, Kadi Bouatouch, Manuela Chessa, Alexis Paljic, Andreas Kerren, Christophe Hurter, Alain Tremeau, and Giovanni Maria Farinella, editors, *Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 73–100, Cham, 2020. Springer International Publishing. [206](#)
- [139] Maxime Reynal, Yvanne Colineaux, André Vernay, and Frédéric Dehais. Pilot flying vs. pilot monitoring during the approach phase: An eye-tracking study. In *Proceedings of the International Conference on Human-Computer Interaction in Aerospace, HCI-Aero '16*, pages 7:1–7:7, New York, NY, USA, 2016. ACM. [206](#)
- [140] Maxime Reynal, Jean-Paul Imbert, Pietro Aricò, Jérôme Toupillier, Gianluca Borghini, and Christophe Hurter. Audio focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility. *International Journal of Human-Computer Studies*, 129:116 – 128, 2019. [196](#), [202](#), [206](#)
- [141] Maxime Reynal, Frank Rister, Sébastien Scannella, Christopher Wickens, and Frédéric Dehais. Investigating pilot’s decision making when facing an unstabilized approach: an eye-tracking study. In *6th Congress of the International Society for Applied Psychology (ISAP 2017)*, pages 1–6, Dayton, US, 2017. [206](#)
- [142] Eckard Riedenklau, Thomas Hermann, and Helge Ritter. Tangible objects and interactive sonification as a scatter plot alternative for the visually impaired. Georgia Institute of Technology, 2010. [65](#)
- [143] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013. [120](#)
- [144] Jonathan C. Roberts and Sabrina Paneels. Where are we with haptic visualization? In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pages 316–323, March 2007. [81](#)
- [145] Barbara Olasov Rothbaum, Larry Hodges, Renato Alarcon, David Ready, Fran Shahar, Ken Graap, Jarrel Pair, Philip Hebert, Dave Gotz, Brian Wills, et al. Virtual reality exposure therapy for ptsd vietnam veterans: A case study. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, 12(2):263–271, 1999. [81](#)
- [146] Diego C. Ruspini, Krasimir Kolarov, and Oussama Khatib. The haptic display of complex graphical environments. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97*, pages 345–352, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. [75](#)
- [147] Jamal Saboune and Juan Manuel Cruz-Hernandez. Haptic effect authoring tool based on a haptification model, May 3 2016. US Patent 9,330,547. [75](#)

- [148] Kenneth Salisbury, David L. Brock, Thomas H. Massie, N. Swarup, and Craig Buchanan Zilles. Haptic rendering: Programming touch interaction with virtual objects. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics, I3D '95*, pages 123–130, New York, NY, USA, 1995. ACM. [xi](#), [75](#), [80](#)
- [149] Sigurd Saue. A model for interaction in exploratory sonification displays. Georgia Institute of Technology, 2000. [55](#), [66](#)
- [150] Anthony Savidis, Constantine Stephanidis, Andreas Korte, Kai Crispian, and Klaus Fellbaum. 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, Assets '96*, pages 117–123, New York, NY, USA, 1996. ACM. [101](#)
- [151] Matti Schwalk, Niko Kalogerakis, and Thomas Maier. 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, 2015. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. [86](#)
- [152] Thomas D. Shannon. Tactile communication attachment, December 18 1973. US Patent 3,780,225. [77](#)
- [153] Edgar A. G. Shaw. *The External Ear*, pages 455–490. Springer Berlin Heidelberg, Berlin, Heidelberg, 1974. [42](#)
- [154] Carl E. Sherrick. Temporal ordering of events in haptic space. *IEEE Transactions on Man-Machine Systems*, 11(1):25–28, March 1970. [77](#)
- [155] Carl E Sherrick and Ronald Rogers. Apparent haptic movement. *Perception & Psychophysics*, 1(3):175–180, 1966. [38](#), [47](#)
- [156] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1997. [72](#), [84](#)
- [157] Brian D Simpson, Douglas S Brungart, Robert H Gilkey, and Richard L McKinley. Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments. Technical report, Wright State University, Department of Psychology, Dayton, OH, 2005. [151](#)
- [158] Alexander W. Skaburskis, Jeffrey S. Shell, Roel Vertegaal, and Connor Dickie. Auramirror: Artistically visualizing attention. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems, CHI EA '03*, pages 946–947, New York, NY, USA, 2003. ACM. [vii](#)
- [159] Aaron E. Sklar and Nadine B. Sarter. Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors*, 41(4):543–552, 1999. PMID: 10774125. [95](#), [100](#), [104](#)
- [160] David Smith, Matthew Donald, Daniel Chen, Daniel Cheng, Changuk Sohn, Aadil Mamuji, David Holman, and Roel Vertegaal. Overhear: Augmenting attention in remote social gatherings through computer-mediated hearing. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05*, pages 1801–1804, New York, NY, USA, 2005. ACM. [55](#), [66](#), [67](#)
- [161] G. Sonet, C. Chnaffa, I. Laumond, C. Castagnié, and R. Mari. On the Influences of k -Vouvray Takes Within a Short Amont of Time. *Transactions on Kinesthetic Interaction*, (3):666, 2019.

- [162] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Estimating focus of attention based on gaze and sound. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, PUI '01, pages 1–9, New York, NY, USA, 2001. ACM. [101](#)
- [163] Robert J. Stone. Haptic feedback: A brief history from telepresence to virtual reality. In *Proceedings of the First International Workshop on Haptic Human-Computer Interaction*, pages 1–16, Berlin, Heidelberg, 2001. Springer-Verlag. [77](#), [79](#), [80](#)
- [164] Bob L Sturm. Surf music: Sonification of ocean buoy spectral data. Georgia Institute of Technology, 2002. [65](#)
- [165] Susanne Tak and Alexander Toet. Towards interactive multisensory data representations. In *GRAPP/IVAPP*, pages 558–561, 2013. [75](#)
- [166] Anthony Talvas, Maud Marchal, Christian Duriez, and Miguel A. Otaduy. Aggregate constraints for virtual manipulation with soft fingers. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):452–461, April 2015. [82](#)
- [167] John C. Tuthill and Eiman Azim. Proprioception. *Current Biology*, 28(5):R194 – R203, 2018. [35](#), [40](#)
- [168] JBF Van Erp, C Jansen, T Dobbins, and HAHC Van Veen. Vibrotactile waypoint navigation at sea and in the air: two case studies. In *Proceedings of EuroHaptics*, pages 166–173, 2004. [73](#), [86](#)
- [169] F.J. van Schaik, J.J.M. Roessingh, G. Lindqvist, and K. Fält. Assessment of visual cues by tower controllers, with implications for a remote tower control centre. *IFAC Proceedings Volumes*, 43(13):123 – 128, 2010. 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems. [15](#), [29](#)
- [170] Joyce Vliegen and A John Van Opstal. The influence of duration and level on human sound localization. *The Journal of the Acoustical Society of America*, 115(4):1705–1713, 2004. [101](#)
- [171] Bruce N. Walker, Amanda Nance, and Jeffrey Lindsay. Spearcons: Speech-based earcons improve navigation performance in auditory menus. Georgia Institute of Technology, 2006. [62](#)
- [172] D. P. Walker. Kepler's celestial music. *Journal of the Warburg and Courtauld Institutes*, 30:228–250, 1967. [59](#)
- [173] Texas M. Ward and Don L. Harlan. Flight telerobotic servicer control from the orbiter. 1989. [vii](#), [77](#)
- [174] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, May 2013. [78](#)
- [175] Karl E. Weick. The vulnerable system: An analysis of the tenerife air disaster. *Journal of Management*, 16(3):571–593, 1990. [146](#), [149](#)
- [176] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *Acoustical Society of America Journal*, 94:111–123, July 1993. [68](#)
- [177] Kenneth Douglas White. *Greek and Roman technology*. Cornell University Press Ithaca, NY, 1984. [59](#)
- [178] Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 86:1–86:12, New York, NY, USA, 2018. ACM. [vii](#), [82](#), [83](#)

- [179] Christopher D Wickens, Anne S Mavor, J McGee, National Research Council, et al. Panel on human factors in air traffic control automation. *Flight to the future: human factors in air traffic control*, 1997. [187](#)
- [180] Gregg Williams. Apple macintosh computer. *Byte*, 9(2):30–31, 1984. [55](#), [66](#)
- [181] Emily Wilson, Melanie Underwood, Olivia Puckrin, Karla Letto, Rebecca Doyle, Holly Caravan, Stacey Camus, and Kate Bassett. The arcsine transformation: has the time come for retirement, 2013. [120](#), [178](#)
- [182] Graham Wilson, Stephen Brewster, Martin Halvey, and Stephen Hughes. Thermal icons: Evaluating structured thermal feedback for mobile interaction. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 309–312, New York, NY, USA, 2012. ACM. [72](#), [84](#)
- [183] Graham Wilson, Thomas Carter, Sriram Subramanian, and Stephen A. Brewster. Perception of ultrasonic haptic feedback on the hand: Localisation and apparent motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1133–1142, New York, NY, USA, 2014. ACM. [72](#), [83](#)
- [184] Graham Wilson, Thomas Carter, Sriram Subramanian, and Stephen A. Brewster. Perception of ultrasonic haptic feedback on the hand: Localisation and apparent motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1133–1142, New York, NY, USA, 2014. ACM. [72](#), [83](#)
- [185] David Worrall. Sonification: A prehistory. Georgia Institute of Technology – International Community on Auditory Display, 2018. [59](#)
- [186] Pavel Zahorik, Douglas S. Brungart, and Adelbert W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420, 2005. [45](#)
- [187] Ji-Hu Zhang, Thomas D. Y. Chung, and Kevin R. Oldenburg. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of Biomolecular Screening*, 4(2):67–73, 1999. PMID: 10838414. [123](#)
- [188] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. *SIGCHI Bull.*, 18(4):189–192, May 1986. [78](#)
- [189] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, CHI '87, pages 189–192, New York, NY, USA, 1987. ACM. [vii](#), [78](#)

APPENDICES

First experimental campaign

Ethics form

CERNI APPLICATION FORM

The CERNI reviews research protocols undertaken by or under the supervision of a tenured researcher or associate professor belonging to the community of research institutions of the Federal University of Toulouse.

Project summary

Project title:

Measurement in a Remote Tower platform (ROT).

Scientific field:

Visual, auditory and vibrotactile perception, Human-Machine Interaction, Embodied Cognition, Sens of presence.

Tenured researcher – project's scientific supervisor:

Christophe HURTER, professor at Interactive Data Visualization group of the French Civil Aviation University (ENAC)

Data, Economics, and Interactive Visualization (DEVI team)

ENAC, Toulouse, France.

Mail: christophe.hurter@enac.fr

Tel: 05 62 25 96 27

Other researchers taking part in the project:

Pietro ARICO, Sapienza University (UniSap), Roma, Italy.

Raïlane BENHACENE, French Civil Aviation University (ENAC), Toulouse, France.

Viviana BETTI, Sapienza University (UniSap), Roma, Italy.

Charles CLERCQ, French Civil Aviation University (ENAC), Toulouse, France.

Jelena DOKIC, Deep Blue Consulting & Research (DBL), Roma, Italy.

Ana FERREIRA, Deep Blue Consulting & Research (DBL), Roma, Italy.

Jean-Paul IMBERT, French Civil Aviation University (ENAC), Toulouse, France.

Han KRUIGER, University of Groningen (RuG), Groningen, Netherlands.

Matteo MARUCCI, Sapienza University (UniSap), Roma, Italy.

Rosa DE PIANO, Deep Blue Consulting & Research (DBL), Roma, Italy.

Simone POZZI, Deep Blue Consulting & Research (DBL), Roma, Italy.

Maxime REYNAL, French Civil Aviation University (ENAC), Toulouse, France.

Alexandru TELEA, University of Groningen (RuG), Groningen, Netherlands.

Michela TERENCE, Deep Blue Consulting & Research (DBL), Roma, Italy.

Jerome TOUPILLIER, French Civil Aviation University (ENAC), René Descartes University, Paris, France.

Research premises:

Ecole Nationale de l'Aviation Civile – French Civil Aviation University

7 avenue Edouard Belin

31400 – Toulouse

France

Main objective:

We would like to study the contribution of new techniques of interaction. The context is a Remote and Augmented Control Tower environment for Air Traffic Control. Its aim is to enhance the localization of spatial events (in our case it will be aircraft located in the airport vicinity). This new interaction uses spatialized sound through headphones, and haptic feedbacks through a wooden chair on which vibrotactile transducers have been attached.

I acknowledge that the recommendation given by the CERNI only relates to the research project described in this document.

Date: 02/09/2017

Electronic signature of the scientific supervisor:



1. SUCCINCT DESCRIPTION OF THE PROJECT

Background and scientific interest

ATM Human Performance research has been traditionally focused on two senses: sight and hearing. Remote tower operations make no exception, with many efforts and resources focused on acquisition of visual images, for instance by means of high-resolution cameras. This situation is often understood by adopting traditional human information processing approaches, where human cognition is described as composed by the three phases of input: acquisition – processing – action, with a clear and neat separation among them. MOTO project (reMOte TOwer) will explore three research opportunities.

The first one is to consider the role of all the human senses for tower operations. The approach of Embodied Cognition (Wilson, 2002) could be applied to achieve a full understanding on the use of other senses for controllers, i.e. one not deprived of important bodily sensations. The second one is that the Embodied Cognition approach also shows how the three phases of human cognition cannot be neatly divided, as decision-making is closely integrated with our perceptual capability and action possibilities. The Embodied Remote Tower could potentially open new possibilities to study (and reproduce) advanced forms of naturalistic decision-making, or attentional mechanism like the cocktail party effect. Third, the understanding of embodied aspects of Air Traffic Management (ATM) Human Performance is a prerequisite to design effective multimodal input and output channels, thereby rethinking the current Human-System interaction model. The end goal is to enhance human performance, by exploiting other channels than the already overloaded visual one.

The experiment described in this document is conducted under the MOTO project (SESAR, 2016) through these last three points. Its goal is to test if a new interaction modality based on spatial hearing and vibrotactile feedbacks can enhance the localization of aircraft in the airport vicinity by augmenting senses of hearing and touch, by giving cues using vibrotactile feedbacks for the latter.

Aims

According to Air Traffic Controllers (ATCOs), vision, touch and hearing sensations are key points for the understanding of how they are immersed into the Control Tower environment. But these sensations are mostly perceived through the sense of vision. However, hearing and touch can also be considered to give information by ways of augmentation, especially into degraded meteorological conditions. In a Control Tower environment, visual information, vibrotactile feedbacks and spatialized sounds come to ATCOs through their visual, touch and hearing channels. The runway environment limits the scope of spatialized sound: it is difficult to perceive sounds coming from flying aircraft (which are farthest compared to aircraft located on the ground). In the context of MOTO project, we propose to use a spatialization paradigm to enrich the perception of aircraft position in the airport vicinity, by associating to them engine sounds played in loop. In another way, this experiment is about increasing the information allowing to deduce the position of aircraft by placing it in a spatialized context with new interaction modalities.

General hypotheses

Spatial hearing consciousness describes the capability to analyze spatial properties of sound sources in a complex auditory environment (i.e., in our case: multiple sound sources, degraded weather conditions). One of the most important feature of this capability is to locate spatial auditory events. They have been studied traditionally (and mainly) in anechoic rooms, where sound target is the only one in the environment. This does not reflect the complexity of auditory scene in real world, in which sound sources are not unique and constantly fluctuant (i.e. sound sources are constantly deleted or added). In the context of a computer simulated environment, it is hard to understand how human listeners are able to cope with multiple and simultaneous sound sources (Brown & Cooke, 1994). Most of our knowledge on spatial hearing is referring to the evaluation of the listeners' ability to isolate sound sources in anechoic environment (Middlebrooks & Green, 1991) (Wightman & Kistler, 1989). A limited number of studies were carried out to analyze location capability of sound sources also in echoic environment (Hartmann, 1983) (Shinn-Cunningham, Kopco, & Martin, 2005) and in multi-sources environment in which targets were masked by one or several interfering sound sources (Brungart & Simpson, 2005) (Good & Gilkey, 1996) (Lorenzi, Gatehouse, & Lever, 1999). In a multi-speaker situation, spatial separation between the interlocutors decreases the cognitive load. Relative sound levels between speakers could be a key point to determine the spatial decrease in cognitive load and more precisely prefrontal activity induced by spatial cue in multi-speaker situations (Andéol, Suied, Scannella, & Dehais, 2017).

When an aircraft is on the approach segment, one of the main difficulties encountered by ATCOs is to locate it precisely. Along with that, we also know that human ear is good at detecting sound events located in the horizontal axis (right, left), but encounters difficulties when it comes from the vertical axis. To enhance this horizontal selectivity of human perception of aircraft, we propose to rely on the information of the operator's head orientation so that we can modify the gain of sounds emitted by the aircraft located into the scene. In extenso, if ATCO's head is directed toward an aircraft, its sound will be louder. We call this interaction "sound focusing cone"; it will be explained in detail further

in the document. In the same way, to enhance the vertical selectivity, we propose to use vibrotactile feedbacks coming from the bottom of the seat or the back of the seat to notify respectively to the ATCo when an aircraft is on the ground or in the air.

Here, we made the hypothesis that the coupling of these two new interaction modalities should enhance the localization of the aircraft in the context of Remote Towers, even in complex environment with multiple sound sources. Furthermore, we think that the interaction modality we introduce in this document could improve the localization of the aircraft into bad meteorological conditions.

Conflicts of interest

No conflict of interest is identified for this project.

2. EQUIPMENT AND METHODS

A. Participants

For the calibration phase (pre-tests), a small number of participants (3 to 4 experienced ATCos) will be tested to calibrate the final experiment. This calibration phase will enable us to evaluate the suitability of the experimental design, and to decide on a possible 5-minute pause between the two sets of trials. At the end, we planned to test a minimum of 18 experienced ATCos for the whole experiment.

Recruitment

Recruitment method: A list of experienced ATCos will be contacted by mail.

Place of recruitment: French Civil Aviation University (ENAC), Toulouse, France.

Selection criteria: Participants will be selected regarding their professional certifications. They will be only ATCos who are certified to be able to control a little airport in Visual Flight Rules (without automatic guidance technology).

Exclusion criteria: Many inter-subject differences in the localization of sounds have been observed, even among young adults with normal audition (Wightman & Kistler, 1989) (Makous & Middlebrooks, 1990) (Wenzel, Arruda, Kistler, & Wightman, 1993). Hence, given the fact that each step of this experiment integrates spatialized sound, it is mandatory here to test every participant to detect those who are not receptive to spatialized sound. To do so, we propose a quick step, just before the experiment will start, during which participants will be asked to localize 10 different sounds. A score over 10 will be assigned to every participant. All those who obtain a score lower than 8/10 will be dismissed.

Possible compensation of subjects:

No compensation will be offered to participants for this experiment.

B. Method

Interaction and feedbacks to enhance the localization of aircraft into a Remote Tower immersive environment

“Sound Focusing Cone” interaction modality

The interaction modality we designed for this purpose and that will be investigated through the experiment described in this document is named Sound Focusing Cone (SFC). This interaction modality was the first we designed. Its principle, described on Figure 1 below, is the following. The head orientation of the participant is retrieved via a gyroscope located upon the headphones. This information is acquired using an Arduino Nano microcomputer and an IMU MPU60501 (Figure 4). Hence, we can detect where the participant is looking at, or at least in which location her or his head is pointing at. A cone defined with a radius of 5 degrees will select sound sources (engine sounds) aligned with its position. Our interaction modality can be qualified with “spatial filtering” (Andéol, Suied, Scannella, & Dehais, 2017): the relative sound levels between sound sources (i.e. aircraft) inside and outside the cone will be adjusted to reduce ATCo’s cognitive load. The gain of the sound sources located inside the cone will be increased, while the gain of sound sources located outside the cone will be decreased. With this type of interaction, we expect that the participant will “play” with sounds she/he is hearing and try to locate them more precisely.

¹ IMUs (for Inertia Measurement Unit), are sensors which are capable to acquire a relative orientation.

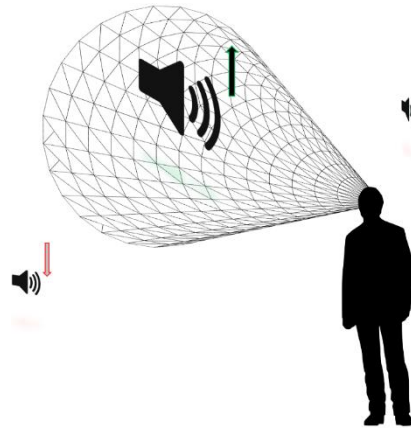


Figure 1. A representation of the Sound Focusing Cone interaction modality (SFC). The gain of every sound sources located in the cone field are increased, while the gain of the other ones, located outside the cone, is decreased.

Vibrotactile feedbacks

The second modality we designed for this experiment is haptic feedbacks (i.e. in our case, vibrotactile), which is here used to map aircraft position on Z axis: down for aircraft located on the ground, and up for aircraft located in the air. We fix two vibrotactile transducers on a chair: one behind the back of the chair and another one under the seat (see Figure 4). The first one is used to notify that an aircraft is in the air, while the second notifies the participant that an aircraft is on the ground. These feedbacks are used here to support SFC interaction modality, especially into bad meteorological conditions. They will be activated only for aircraft located into the field defined by the focusing cone.

Stimuli

Auditory channel: engine sounds. Participants will hear spatialized sounds associated with every aircraft and linked to their static position. The stimulus is a unique engine sound, from a common VFR aircraft, like Robin DR400. Typically, the interaction explained below will act on this/these engine sound(s), helping to locate the related aircraft in a 3-dimensionnal environment, aiming at decreasing ATCos' cognitive load while enhancing their situation awareness. As on average, the minimum time required to locate a sound is approximately 100 ms (Vliegen & Van Opstal, 2004). Engine sounds will be played in loop (i.e. continuously) until the participant is able to locate it in space.

Touch channel: vibrotactile feedbacks. As described before, human is good at locating sounds along the horizontal axis; and the sound interaction described in this document relies on this aptitude. Haptics, and more precisely vibrotactile feedbacks, will come with this modality with the goal to map vertical axis by giving information on the position of aircraft located on the ground (i.e., “down” modality) or in the air (i.e., “up” modality). With this second modality, we want to increase vertical selectivity by adding the information which will notify the participant if the aircraft she/he is aiming at (i.e. the aircraft which is inside the focusing cone, which will be located on the ground or in the air).

Visual channel: Flight Gear graphical simulated environment. The experiment will take place in a 3D environment generated with Flight Gear (FG) flight simulator. Hence, visual stimuli are also to be considered. Participants will probably be able to see the aircraft through FG graphics: visual cues about aircraft position(s) will be implicitly given to them. A similar approach would be adopted in a “normal” ATC environment: visuals are here given to the participants to reproduce their working environment as ecologically as possible.

Material

Airport discrimination with selectable areas

In an airport, aircraft are following a specific circuit. For this experiment, in which we want to test our augmentation means into a simple context (one aircraft), or in a more complex one (multiple simultaneous aircraft), Muret airport have been discriminated into five different areas (Figure 2). From the point of view of its Control Tower, which is located in front of the runway, we want to discriminate the right, the left, the space located in front of the Control Tower, and a farther one. Regarding this approach, and regarding how spatialized sound can be manipulated, the choices we made for these five areas are the following ones:

- *Take-Off* and *Crosswind* legs (purple);
- *Downwind* leg (red);
- *Base and Final* legs (yellow);
- *Runway, west part* (green);
- *Runway, east part* (blue).

Aircraft (i.e. sound sources) will be located in the center of each area. In this way, there cannot be more than one plane

in the axis of the participant. An area cannot contain more than one aircraft at a time.



Figure 2. Generic scheme of an airport circuit (left), and discrimination of the Muret airport vicinity (right) into five distinct areas regarding this circuit.

Answering HMI

Multiple static aircraft configurations (combinations) will be displayed through FG, regarding these five areas and the two difficulty levels. All the participants will hear the same configurations, in a random order for each modality and difficulty level. By clicking in the corresponding area(s) through a specific HMI, participants will be able to explain where the sound(s) they hear is (are) coming from, according to their perception. With the purpose of simplifying the installation, they will take place in front of the separation between west and east parts of the runway (*Runway east* and *Runway west* areas).

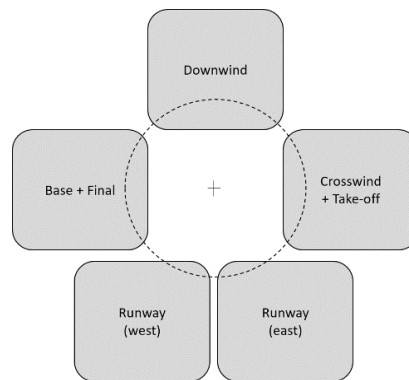


Figure 3. A scheme of the HMI representing the five distinct areas on which participants will have to fill their answers.

As shown on the figure above, this HMI will be used to register the participants' answers. The distance between the center of this HMI and each of its button is constant to minimize and to standardize participants' movements while they are answering. In addition, every button is the same color to not influence their answers. Buttons will have two positions: selected (dark grey) or not (light grey). A validation button will come at the bottom of the window to validate their response. This last button will be displayed in green, and will be positioned at the right for right-handed participants, or at the left for left-handed ones (see Figure 5).

Setup

The setup for the experiment is composed of: 8 UHD screens (Iiyama Prolite X4071), an Alienware Area51 computer equipped with two nVidia GeForce GTX 1080 graphic cards, an "haptic chair" (i.e. a wooden chair on which two Clark Synthesis T239 Gold tactile transducers have been attached, one behind the back to code the "up" modality, and another one under the seat to code the "down" modality) and a stereo headphone (Plantronics BackBeat Pro) equipped with an IMU connected to a microcomputer Arduino Nano (Figure 4).

The 3D environment is made up with Flight Gear2 (FG) flight simulator. Positions taken by aircraft in the different combinations are parametrized through FG configuration files. Spatial sound software is developed at ENAC, using the OpenAL 1.13 library. Network communications are developed using ENAC Rejeu and Ivy4 bus technologies. All the modules used for the experiment will be synchronized with NTP 5 protocol.

² Flight Gear is a free flight simulator distributed under GNU licence. It is mainly written in C++.

³ OpenAL 1.1 (for *Open Audio Library*) is a free multiplatform sound library, giving an API (Application Programming Interface) mainly used for video games, but also for any other purposes that need to spatialize sound sources.

⁴ Ivy is a simple protocol and a set of open-source (LGPL) libraries and programs that allows applications to broadcast information through text messages, with a subscription mechanism based on regular expressions (regex).

⁵ NTP, for *Network Time Protocol*, is a network protocol used to synchronized all entities connected to the same network.



Figure 4. The 8 screens setup (top left) and Alienware Area51 computer used for the simulation (top middle) coupled with Nvidia GeForce GTX 1080 graphic card (top right top) will display FG graphical environment. Plantronics BackBeat Pro stereo headphones (top right bottom) will play spatialized engine sounds, and the haptic chair set up (bottom left) will give up and down cues through vibrotactile feedbacks using two ClarkSynthesis T239 Gold transducers (bottom middle). An Arduino Nano microcomputer associated to an IMU (bottom right) will retrieve the participant head orientation.

Participants will have another screen in front of them on which the answering HMI will be displayed (Figure 5).

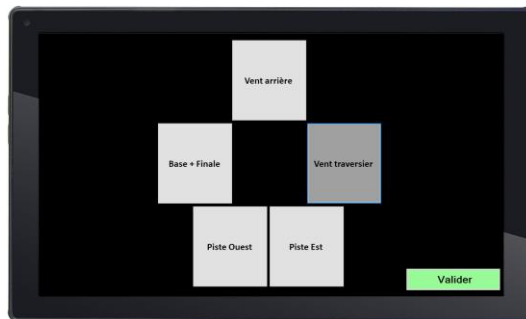


Figure 5. Answering HMI prototype on a tactile tablet (right-handed version).

Finally, the complete experimental setup is visible on Figure 6 below.



Figure 6. Photograph of the complete experimental setup used for the experiment (current setup, prototype).

Experimental flow

Discovery and familiarization

First, all pre-selected participants will be informed by a scientific officer beforehand of the objectives of the study, its methodology, duration, constraints, and foreseeable risks. Participants will be asked to fill the Consent Form (Appendix 1). The experiment will start with a presentation of the stimulus (engine sound). To do so, the engine sound will be listened by the participants from each area. Vibrotactile feedbacks will also be presented together with additional oral explanations. Finally, SFC interaction will be presented, and the participant will have the opportunity to interact with it during approximately 3 minutes with and without vibrotactile feedbacks. At the end of these first presentations, the experimental phase will start. Participants will enter their age, gender, profession, experience (in hours), the results from their last audiogram, and their manual laterality. The experiment will be articulated regarding the Modality condition, and the sets of trials. First, the baseline (3D) will be presented to the participants following four ordered sets of trials. In a second time, it will be the same for the second modality (SFC), and then, in a third time, for the third modality (SFC + V).

Baseline (1st condition)

The experimental steps will start and the participant will be asked to localize different combination of spatialized sounds over the five airport areas described before (see previous section). The same combinations, in the same order, but previously randomly organized, will be presented to all the participants. All the stimuli will be presented randomly following Difficulty level and Weather conditions. This first organization will be the first set of trials. Second, third and fourth ones will be organized following the same criteria, but in different orders (same for all participants).

SFC (2nd condition)

The flow here is the same as the one described at step 1 (baseline), with the difference that SFC interaction will be activated. As well as for step 1, this step will be repeated four times (four sets of trials).

SFC + V (3rd condition)

This step is the same than step 2 (SFC), but here, vibrotactile feedbacks will be activated to support SFC interaction. Vibrations of the seat (“down”) or the back (“up”) in the haptic chair will give additional information concerning the vertical position of aircraft located into the focalization cone.

Debriefing and questionnaires

At the end of the experiment, participants will be asked to fill two different questionnaires: NASA-TLX (Hart and Staveland, 1988), to evaluate their mental subjective load, and SART (Taylor, 1990), for a situation awareness auto-evaluation (Appendixes 2 and 3). Few minutes of directed interview (recorded) could also be considered, to ask participants if they perceived differences between the three phases of the experiment. The HMI on which participants will fill their answers will display black screens between every trial (during the rest time), to notify them that the current trial is over and to prepare for the next one. At the end of each set of trials, participants will be asked if they want to have a five-minute break.

Finally, the next scheme synthesized the whole experiment flow.

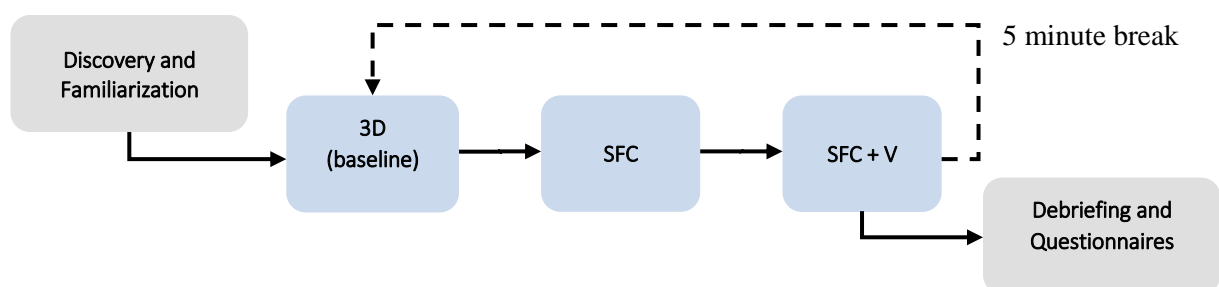


Figure 7. General flow of the experiment.

Study premises

Evaluations/observations schedule

A series of pre-tests will be conducted at the end during November 2017 to calibrate the experiment. We plan to use these tests to refine the experimental protocol. Final experiment will take place during December 2017.

Duration of the study

One evaluation session, including four sets of trials, will be presented to every participant. In the table below, we tried to estimate the total duration time. We estimated 6 seconds for the participant to answer to a combination of sounds placed into airport areas in baseline condition (3D), and 12 seconds to answer for the two others (SFC and SFC + V). Considering a 3-seconds break between every trial, the total experiment duration could be about 47

minutes. If we add 15 minutes for the Discovery and Familiarization phase and 15 minutes for the Debriefing and Questionnaires phase, we can status on an approximate total duration of 1 hour and 20 minutes per participant.

Difficulty level	Weather	Modality	Number of combinations	Number of trials	Estimated duration for one trial (sec)	Resting time between trials (sec)	Estimated total (sec)
1 (easy)	Normal	3D	3	4	6	3	108
		SFC	3	4	12	3	180
		SFC + V	3	4	12	3	180
	Degraded	3D	3	4	6	3	108
		SFC	3	4	12	3	180
		SFC + V	3	4	12	3	180
2 (hard)	Normal	3D	6	4	6	3	216
		SFC	6	4	12	3	360
		SFC + V	6	4	12	3	360
	Degraded	3D	6	4	6	3	216
		SFC	6	4	12	3	360
		SFC + V	6	4	12	3	360
Estimated total (min)							46,8

Table 1. Calculation of the total duration of the experiment for one trial.

We plan to have about 3 participant a day. Thereby, the study should take about 2 weeks to collect data from the 18 participants or more.

Data analysis

Experimental conditions. This experiment aims to quantify the contribution of SFC interaction and vibrotactile feedbacks to the search of aircraft located into the airport vicinity (which are associated with sound sources). Only spatialized sounds will be played during the baseline condition. We consider the contribution of vibrotactile feedbacks alone (i.e. not coupled with SFC interaction) trivial, because it only gives information about the aircraft location on the runway (i.e., “down”, vibrotactile feedbacks spread through the chair seat transducer) or in the air (i.e., “up”, vibrotactile feedbacks spread through the chair back transducer). Thereby, the following conditions are tested:

- Spatialized sound only;
- Spatialized sound with SFC interaction;
- Spatialized sound with SFC interaction and vibrotactile feedbacks.

Metrics. Participants’ performance is measured along two difficulty levels (materialized with the number of simultaneous aircraft) using two different scores (dependent variables):

- *Accuracy to the task*, which is percentage quantifying the number of right answers following the difficulty level;
- *Reaction time*, which is the time took by the participants to detect aircraft into the right area.

Three different aspects will be manipulated during this experiment (independent variables): interaction type, normal or degraded meteorological conditions, and difficulty level (Easy for one aircraft, and Hard for three simultaneous aircraft). More precisely, their values are:

- Visibility (meteorological conditions):
 - *Value 1* – normal/good visibility;
 - *Value 2* – bad visibility (fog/rain);
- Experimental conditions (modality):
 - *Value 1* – Spatialized sound only (baseline, designed with 3D);
 - *Value 2* – SFC interaction (SFC);
 - *Value 3* – SFC interaction and vibrotactile feedbacks (SFC + V);
- Number of aircraft (difficulty):
 - *Value 1* – Easy: 1 engine sound;
 - *Value 2* – Hard: 3 simultaneous engine sounds coming from 3 different aircraft having different positions in the airport.

As explained before, the visibility conditions are here to prove the contribution of SFC interaction and vibrotactile feedbacks. Given that participants will be able to see the airport vicinity through FG, we think that these modalities will be useful into a context of bad visibility (e.g. fog, rain) where they will not have visual cues about aircraft location(s).

C. Predictable and known benefits and dangers for mental and physical health (self-esteem, etc.) and social life (reputation)

This experiment tests new interaction modalities which would be interesting, if results are positive, to integrate into future Remote Towers because it could increase the detection of new aircraft in the ATCo environment. Through spatialized sound, if immersion is enhanced, ATCo's tasks could be executed more accurately and more quickly. If ATCos have a more detailed consciousness of their working environment, security could be increased.

No	Does your protocol include an experimental setup aimed at concealing part of the objective or methodology to subjects or make them believe in other objectives or methodologies?
No	Questions or situations which can put participants ill at ease?
No	Materials which can be considered menacing, shocking or disgusting by participants?
No	Possibility that the participant's private life or his/her family's will be affected, including the use of personal information?
Yes	Use of physical stimuli (auditive, visual, haptic, etc.) other than stimuli associated with normal activities?
No	Deprivation of physiological needs (drinking, eating, sleeping, etc.)
No	Manipulation of psychological or social factors such as sensory privation, social isolation or psychological
No	Physical efforts beyond what can be considered moderate for the participant?
No	Exposure to drugs, chemicals or potentially toxic agents?

No risk has been identified for this experiment. The tasks we ask participants to perform are not correlated with their ability to control airspace. This an exploratory research, which is not directly connected to the real situation and sufficiently far from it. In all cases, personal results will be not available to the participants due to their anonymity.

D. Vigilance/Early interruption of the study

Interruption criteria for a participating subject

Participants have the right to withdraw from the experiment at any time. The experiment is based on voluntary work.

3. DATA PROCESSING – RESPECT OF THE PARTICIPANT'S PRIVATE LIFE

A. Confidentiality

De-identification process

Each participant will be attributed a personal identification number, automatically and randomly generated at the beginning of each session by the module which control the answering HMI.

People having access to data

The results of the sessions, and the questionnaires will be reviewed by Maxime REYNAL. The analysis of the latter results will be part of the deliverable D2.2 of the MOTO project (H2020-SESAR-2015-1, Grant Agreement number: 699379 — MOTO). Christophe HURTER and Jérôme TOUPILLIER will also require access to these data.

B. Storage

Type of data stored (specify if identifying data, directly or by cross checking): Collected information will be anonymously saved into text files. Those text files will be stored into a secure environment, at French Civil Aviation University, with badged access.

Storage premises: French Civil Aviation University (ENAC), Toulouse, France.

Person in charge: Maxime REYNAL

Possibility to destroy data upon participant's request: Yes

4. INFORMED CONSENT FORM INCLUDING INFORMATION FOR PARTICIPANTS

Precisions on the information given to participants:

Any participant preselected will be previously informed by the scientific supervisor of the objectives of the study, of its methodology and duration, of constraints and of predictable risks. A summary of the information given by the

scientific advisor will be provided in the consent form (Appendix 1).

Precisions on the signature and delivery to the participant of the informed consent:

After making sure the information provided is well understood, the scientific supervisor will request from the participant his/her consent to take part in the study. If he/she accepts, the participant will sign the consent form in two copies prior to the study (Appendix 1). The scientific supervisor will keep one copy and give the other to the participant.

Specify how the informed consent will be obtained:

Participants will be asked to fill the Consent Form just before starting the experiment. They will receive a copy of the document.

Bibliography

- 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*, 457-464.
- Arduino. (s.d.). Récupéré sur Wikipedia: <https://en.wikipedia.org/wiki/Arduino>
- Brown, G. J., & Cooke, M. (1994). Computational auditory scene analysis. *Computer Speech & Language*, *8*, 297-336.
- Brungart, D. S., & Simpson, B. D. (2005). Optimizing the spatial configuration of a seven-talker speech display. *ACM Transactions on Applied Perception (TAP)*, *2*, 430-436.
- Chocholle, R. (1940). Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *L'ann{\'e}e psychologique*, *41*, 65-124.
- Good, M. D., & Gilkey, R. H. (1996). Sound localization in noise: The effect of signal-to-noise ratio. *The Journal of the Acoustical Society of America*, *99*, 1108-1117.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.
- Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, *74*, 1380-1391.
- Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *105*, 3454-3463.
- Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The journal of the Acoustical Society of America*, *87*, 2188-2200.
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual review of psychology*, *42*, 135-159.
- SESAR. (2016, June). *MOTO, the Embodied Remote Tower*. Récupéré sur www.moto-project.eu: <http://www.moto-project.eu/wp-content/uploads/2016/11/MOTO-project-Factsheet-SESAR-Innovation-Days-SIDs-2016.pdf>
- Shinn-Cunningham, B. G., Kopco, N., & Martin, T. J. (2005). Localizing nearby sound sources in a classroom: Binaural room impulse responses. *The Journal of the Acoustical Society of America*, *117*, 3100-3115.
- Taylor, R. M. (1990). Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53)*.
- 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*, 1705-1713.
- 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*, 111-123.
- Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, *85*, 868-878.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625-636.

APPENDIX 1 – INFORMED CONSENT AND INFORMATION FORM

Project title: Study of the location of spatialized elements in a remote and augmented control tower environment.

Tenured researcher – project's scientific supervisor: Christophe Hurter, professor at Interactive Data Visualization group of the French Civil Aviation University (ENAC), Data, Economics, and Interactive Visualization (DEVI team)

Mail: christophe.hurter@enac.fr / **Tel:** 05 62 25 96 27

Premises for the study: French Civil Aviation University (ENAC), Toulouse, France.

Aim of the research project: We wish to study the contribution of a new form of interaction which aim is to improve the detection of elements (aircraft) placed in a Remote Control Tower environment. This interaction uses spatialized sound through a stereo headphone, and so-called vibrotactile returns, which are vibrations that you sense in the seat and back of the chair on which you will be seated.

What is expected from you (methodology)

If you accept to participate in this study, you will take part in an experiment where you will be listening to one or more spatialized engine sounds simultaneously. You will have the possibility, at different times, to use a new interaction that will act on this sound spatialization. You will also sense vibrations under the seat of the chair, and behind its back. You will be placed into environment that will materialize a Remote Control Tower context. For each sound you will be listening to, you will be asked to locate it/them and fill your answer by clicking on different buttons corresponding to areas which are discriminating the airport vicinity. We will record your answers, reaction times, and date and hours of every trials. At the end of the experiment, you will fill out a questionnaire in which you will provide details about your feelings during the experimentation.

Rights to withdraw from the study at any time

By agreeing to participate in this experience, you understand that:

- Your participation is entirely voluntary, and will not be remunerated;
- You may withdraw or ask to cease the experiment at any time for any reason whatsoever;
- A possible termination of your participation will have no impact on your notes, your status or your future relations with the team of the ENAC Department of Data, Economy and Interactive Visualization (DEVI).

Rights to confidentiality and respect of private life

By agreeing to participate in this research, you understand that:

- The data to be obtained will be treated with the utmost confidentiality;
- Your identity will be veiled, using a random number;
- As a result, no link can be established between your identity and the data collected;
- The data will be kept in a secure location, and only the Scientific Officer, as well as the assistant researchers, will have access to it;
- Due to the complete anonymization of your data, their destruction or rectification will not be possible once the experimental campaign has ended.

Benefits

The expected benefits of this study are to gain a better understanding of factors influencing the way ATCos searches for localized aircraft in the airport vicinity. A better understanding of these factors will contribute to improving control tasks in a Remote Control Tower context.

Potential risks

To our knowledge, this study entails no risk or inconvenience other than in ATCos' professional life. Head movements are recorded using a device which retrieve relative head orientation. Vibrotactile feedbacks will give cues about aircraft vertical position with little vibrations, like vibrations we can sense with smartphones notifications.

Diffusion

The results of this research may be published in a scientific publication, in a national or international conference, or in a specialized journal. You can ask questions about the research at any time by contacting the project's Scientific Officer by e-mail at christophe.hurter@enac.fr, or by phone at 05 62 25 96 27.

Rights to ask questions at all time

You may ask questions about the study at all time by writing to the scientific supervisor of the project by email at maxime.reynal@enac.fr (05 62 25 96 39) or christophe.hurter@enac.fr (05 62 25 96 27).

Participation consent

By signing this consent form, you certify that you have read and understood the above information, that your questions have been answered in a satisfactory manner and that you have been advised that you are free to cancel your consent or withdraw from this search at any time, without prejudice.

I have read and understood the information above and I accept of my own free will to take part in this study.

First name, last name – Date – Signature:

SITUATION AWARENESS RATING TECHNIQUE (SART; Taylor, 1990)

Instability of Situation
How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?

|-----|
1 2 3 4 5 6 7

Complexity of Situation
How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?

|-----|
1 2 3 4 5 6 7

Variability of Situation
How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing (Low)?

|-----|
1 2 3 4 5 6 7

Arousal
How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?

|-----|
1 2 3 4 5 6 7

Concentration of Attention
How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

|-----|
1 2 3 4 5 6 7

Division of Attention
How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

|-----|
1 2 3 4 5 6 7

Spare Mental Capacity
How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?

|-----|
1 2 3 4 5 6 7

Information Quantity
How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?

|-----|
1 2 3 4 5 6 7

Familiarity with Situation
How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?

|-----|
1 2 3 4 5 6 7

CERNI form (in French)



Toulouse, le lundi 25 septembre 2017

Pr Christophe Hurter
ENAC Toulouse

CERNI : Comité d’Ethique sur les Recherches Non-Interventionnelles

Le bureau (Courriel : bureau-cerni@univ-toulouse.fr)

Objet : Avis sur le projet « Visual, auditory and vibrotactile perception, Human-Machine Interaction, Embodied Cognition, Sens of presence » présenté par Pr Christophe Hurter de l’ENAC, Toulouse.

Monsieur,

Compte tenu des éléments fournis dans votre demande, le Comité d’Ethique pour les Recherches Non-Interventionnelles émet un **AVIS FAVORABLE**. Néanmoins, nous vous signalons qu’il convient de présenter aux participants un formulaire de consentement éclairé en français.

Nous rappelons, par ailleurs, qu’il relève de la responsabilité des chercheurs de se conformer à leurs obligations légales notamment en ce qui concerne les aspects « informatique et liberté », ou encore l’homologation du lieu de recherche.

Numéro de l’avis : CERNI-Université fédérale de Toulouse-2017-046

A Toulouse,

Le bureau du CERNI : Pr Maria Teresa Munoz Sastre

Pr Jacques Py

CERNI - Université Fédérale Toulouse Midi-Pyrénées

41, Allées Jules Guesde - CS 61321 - 31013 Toulouse CEDEX 6 - Tél. : 05 61 14 80 10 - Télécopie : 05 61 14 80 20

Courriel : contact@univ-toulouse.fr

www.univ-toulouse.fr

Consent form (in French): Formulaire de consentement pour participer à une expérience

Formulaire de consentement pour participer à une expérience

Titre du projet : MOTO – the reMOte TOWER

Nom de l'expérience : Etude de la localisation d'éléments spatialisés à l'aide de modalités d'interaction sonore

Chercheur titulaire, responsable scientifique du projet : Christophe Hurter, professeur à l'Ecole Nationale de l'Aviation Civile (ENAC) au sein du groupe de recherche Visualisation Interactive de Données, équipe Données, Economie et Visualisation Interactive (DEVI).

Mail : christophe.hurter@enac.fr / **Tél. :** 05 62 25 96 27

Lieu de l'expérience : Ecole Nationale de l'Aviation Civile (ENAC), Toulouse, France.

But du projet de recherche

Nous souhaitons étudier la contribution d'une nouvelle modalité d'interaction dont le but est d'améliorer la détection d'éléments (avions) placés dans un environnement pseudo-écologique de tour de contrôle déportée. Cette modalité agit sur le son, qui est spatialisé à l'aide des hauts parleurs des écrans qui retranscrivent l'environnement de l'aéroport. Elle est également couplée à des feedbacks vibratoires diffusés au-travers de la chaise sur laquelle vous serez assis durant l'expérience.

Ce qui est attendu de vous

Si vous acceptez de participer à cette étude, vous prendrez part à une expérience durant laquelle vous serez immergé au sein d'un banc d'essai simulant l'environnement d'une tour de contrôle déportée. Vous devrez vous concentrer sur ces sons spatialisés de moteurs provenant simultanément d'un ou de plusieurs avions. Vous aurez la possibilité, à différents moments, d'utiliser une modalité d'interaction qui agit sur cette spatialisation sonore à l'aide d'un casque (Microsoft HoloLens) que vous devrez porter sur votre tête. Vous ressentirez, également à différents moments, des vibrations sous l'assise et dans le dossier de la chaise sur laquelle vous serez assis. Pour chacun des sons que vous percevrez, vous devrez renseigner leur localisation à l'aide d'une IHM de réponse proposant cinq boutons différents. Ces cinq boutons correspondent aux cinq zones qui discriminent l'aéroport. Nous vous expliquerons en début de passation comment l'expérience est conçue. Nous enregistrerons votre temps de réaction, la justesse de votre réponse, ainsi que la date et l'heure de tous les trials qui vous seront présentés. Pour certains participants, l'activité du cerveau sera également enregistrée grâce à un électro-encéphalogramme (EEG) qui sera positionné sur votre tête à l'aide d'un bonnet. Si c'est votre cas, nous devons utiliser un gel spécial pour améliorer la conductivité des électrodes, qui se déposera sur vos cheveux. Il faudra donc prévoir un temps supplémentaire après l'expérience pour pouvoir de les laver. Dans tous les cas, après l'expérience vous serez invité à répondre à des questionnaires pour nous donner des détails sur votre ressenti.

Droit de vous retirer à tout moment

En acceptant de participer à cette expérience, vous comprenez que :

1. Votre participation est tout à fait volontaire, et ne sera pas rétribuée ;
2. Vous pouvez vous retirer ou demander de cesser l'expérimentation à tout moment et pour quelque motif que ce soit ;
3. Un arrêt éventuel de votre participation n'aura aucun impact sur votre statut ou vos relations futures avec l'équipe Données, Economie et Visualisation de Données de l'ENAC.

Bénéfices

Les bénéfices attendus par cette étude sont d'obtenir une meilleure compréhension de l'environnement de l'aéroport, notamment en situation météorologique dégradée (très faible visibilité), ce qui pourrait être une voie d'augmentation de la tâche des contrôleurs aériens en tour déportée.

Risques potentiels

A notre connaissance, cette étude ne fait encourir aucun risque ou gêne aux personnes qui y participent.

Droit de confidentialité et respect de la vie privée

En acceptant de participer à cette étude, vous comprenez que :

1. Les données qui seront obtenues seront traitées avec la plus entière confidentialité ;
2. Votre identité sera voilée, à l'aide d'un numéro ou d'une lettre anonymisé ;
3. De ce fait, aucun lien ne pourra être établi entre votre identité et les données recueillies ;
4. Ces données seront conservées dans un endroit sécurisé, et seul le responsable scientifique, ainsi que les chercheurs adjoints y auront accès ;
5. Compte tenu de l'anonymisation totale des données, leur destruction ou rectification ne sera pas possible une fois la campagne d'expérimentation terminée.

Diffusion

Les résultats de cette recherche pourront faire l'objet d'une ou de plusieurs publications scientifiques, dans un colloque ou une conférence à portée nationale ou internationale, et/ou une revue spécialisée.

Droit de se renseigner

À tout moment, vous pourrez poser des questions au sujet de la recherche en communiquant avec le responsable scientifique du projet ou les personnes étant impliquées, par courrier électronique ou téléphone à maxime.reynal@enac.fr (05 62 25 96 39) ou christophe.hurter@enac.fr (05 62 25 96 27).

Consentement à la participation

En signant ce formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, qu'on a répondu à vos questions de façon satisfaisante et qu'on vous a avisé que vous étiez libre d'annuler votre consentement ou de vous retirer de cette recherche en tout temps, sans préjudice.

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré de participer à cette étude.

Nom, Prénom – Date – Signature :

Welcome text (in French)

Bonjour,

Vous avez accepté de prendre part à notre expérience, et nous vous en remercions !

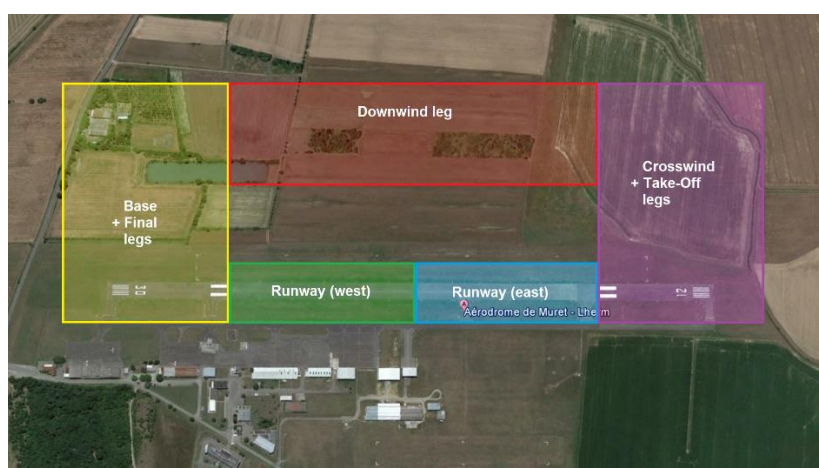
Comme vous le savez peut-être déjà, cette étude s'inscrit dans le cadre du projet **MOTO**. Celui-ci vise à concevoir une **tour déportée** pour ensuite lui ajouter de **nouvelles fonctionnalités** qui pourraient améliorer les outils et l'environnement de travail des contrôleurs aériens.

Dans ce but, cette expérience nous sert à tester une nouvelle forme d'interaction que nous avons imaginé. L'idée est **d'améliorer la localisation d'avions** situés dans l'environnement de l'aéroport, notamment **lorsque la visibilité est mauvaise**.

Cette interaction agit sur le son, qui est spatialisé. L'expérience va consister en ceci : par le biais de notre banc de test, nous allons vous présenter une certaine quantité de stimuli **visuels, auditifs et vibratoires**. Vous devrez nous indiquer où vous percevrez les avions, suivant les canaux sensoriels qui seront mis en jeux.

Pour des raisons de simplification, **les avions auront des positions statiques**. Ils seront positionnés dans des zones qui discriminent l'aéroport. Il n'y aura **pas plus d'un avion en même temps au sein d'une même zone**, et **1 seul, 2 ou bien 3 avions simultanément** en tout et pour chaque trial, **pas plus**.

L'aéroport (Muret) est découpé de la manière suivante :



Les 5 zones sont donc, en français :

- **Piste Ouest** (en vert)
- **Piste Est** (en bleu),
- **Vent traversier** (en mauve),
- **Vent arrière** (en rouge),
- **Base + Finale** (en jaune).

L'interface que vous devrez utiliser pour répondre reprend exactement ces 5 zones. Vous serez positionné **derrière la séparation entre les deux pistes** (Ouest et Est). **Tous les avions seront donc situés devant vous** (vous n'entendrez aucun son derrière vous).

Au cours de l'expérience, vous aurez l'occasion d'utiliser **3 types de modalité** : **son spatialisé uniquement**, un « **Cône de Focalisation Sonore** », et **ce même cône couplé à des vibrations** dans la chaise sur laquelle vous êtes assis. Le Cône de Focalisation Sonore amplifiera le son **provenant des avions qui seront situés dans l'axe de votre tête**, tandis que **le son des autres avions sera diminué**. Les vibrations, lorsqu'elles seront activées, ne pourront être perçue **que lorsqu'un avion sera situé dans ce « cône »**, pour vous signifier que l'avion concerné est **en l'air** (auquel cas vous ressentirez des vibrations dans le dossier de la chaise), ou bien **au sol** (vibrations sous l'assise). Vous serez informé lorsque ces différentes modalités seront activées.

Pour terminer, les **conditions météorologiques** alterneront entre **bonne visibilité** (vous pourrez voir les avions), et **brouillard épais** (vous ne verrez rien).

Si vous avez des questions, n'hésitez pas. Installez-vous confortablement, nous allons pouvoir commencer l'expérience !

Post-experiment questionnaires (in French)

The next figures are all the pages of the post-run questionnaire provided to the participants at the end of the experiment during the first campaign

- Individual informations: Figure 1 (p. 246);
- General questions: Figure 2 (p. 247);
- Usability: Figure 3 (p. 248);
- Fatigue: Figure 4 (p. 249);
- NASA-TLX (3DS): Figure 5 (p. 250);
- NASA-TLX (AF): Figure 6 (p. 252);
- NASA-TLX (AF+V): Figure 7 (p. 252);

Vous avez participé à une expérience visant à tester la contribution de nouvelles modalités sonores et haptiques dans le cadre des tours de contrôle déportées et augmentées.

Pour terminer complètement l'expérience, nous avons besoin de recueillir vos impressions et ressentis. Pour ce faire, nous vous demandons de répondre aux questionnaires ci-après.

Ceci ne vous prendra qu'une dizaine de minutes environ.

Merci encore !

Code participant

Votre réponse

Sexe

Sélectionner ▼

Âge

Votre réponse

Dénomination exacte du statut (ICA, Instructeur TS, Assistant sub, élève, etc.)

ICA

Instructeur TS

Assistant sub

PC

Élève ICNA

Autre : _____

Nombre d'années de PC en tour

Votre réponse

Si vous n'exercez plus de mention d'unité, depuis combien de temps ?

Votre réponse

Figure 1

Questions d'ordre général

En opérationnel

Avez-vous déjà mal localisé un avion ?

- Oui
 Non

Si oui, par quel(s) moyen(s) avez-vous détecté cette erreur de localisation ?

Votre réponse

Durant cette expérience

Quelle modalité vous a-t-elle le plus aidé à localiser les avions :

- Spatialisation sonore seule
 Cône de focalisation sonore
 Cône de focalisation sonore couplé aux vibrations

De manière générale durant l'expérience, diriez-vous que vous avez mal localisé :

- Aucun avion
 Un nombre faible d'avions
 Quelques avions
 Beaucoup d'avions
 Un très grand nombre d'avions

Spatialisation sonore seule

Pensez-vous que la spatialisation sonore puisse constituer une aide ou au contraire une gêne pour localiser des avions ?

	1	2	3	4	5	6	7	
Une gêne	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Une aide

Auriez-vous des remarques par rapport au concept de la spatialisation des sons de moteurs d'avions ?

Votre réponse

Cône de Focalisation Sonore

Pensez-vous que le cône de focalisation sonore puisse constituer une aide ou au contraire une gêne pour localiser des avions ?

	1	2	3	4	5	6	7	
Une gêne	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Une aide

Auriez-vous des remarques par rapport à ce concept de cône de focalisation sonore ?

Votre réponse

Vibrations

Pensez-vous que le cône de focalisation sonore couplé aux vibrations puisse constituer une aide ou au contraire une gêne pour localiser des avions ?

	1	2	3	4	5	6	7	
Une gêne	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Une aide

Auriez-vous des remarques par rapport au concept de cône de focalisation sonore couplé aux vibrations ?

Votre réponse

Figure 2

Utilisabilité

Sur une échelle de 1 (Très difficile) à 7 (Très facile), évaluez la facilité d'utilisation des différentes modalités d'interaction qui vous ont été présentées durant l'expérience.

Spatialisation sonore seule :

	1	2	3	4	5	6	7	
Très difficile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Très facile

Cône de focalisation sonore :

	1	2	3	4	5	6	7	
Très difficile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Très facile

Cône de focalisation sonore couplé aux vibrations :

	1	2	3	4	5	6	7	
Très difficile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Très facile

Figure 3

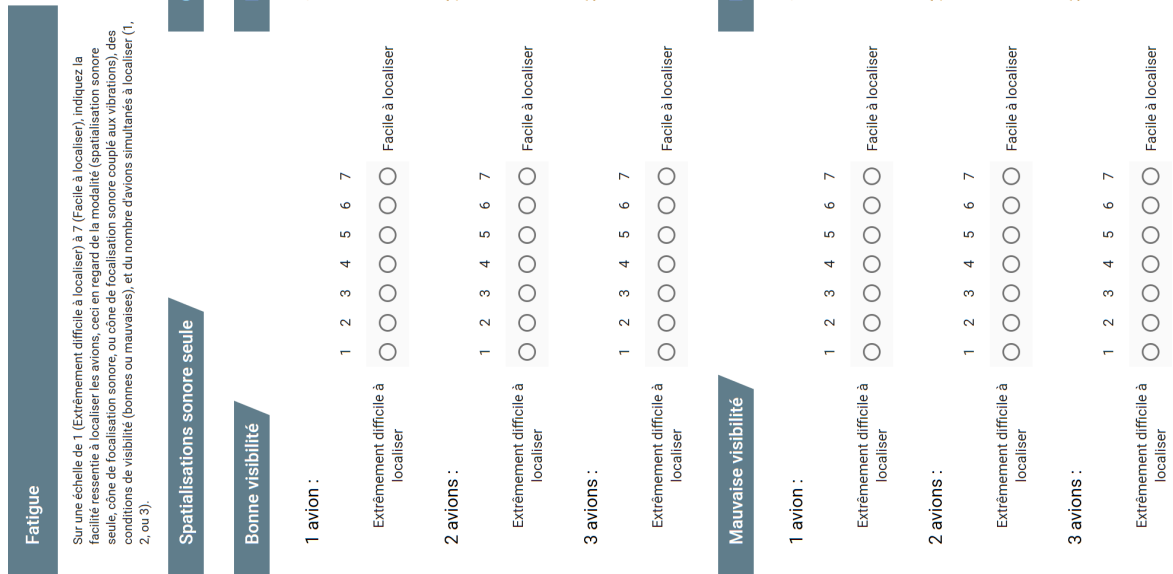


Figure 4

Spatialisation sonore seule

Exigence mentale

Lorsque la modalité "spatialisation sonore seule" été activée, quel degré d'activité mentale était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence physique

Lorsque la modalité "spatialisation sonore seule" été activée, quel degré d'activité physique était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence temporelle

Lorsque la modalité "spatialisation sonore seule" été activée, quelle pression temporelle avez-vous ressenti pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevée

Performance

Lorsque la modalité "spatialisation sonore seule" été activée, avec quel succès pensez-vous avoir réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Echec Réussite

Effort

Lorsque la modalité "spatialisation sonore seule" été activée, quel effort deviez-vous fournir pour accomplir la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Frustration

Lorsque la modalité "spatialisation sonore seule" été activée, avez-vous ressenti, durant votre tâche, de l'insécurité, du découragement, de l'irritation, du stress ou de l'agacement ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Figure 5

Cône de focalisation sonore

Exigence mentale

Lorsque la modalité "cône de focalisation sonore" été activée, quel degré d'activité mentale était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence physique

Lorsque la modalité "cône de focalisation sonore", quel degré d'activité physique était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence temporelle

Lorsque la modalité "cône de focalisation sonore" été activée, quelle pression temporelle avez-vous ressenti pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Performance

Lorsque la modalité "cône de focalisation sonore" été activée, avec quel succès pensez-vous avoir réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Echec Réussite

Effort

Lorsque la modalité "cône de focalisation sonore" été activée, quel effort deviez-vous fournir pour accomplir la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Frustration

Lorsque la modalité "cône de focalisation sonore" été activée, avez-vous ressenti, durant votre tâche, de l'insécurité, du découragement, de l'irritation, du stress ou de l'agacement ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Figure 6

Cône de focalisation sonore couplé aux vibrations

Exigence mentale

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations" été activée, quel degré d'activité mentale était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence physique

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations", quel degré d'activité physique était exigé pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Exigence temporelle

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations" été activée, quelle pression temporelle avez-vous ressenti pour réaliser la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Performance

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations" été activée, avec quel succès pensez-vous avoir réalisé la tâche ?

1 2 3 4 5 6 7 8 9 10

Echec Réussite

Effort

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations" été activée, quel effort deviez-vous fournir pour accomplir la tâche ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Frustration

Lorsque la modalité "cône de focalisation sonore couplé aux vibrations" été activée, avez-vous ressenti, durant votre tâche, de l'insécurité, du découragement, de l'irritation, du stress ou de l'agacement ?

1 2 3 4 5 6 7 8 9 10

Très faible Très élevé

Figure 7

Second experimental campaign

Experimental scripts

The times in the margin correspond to the timestamps in the simulation (10 min per scenario). Aircraft call signs are of the form NAK10X for Muret airport (LFBR), and NAK11X for Lasbordes airport (LFCL). The following abbreviations are used: horodatage

Abbreviation	Meaning
Dct	“direct”
Eng.	“engines”
H-Pnt	“holding point”
L/Up	“line-up”
RQST	“request”
T/O	“take-off”
W/O	“without”

SRT1 and SART1

The script used for scenarios SRT 1 and SART 1 was:

```
--Briefing
102, 103, 104, 105 are parked; 106 will be in DOWNWIND at 3'
--Initialization
    06:55:30 - NAK101: Eng.Start
    06:55:30 - NAK106: Eng.Start
--Start
    06:56:20 - NAK104: Eng.Start
    06:56:30 - NAK102: Eng.Start RQST Eng.Test
    06:58:00 - NAK102: Inform Eng.OK
    06:58:30 - NAK102: RQST Taxi to H-Pnt
    06:59:00 - NAK103: RQST Eng.Start
    06:59:20 - NAK106: Report DOWNWIND
    06:59:30 - NAK103: Start to move W/O clearance (SPATIAL)
    07:00:45 - NAK106: Report BASE and RQST Landing (RWY)
```

07:01:20 - NAK102: Cross H-Pnt W/O clearance (RWY)
07:01:30 - NAK106: Report FINAL
07:01:50 - NAK104: RQST Taxi to H-Pnt
07:02:30 - NAK102: RQST L/Up and T/O
07:02:30 - NAK101: Report DOWNWIND
07:03:20 - NAK101: Report BASE and RQST Landing
07:03:40 - NAK106: Report DOWNWIND
07:04:10 - NAK106: Report BASE and RQST Landing
07:04:35 - NAK101: Report FINAL
07:05:00 - NAK106: Report on FINAL
07:05:30 - NAK104: RQST L/Up and T/O
07:06:00 - NAK105: RQST Eng.Start

--End

SRT2 and SART2

The script used for scenarios SRT 2 and SART 2 was:

--Briefing

101, 102 103 and 106 are parked; 104 will be in DOWNWIND at 3'50'';
105 will be in DOWNWIND at 9'30''

--Start

06:55:30 - NAK104: Eng.Start
06:55:30 - NAK105: Eng.Start
06:56:30 - NAK103: Eng.Start
06:56:30 - NAK101: Eng.Start
06:56:50 - NAK101: RQST Taxi to H-Pnt
06:58:20 - NAK102: Eng.Start
06:58:30 - NAK103: RQST Taxi to H-Pnt
06:59:10 - NAK102: Moves W/O clearance (SPATIAL)
06:59:40 - NAK104: Dct BASE
06:59:50 - NAK104: Report DOWNWIND
07:00:00 - NAK102: RQST Taxi to H-Pnt
07:00:50 - NAK101: RQST L/Up and T/O
07:01:15 - NAK104: Report BASE and RQST Landing
07:01:55 - NAK103: Cross H-Pnt W/O clearance (RWY)
07:02:00 - NAK104: Report FINAL (RWY)
07:02:50 - NAK104: Dct DOWNWD
07:03:00 - NAK103: RQST L/Up and T/O
07:03:30 - NAK104: Report DOWNWIND Dct BASE
07:04:00 - NAK102: RQST L/Up and T/O
07:05:00 - NAK104: Report BASE and RQST Landing Dct FINAL
07:05:30 - NAK105: Report DOWNWIND

07:06:00 - NAK102: Say goodbye
--End

MRT1 and MART1

LFBR (Muret) airport

The script used for scenarios MRT 1 and MART 1 at LFBR (Muret) airport was:

--Briefing
101 will be in BASE at about 1'; 102 is parked; 104 will be in
DOWNWIND at 3'40''
--Start
06:55:30 - NAK101: Eng.Start
06:55:30 - NAK104: Eng.Start
06:55:30 - NAK105: Eng.Start
06:56:30 - NAK102: Eng.Start
06:57:00 - NAK101: Report BASE and RQST Landing
06:58:30 - NAK102: RQST Taxi to H-Pnt (CALL)
06:59:40 - NAK104: Report DOWNWIND
07:01:25 - NAK104: Report BASE and RQST Landing
07:02:00 - NAK104: Report FINAL (RWY)
07:02:00 - NAK102: Cross H-Pnt W/O clearance (RWY)
07:02:40 - NAK102: RQST L/Up and T/O
07:04:10 - NAK104: Report CROSSWIND
07:04:40 - NAK104: Report DOWNWIND
07:04:50 - NAK105: Say Hello Annonce DOWNWIND proche
07:05:20 - NAK105: Report DOWNWIND
--End

LFCL (Lasbordes) airport

The script used for scenarios MRT 2 and MART 2 at LFCL (Lasbordes) airport was:

--Briefing
111, 112, 113 and 114 are parked
--Start
06:55:30 - NAK112: Eng.Start
06:58:10 - NAK111: Eng.Start
06:58:40 - NAK111: RQST Taxi to H-Pnt
06:59:00 - NAK112: Say Hello Confirm F/P (CALL)
06:59:30 - NAK111: Report at H-Pnt and RQST Eng.Test
07:01:00 - NAK113: Eng.Start
07:02:30 - NAK112: Report on BASE and\RQST Landing

07:03:00 - NAK113: Moves W/O clearance (SPATIAL)
07:03:40 - NAK111: Report Eng.OK
07:04:30 - NAK111: RQST L/Up and T/O

--End

MRT2 and MART2

LFBR (Muret) airport

The script used for scenarios MRT 2 and MART 2 at LFBR (Muret) airport was:

--Briefing

102 is on its way to H-Pnt ; 102 and 103 are parked; 109 will be in
DOWNWIND at 2'

--Start

06:55:30 - NAK101: Eng.Start
06:55:30 - NAK104: Eng.Start
06:55:30 - NAK109: Eng.Start
06:57:55 - NAK109: Report on BASE and RQST Landing
06:58:30 - NAK109: Report on FINAL (RWY)
06:58:30 - NAK101: Cross H-Pnt W/O clearance (RWY)
06:58:55 - NAK109: Dct DOWNWIND
06:59:30 - NAK102: Eng.Start
07:00:00 - NAK101: Report Eng.Trouble RQST Taxi to apron
07:00:30 - NAK102: Moves W/O clearance (SPATIAL)
07:01:00 - NAK103: Eng.Start
07:01:10 - NAK109: Report DOWNWIND Dct BASE
07:02:00 - NAK102: RQST Taxi ti H-Pnt
07:02:45 - NAK109: Report BASE and RQST Landing
07:03:00 - NAK103: RQST Taxi to H-Pnt (CALL)
07:03:20 - NAK109: Report FINAL
07:05:30 - NAK102: RQST L/Up and T/O

--End

LFCL (Lasbordes) airport

The script used for scenarios MRT 2 and MART 2 at LFCL (Lasbordes) airport was:

--Briefing

111 will be in FINAL at 1'30''; 112 and 113 are parked

--Start

06:55:30 - NAK111: Eng.Start
06:57:30 - NAK111: Report on FINAL
06:59:00 - NAK111: Ask for stand number

07:03:00 - NAK112: RQST Eng.Start (CALL)

07:04:30 - NAK112: RQST Taxi to H-Pnt:

07:05:50 - NAK112: RQST L/Up and T/O: (CALL)

07:03:50 - NAK112: Eng.Start

--End

SASHA questionnaire (SA)

Situation Awareness

Please answer the following questions according to your experience in the scenario you have just completed.

*

	Never	Seldom	Often	More than often	Very Often	Always
I had to search for an item of information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was able to plan and organize my work as I wanted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was ahead of the traffic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was surprised by an event I did not expect	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I started to focus on a single problem or a specific area of the sector	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There was a risk of forgetting something important (like transferring an a/c on time or communicating a change to an adjacent sector)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How would you rate your overall Situation Awareness during the run? *

	1	2	3	4	5	6	7	8	9	10	
Very poor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very good

In case you rated your Situation Awareness as poor, please explain what factor(s) contributed mostly to that:

Your answer

Figure 8

Questionnaires (post-experiment guided interview)

1. Audio Focus solution

This solution is an amplification of the sound of the aircraft which are located in the head direction of the ATCO in the airfield.

1 **USEFULNESS.** The current augmented solution is a useful aid for RT operations.

Strongly Disagree

1	2	3	4	5
---	---	---	---	---

 Strongly Agree

Comments:

2 **ACCURACY.** The current augmented solution is accurate enough to support you during the RT operations.

Strongly Disagree

1	2	3	4	5
---	---	---	---	---

 Strongly Agree

Comments:

3 **SITUATION AWARENESS.** The current augmented solution improves your situation awareness in RT Operations.

Strongly Disagree

1	2	3	4	5
---	---	---	---	---

 Strongly Agree

Comments:

4 **WORKLOAD.** The current augmented solution does not have a negative impact on your workload in RT Operations.

Strongly Disagree

1	2	3	4	5
---	---	---	---	---

 Strongly Agree

Comments:

5 **SENSE OF PRESENCE.** The current augmented solution improves your sense of presence.

Strongly Disagree

1	2	3	4	5
---	---	---	---	---

 Strongly Agree

Comments:

6 Could you rate the suitability of this solution in the following operational contexts?

Current Control Tower Operations

Not at all

1	2	3	4	5
---	---	---	---	---

 Very Useful

Single Remote Tower

	<p>Not at all <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Multiple Remote Tower</p> <p>Not at all <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Low Visibility Operations</p> <p>Not at all <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Contingency Operations</p> <p>Not at all <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p>	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	2	3	4	5																	
1	2	3	4	5																	
1	2	3	4	5																	
1	2	3	4	5																	
7	Please provide here any further comments or suggestions for improvement of this solution:																				

2. Runway Incursion Alert

Alert informing the ATCO that there is an aircraft in the vicinity of the runway. The alert is provided in the form of a chair vibration coupled with a spatial sound alert coming from the direction of the aircraft.

1	<p>USEFULNESS. The current augmented solution is a useful aid for RT operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
2	<p>ACCURACY. The current augmented solution is accurate enough to support you during the RT operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
3	<p>SITUATION AWARENESS. The current augmented solution improves your situation awareness in RT Operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
4	<p>WORKLOAD. The current augmented solution does not have a negative impact on your workload in RT Operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
5	<p>SENSE OF PRESENCE. The current augmented solution improves your sense of presence.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

6	<p>Could you rate the suitability of this solution in the following operational contexts?</p> <p style="text-align: center;">Current Control Tower Operations</p> <p>Not at all <table border="1" style="display: inline-table; text-align: center;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Single Remote Tower</p> <p>Not at all <table border="1" style="display: inline-table; text-align: center;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Multiple Remote Tower</p> <p>Not at all <table border="1" style="display: inline-table; text-align: center;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Low Visibility Operations</p> <p>Not at all <table border="1" style="display: inline-table; text-align: center;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p> <p style="text-align: center;">Contingency Operations</p> <p>Not at all <table border="1" style="display: inline-table; text-align: center;"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Very Useful</p>	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
7	<p>Please provide here any further comments or suggestions for improvement of this solution:</p>																									

3. Spatial Alert solution

The aim this solution is to inform the ATCO that there is an event which is requiring his attention using an audio alert.

1	<p>USEFULNESS. The current augmented solution is a useful aid for RT operations.</p> <p>Strongly Disagree <table border="1" data-bbox="539 533 1066 566"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
2	<p>ACCURACY. The current augmented solution is accurate enough to support you during the RT operations.</p> <p>Strongly Disagree <table border="1" data-bbox="539 857 1066 891"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
3	<p>SITUATION AWARENESS. The current augmented solution improves your situation awareness in RT Operations.</p> <p>Strongly Disagree <table border="1" data-bbox="579 1171 1106 1205"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
4	<p>WORKLOAD. The current augmented solution does not have a negative impact on your workload in RT Operations.</p> <p>Strongly Disagree <table border="1" data-bbox="579 1518 1106 1552"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		
5	<p>SENSE OF PRESENCE. The current augmented solution improves your sense of presence.</p> <p>Strongly Disagree <table border="1" data-bbox="539 1832 1066 1865"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p>	1	2	3	4	5
1	2	3	4	5		

	Comments:																									
6	<p>Could you rate the suitability of this solution in the following operational contexts?</p> <p style="text-align: center;">Current Control Tower Operations</p> <p>Not at all <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table> Very Useful</p> <p style="text-align: center;">Single Remote Tower</p> <p>Not at all <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table> Very Useful</p> <p style="text-align: center;">Multiple Remote Tower</p> <p>Not at all <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table> Very Useful</p> <p style="text-align: center;">Low Visibility Operations</p> <p>Not at all <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table> Very Useful</p> <p style="text-align: center;">Contingency Operations</p> <p>Not at all <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">1</td><td style="width: 20px; text-align: center;">2</td><td style="width: 20px; text-align: center;">3</td><td style="width: 20px; text-align: center;">4</td><td style="width: 20px; text-align: center;">5</td></tr></table> Very Useful</p>	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
1	2	3	4	5																						
7	Please provide here any further comments or suggestions for improvement of this solution:																									

4. Secondary airport call solution

The function is activated when the operator receives a new communication from the secondary airport while his attention is focused on the primary one. When there is an incoming call, the system provides the operator with continuous low vibration while the ATCO receives the aircraft communication.

1	<p>USEFULNESS. The current augmented solution is a useful aid for RT operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

2	<p>ACCURACY. The current augmented solution is accurate enough to support you during the RT operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

3	<p>SITUATION AWARENESS. The current augmented solution improves your situation awareness in RT Operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

4	<p>WORKLOAD. The current augmented solution does not have a negative impact on your workload in RT Operations.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

5	<p>SENSE OF PRESENCE. The current augmented solution improves your sense of presence.</p> <p>Strongly Disagree <table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table> Strongly Agree</p> <p>Comments:</p>	1	2	3	4	5
1	2	3	4	5		

6	<p>Could you rate the suitability of this solution in the following operational contexts?</p> <p style="text-align: center;">Current Control Tower Operations</p>
---	--

	Not at all	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>	1	2	3	4	5	Very Useful
1	2	3	4	5				
		Single Remote Tower						
	Not at all	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>	1	2	3	4	5	Very Useful
1	2	3	4	5				
		Multiple Remote Tower						
	Not at all	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>	1	2	3	4	5	Very Useful
1	2	3	4	5				
		Low Visibility Operations						
	Not at all	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>	1	2	3	4	5	Very Useful
1	2	3	4	5				
		Contingency Operations						
	Not at all	<table border="1"><tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr></table>	1	2	3	4	5	Very Useful
1	2	3	4	5				
7	Please provide here any further comments or suggestions for improvement of this solution:							

Abstract

In an effort to simplify human resource management and reduce operational costs, control towers are now increasingly designed to not be implanted directly on the airport but remotely. This concept, known as *remote tower*, offers a “digital” working context: the view on the runways is broadcast remotely using cameras located on site. Furthermore, this concept could be enhanced to the control of several airports simultaneously from one remote tower facility, by only one air traffic controller (*multiple* remote tower). These concepts offer designers the possibility to develop novel interaction forms. However, the most part of the current augmentations rely on sight, which is largely used and, therefore, is sometimes becoming overloaded. In this Ph.D. work, the design and the evaluation of new interaction techniques that rely on non-visual human senses have been considered (e.g. hearing, touch and proprioception). Two experimental campaigns have been led to address specific use cases. These use cases have been identified during the design process by involving experts from the field, appearing relevant to controllers due to the criticality of the situation they define. These situations are a) poor visibility (heavy fog conditions, loss of video signal in remote context), b) unauthorized movements on ground (when pilots move their aircraft without having been previously cleared), c) runway incursion (which occurs when an aircraft crosses the holding point to enter the runway while another one is about to land), and d) how to deal with multiple calls associated to distinct radio frequencies coming from multiple airports. The first experimental campaign aimed at quantifying the contribution of a multimodal interaction technique based on spatial sound, kinaesthetic interaction and vibrotactile feedback to address the first use case of poor visibility conditions. The purpose was to enhance controllers' perception and increase overall level of safety, by providing them a novel way to locate aircraft when they are deprived of their sight. 22 controllers have been involved in a laboratory task within a simulated environment. Objective and subjective results showed significantly higher performance in poor visibility using interactive spatial sound coupled with vibrotactile feedback, which gave the participants notably higher accuracy in degraded visibility. Meanwhile, response times were significantly longer while remaining acceptably short considering the temporal aspect of the task. The goal of the second experimental campaign was to evaluate 3 other interaction modalities and feedback addressing 3 other critical situations, namely unauthorized movements on ground, runway incursion and calls from a secondary airport. We considered interactive spatial sound, tactile stimulation and body movements to design 3 different interaction techniques and feedback. 16 controllers' participated in an ecological experiment in which they were asked to control 1 or 2 airport(s) (Single Vs. Multiple operations), with augmentations activated or not. Having no neat results regarding the interaction modalities into multiple remote tower operations, behavioural results shown a significant increase in overall participants' performance when augmentation modalities were activated in single remote control tower operations. The first campaign was the initial step in the development of a novel interaction technique that uses sound as a precise means of location. These two campaigns constituted the first steps for considering non-visual multimodal augmentations into remote tower operations.

Keywords: Human-Computer Interaction, Haptics, Kinesthetics Interaction, Sonification, Remote Control Tower

Résumé

Afin de simplifier la gestion des ressources humaines et de réduire les coûts d'exploitation, certaines tours de contrôle sont désormais conçues pour ne pas être implantées directement sur l'aéroport. Ce concept, connu sous le nom de *tour de contrôle distante* (*remote tower*), offre un contexte de travail « digital » : la vue sur les pistes est diffusée via des caméras situées sur le terrain distant. Ce concept pourrait également être étendu au contrôle simultanés de plusieurs aéroports à partir d'une seule salle de contrôle, par un contrôleur seul (tour de contrôle distante *multiple*). Ces notions nouvelles offrent aux concepteurs la possibilité de développer des formes d'interaction novatrices. Cependant, la plupart des augmentations actuelles reposent sur la vue, qui est largement utilisée et, par conséquent, parfois surchargée. Nous nous sommes ainsi concentrés sur la conception et l'évaluation de nouvelles techniques d'interaction faisant appel aux sens non visuels, plus particulièrement l'ouïe, le toucher et la proprioception. Deux campagnes expérimentales ont été menées. Durant les processus de conception, nous avons identifié, avec l'aide d'experts du domaine, certaines situations pertinentes pour les contrôleurs aériens en raison de leur criticité: a) la mauvaise visibilité (brouillard épais, perte de signal vidéo), b) les mouvements non autorisés au sol (lorsque les pilotes déplacent leur appareil sans y avoir été préalablement autorisés), c) l'incursion de piste (lorsqu'un avion traverse le point d'attente afin d'entrer sur la piste alors qu'un autre, simultanément, s'apprête à atterrir) et d) le cas des communications radio simultanées provenant de plusieurs aéroports distants. La première campagne expérimentale visait à quantifier la contribution d'une technique d'interaction basée sur le son spatial, l'interaction kinesthésique et des stimuli vibrotactiles, afin de proposer une solution au cas de perte de visibilité sur le terrain contrôlé. L'objectif était d'améliorer la perception de contrôleurs et d'accroître le niveau général de sécurité, en leur offrant un moyen différent pour localiser les appareils. 22 contrôleurs ont été impliqués dans une tâche de laboratoire en environnement simulé. Des résultats objectifs et subjectifs ont montré une précision significativement plus élevée en cas de visibilité dégradée lorsque la modalité d'interaction testée était activée. Parallèlement, les temps de réponse étaient significativement plus longs relativement courts par rapport à la temporalité de la tâche. L'objectif de la seconde campagne expérimentale, quant à elle, était d'évaluer 3 autres modalités d'interaction visant à proposer des solutions à 3 autres situations critiques : les mouvements non autorisés au sol, les incursions de piste et les appels provenant d'un aéroport secondaire contrôlé. Le son spatial interactif, la stimulation tactile et les mouvements du corps ont été pris en compte pour la conception de 3 autres techniques interactives. 16 contrôleurs aériens ont participé à une expérience écologique dans laquelle ils ont contrôlé 1 ou 2 aéroport(s), avec ou sans augmentation. Les résultats comportementaux ont montré une augmentation significative de la performance globale des participants lorsque les modalités d'augmentation étaient activées pour un seul aéroport. La première campagne a été la première étape dans le développement d'une nouvelle technique d'interaction qui utilise le son interactif comme moyen de localisation lorsque la vue seule ne suffit pas. Ces deux campagnes ont constitué les premières étapes de la prise en compte des augmentations multimodales non visuelles dans les contextes des tours de contrôles déportées Simples et Multiples.

Mots-clefs: Interaction Homme-Machine, Interaction haptique, Interaction kinesthésique, Sonification, Tour de Contrôle Déportée