



HAL
open science

Manipulations gestuelles d'un objet virtuel sonifié pour le contrôle d'une interface en situation de conduite

Benjamin Bressolette

► **To cite this version:**

Benjamin Bressolette. Manipulations gestuelles d'un objet virtuel sonifié pour le contrôle d'une interface en situation de conduite. Acoustique [physics.class-ph]. Ecole Centrale de Marseille, 2018. Français. NNT: . hal-01729222

HAL Id: hal-01729222

<https://hal.science/hal-01729222>

Submitted on 25 Jan 2019

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.

École Doctorale : Sciences pour l'ingénieur : Mécanique, Physique, Micro et Nanoélectronique (ED353)

Laboratoire PRISM : Perception, Représentations, Image, Son, Musique (UMR 7061)

THÈSE DE DOCTORAT

pour obtenir le grade de

DOCTEUR de l'ÉCOLE CENTRALE de MARSEILLE

Discipline : Acoustique

**Manipulations gestuelles d'un objet virtuel sonifié
pour le contrôle d'une interface en situation de conduite**

par

BRESSOLETTE Benjamin

Directeurs de thèse : KRONLAND-MARTINET Richard, YSTAD Sølvi

Thèse réalisée en cotutelle avec Groupe PSA et encadrée par ROUSSARIE Vincent.

Soutenue le 12 novembre 2018

devant le jury composé de :

SERAFIN Stefania	Professeur - Université Aalborg	Rapporteur
PETIOT Jean-François	Professeur - École Centrale de Nantes	Rapporteur
BEAUDOUIN-LAFON Michel	Professeur - Université Paris-Sud	Examineur
SUSINI Patrick	Directeur de recherche CNRS - IRCAM	Examineur
DENJEAN Sébastien	Docteur - Groupe PSA	Représentant industriel
YSTAD Sølvi	Directeur de recherche CNRS - PRISM	Directrice de thèse
KRONLAND-MARTINET Richard	Directeur de recherche CNRS - PRISM	Directeur de thèse

Remerciements

Je tiens tout d'abord à remercier Vincent Roussarie pour m'avoir initialement accepté en stage ingénieur au sein de l'équipe Sciences Cognitives et Facteurs Humains de PSA, puis pour m'avoir proposé ce sujet de thèse fort original. Ces travaux et les études menées dans l'équipe m'ont permis d'apprendre beaucoup sur la perception multisensorielle dans ce contexte si particulier qu'est la conduite automobile. Je voudrais également particulièrement remercier le trio marseillais composé de Mitsuko, Sølvi et Richard qui ont encadré ma thèse du côté laboratoire, et qui m'ont suivi au long de ces trois années en veillant à faire germer l'intérêt pour la recherche qui me porte aujourd'hui. Je remercie Stefania Serafin, Jean-François Petiot et Patrick Susini pour avoir fait parti de mon jury, ainsi que Michel Beaudouin-Lafon pour avoir présidé ce jury d'une part, et je le remercie d'autre part pour me permettre de continuer de travailler sur les interfaces sonores dans son équipe de recherche Ex-situ du Laboratoire de Recherche en Informatique.

Je remercie chaleureusement Sébastien pour l'ensemble de son *shadow tutoring*, ses précieux conseils côtés techniques ou humains qui m'ont permis de gravir cette montagne qu'est la thèse. Les fameux « bière thèse » que l'on organisait en compagnie d'Alix étaient des moments essentiels qui m'ont permis de ne pas lâcher prise, et j'en profite également pour saluer et féliciter cette autre grimpeuse ! Du côté de l'Institut des Sciences et du Mouvement, je voudrais particulièrement remercier Christophe et Lionel pour leurs conseils qu'ils m'ont prodigué, lors des réunions Sonimove notamment.

Emmanuelle m'a également beaucoup aidé à y voir plus clair dans cet outil essentiel aux études perceptives qu'est l'analyse statistique, et je voudrais grandement la remercier pour cela et pour avoir supporté les longues réunions qui étaient nécessaires pour la préparation et le traitement des expériences menées dans cette thèse. J'ai ainsi pu me rendre compte de l'importance de ces méthodes, pour que des résultats rigoureusement obtenus puisse être utilisés comme de véritables points d'appuis lors de l'avancement de la thèse.

Je veux également rendre hommage à l'équipe SCFH pour sa bonne humeur générale et la disponibilité de ses membres, Julien avec qui je partageais un bureau, Luciano, Jean-Christophe, Rémi, Claire, Benjamin, Jérémy, Rudolf, Philippe, Gilles, Vincent, Malisa, Clément, Thomas, ainsi que les anciens de l'équipe : Morgane, Anca, Alexandre, Francine, Claire, Laure, Jonathan, Ludovic, Mitsuaki, Florent, Pierre, ou Florian, en espérant n'avoir oublié personne !

Un grand merci à Mihai, le "papa" des thésards de PSA, pour son encadrement, sa bienveillance et ses encouragements.

J'en profite pour féliciter Charlotte pour sa thèse, et pour saluer et encourager les thésards de l'équipe : Daniele, Lucas, Jocelyn, Daniele, Enguerrand, et Corentin à qui je passe le relais pour perpétuer les thèses PSA-PRISM!

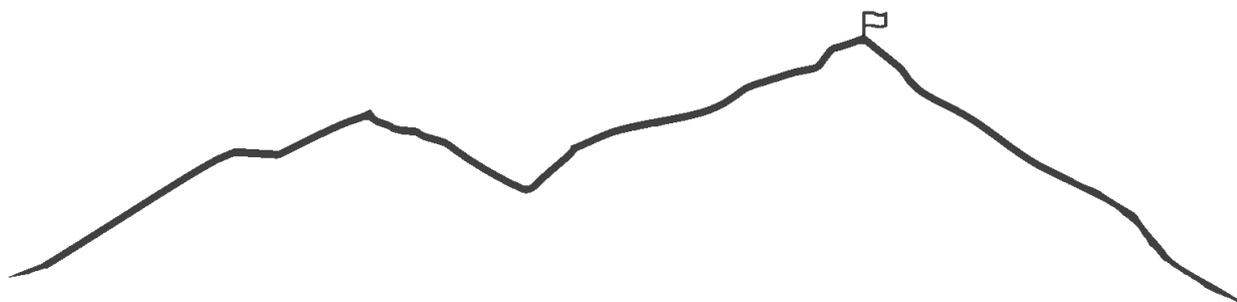
Je voudrais également remercier l'ensemble des sujets qui ont participé à mes expériences, sans qui rien n'aurait été possible!

Du côté laboratoire, je remercie Gaëtan et Laurent pour leur aide et leur bonne humeur, je salue également l'ensemble de l'équipe et les thésards dont j'ai croisé l'ascension : Lennie, Jocelyn, Thomas et Antoine, je souhaite bon courage à ceux qui n'ont pas terminé et aux nouveaux arrivants!

Un clin d'œil à l'ensemble de mes amis qui m'ont soutenu de près ou de loin pendant ces trois années, notamment à Gif et Saille pour leurs conseils dans le moment crucial avant la soutenance, à Rémy pour la relecture de mon abstract, au groupe des « depressed doctors » pour ce soutien mutuel que nous nous sommes donnés, et aux différents hôtes qui m'ont hébergés à Marseille!

Je voudrais grandement remercier mes parents pour l'ensemble de leur modeste œuvre, pour m'avoir soutenu dans ces années de thèse qui peuvent être difficiles. Je les remercie, ainsi que mon frère Pierre, d'être venu assister à ce moment assez spécial qu'est la soutenance, et d'avoir fait découvrir la culture auvergnate aux membres du jury et au public par la même occasion. Mention spéciale à mon père qui a lu et corrigé l'ensemble du manuscrit!

Enfin, je voudrais rendre un vibrant hommage à Gwenaëlle ma compagne, pour son soutien sans faille qui m'a permis de voir et d'atteindre le sommet de ce travail de thèse. Ses nombreuses corrections m'ont été d'une grande aide, et je suis reconnaissant qu'elle m'ait écouté lors de mes multiples présentations. Je la remercie de m'avoir supporté tout ce temps, de m'avoir encouragé lors des moments difficiles. Sa passion pour la recherche scientifique m'a également grandement inspiré, et je lui souhaite une grande réussite dans ses travaux qu'elle mérite sans aucun doute.



Introduction générale	1
I Contexte de la thèse et état de l'art	13
1 Le dialogue Homme – machine	15
1.1 Les gestes comme médiateurs d'interactions	17
1.1.1 Interfaces : cas général	18
1.1.2 Les gestes sans contact en contexte automobile	22
1.2 Le son comme retour d'information	24
1.2.1 Icônes sonores	24
1.2.2 Sonification des données	27
1.2.3 Sonification appliquée à l'automobile	28
2 Perception et association de modalités	31
2.1 Systèmes perceptifs gestuel et sonore	33
2.1.1 Système sensorimoteur	33
2.1.2 Système auditif humain et localisation des sons	34
2.2 Cas général de l'intégration et d'interactions multisensorielles	38
2.2.1 Une forte influence de la modalité visuelle	39
2.2.2 Modèles de jugements perceptifs	40
2.2.3 Effets attentionnels sur la perception	41
2.3 Approche écologique de la perception auditive	44
2.3.1 Écoutes quotidienne et musicale	44
2.3.2 L'événement comme origine	45
2.3.3 Les invariants perceptifs comme base de l'écoute écologique	45
2.3.4 Synthèse de sons quotidiens	48
2.4 Relations entre mouvements et sons	51
2.4.1 Évocation du mouvement par les sons	51
2.4.2 Interactions sonores et gestuelles	53

3	Aspects cognitifs et distractions en conduite	59
3.1	La construction de schémas cognitifs	61
3.1.1	Mémoire et charge cognitive	61
3.1.2	Mécanismes d'apprentissage	62
3.1.3	Techniques de mesure de la charge cognitive	65
3.2	Les phénomènes distractifs en conduite	68
3.2.1	Le simulateur de conduite : avantages et limites	69
3.2.2	Distractions motrices	72
3.2.3	Distractions visuelles	72
3.2.4	Distractions cognitives	73
3.2.5	Comparaison et conséquences sur la conduite	74
II	Contribution expérimentale	77
4	Construction d'un objet virtuel sonore	79
4.1	Système mis en place	81
4.1.1	Choix du capteur et logiciels	81
4.1.2	Construction de l'objet virtuel	82
4.2	Expérience 1 : liens entre paramètres de l'objet virtuel et paramètres sonores	86
4.2.1	Hypothèses initiales et motivations	86
4.2.2	Méthodologie	86
4.2.3	Résultats	89
4.2.4	Discussion	90
4.2.5	Conclusion	92
4.3	Expérience 2 : hiérarchisation des stratégies sonores	92
4.3.1	Hypothèses initiales et motivations	93
4.3.2	Méthodologie	94
4.3.3	Résultats	98
4.3.4	Discussion	101
4.3.5	Conclusion	104
5	Prise en main de l'objet virtuel	107
5.1	Premières briques de l'interface gestuelle	109
5.1.1	Choix de cas d'usages	109
5.1.2	Design des sons	111
5.2	Expérience 3 : évocations sonores	117
5.2.1	Évolution de l'objet virtuel	117
5.2.2	Hypothèses initiales et motivations	117
5.2.3	Méthodologie	117
5.2.4	Résultats	120
5.2.5	Discussion	126
5.2.6	Conclusion	127
5.3	Expérience 4 : prise en main des stratégies sonores	129
5.3.1	Hypothèses initiales et motivations	130
5.3.2	Méthodologie	130
5.3.3	Résultats	132

5.3.4	Discussion	143
5.3.5	Conclusion	148
6	Utilisation de l'interface gestuelle en situation de conduite	151
6.1	Développement d'une interface gestuelle complète	153
6.1.1	Fonctions de ventilation	153
6.1.2	Fonctions de gestion des médias	154
6.1.3	Choix des gestes et manipulations de l'objet virtuel associées	155
6.1.4	Stimuli audio-visuels et construction des interfaces	158
6.2	Intégration du système et simulateur de conduite	162
6.2.1	Simulateur de conduite	162
6.2.2	Verrouillage de l'interface	164
6.3	Expérience 5 : comparaison d'interfaces en situation de conduite	165
6.3.1	Hypothèses initiales et motivations	165
6.3.2	Méthodologie	166
6.3.3	Résultats	176
6.3.4	Discussion	190
6.3.5	Conclusion	195
	Conclusion et perspectives	199
	Bibliographie	I
	Annexes	XXVII

1	Une boucle d'interaction	2
2	Boucle d'interaction entre l'utilisateur et l'IHM	5
3	Comparaison des situations réelles avec notre situation	5
4	Schéma du dispositif liant gestes, sons et fonctions de l'interface	7
5	Schéma récapitulatif de la problématique de la thèse	9
1.1	Vue intérieure d'une Peugeot 308 de première génération	18
1.2	L'écran central dans une Peugeot 308 de seconde génération	19
1.3	Lev Termen jouant du Thérémine.	20
1.4	Exemple de Thérémine actuel vendu dans le commerce.	20
1.5	Le capteur Leapmotion	21
1.6	Exemple de micro-geste de zoom au volant tiré de [Mahr <i>et al.</i> , 2011]	22
1.7	Exemple de deux <i>earcons</i> tirés de [Blattner <i>et al.</i> , 1989].	26
1.8	Le principe de la boucle action - perception pour la sonification interactive	28
1.9	Vue arrière donnée sur l'écran central d'une Peugeot 308	29
2.1	Représentation des sensations provenant des sources sensorimotrices	33
2.2	Représentation schématique de l'oreille	34
2.3	Schéma illustrant les différences interaurales de niveau et de temps	35
2.4	Le cône de confusion	36
2.5	Grandes étapes du traitement des informations perceptives	41
2.6	Stimuli visuels et auditifs de l'étude de Duncan [Duncan <i>et al.</i> , 1997]	43
2.7	Liens entre le barycentre d'intensité spécifique et la taille ou la vitesse de billes roulantes	47
2.8	Illustration de l'analyse – synthèse	48
2.9	Interface proposée du synthétiseur de sons naturels	49
2.10	Schéma listant différentes informations évoquées par le son	50
2.11	Exemple de l'interface permettant des dessins de trajectoires	52
2.12	Sujet au cours d'une expérience de Rosati <i>et al.</i>	54
2.13	Stimuli et résultats de l'expérience menée par Thoret <i>et al.</i>	55
2.14	Le dispositif du « Ballancer »	56
3.1	Résumé des différentes catégories de charges cognitives	64
3.2	Mécanismes intervenant lors de la présentation d'une tâche	65

3.3	Performances en fonction de la demande	66
4.1	Schéma des connexions des différentes briques du système mis en place	82
4.2	Schéma résumant les stimuli disponibles lors de l'expérience 1	87
4.3	Vue du dessus du dispositif donnée par l'interface de l'outil de spatialisation	87
4.4	Les 6 différents mappings sélectionnés	88
4.5	Barycentre d'intensité spécifique calculé pour la stratégie sonore 6	91
4.6	Illustration des trois zones à atteindre	94
4.7	Interface visuelle mise à la disposition des sujets lors de l'expérience 2	95
4.8	Construction de l'indice	97
4.9	Performances par stratégie sonore pour les deux phases de l'expérience 2	98
4.10	Indice par stratégie sonore pour les deux phases de l'expérience 1	99
4.11	Indice détaillé par raideur et zone en phase 2	100
4.12	Évolution de moyennes glissantes sur quatre indices de quatre stratégies sonores	103
5.1	Comparaison de sons pour la stratégie liant la vitesse de l'OV à la brillance	113
5.2	Module du synthétiseur de sons convoluant une action à une texture sonore	113
5.3	Comparaison de sons pour la stratégie liant la distance main-OV à la hauteur	115
5.4	Code de l'analyse thématique	120
5.5	Résultats du questionnaire concernant le caractère agréable des stratégies sonores	125
5.6	Liens entre zone atteinte par l'objet virtuel et changement de température	129
5.7	Nouvelle construction de l'indice à partir des positions finales de l'objet virtuel	132
5.8	ACP pour la stratégie sonore liant la distance main-OV à la hauteur	133
5.9	ACP pour la stratégie sonore liant la vitesse de l'OV au paramètre de roulement	134
5.10	Évolution moyenne de l'indice selon les groupes de lancers	136
5.11	ACP sur la phase 1 de l'expérience pour la stratégie de sonification de la distance	137
5.12	ACP sur la phase 1 de l'expérience pour la stratégie de sonification de la vitesse	138
5.13	Évolution moyenne de l'indice selon les zones pour la seconde partie de l'expérience	140
5.14	ACP sur la phase 2 de l'expérience pour la stratégie de sonification de la distance	141
5.15	ACP sur la phase 2 de l'expérience pour la stratégie de sonification de la vitesse	142
5.16	Évolution moyenne de l'indice selon les zones et la phase de l'expérience	143
5.17	Évolution de l'indice selon les stratégies sonores et la phase de l'expérience	147
6.1	Schéma récapitulatif des liens entre l'OV, les gestes et les fonctions	157
6.2	Le simulateur de conduite que nous avons utilisé pour l'expérience 5	162
6.3	Vue alternative du simulateur de conduite utilisé en expérience 5.	163
6.4	Positions relatives des deux écrans dans le simulateur de conduite.	163
6.5	Schéma résumant les liens OV - gestes - fonctions - sons	168
6.6	Vue d'ensemble du circuit proposé	174
6.7	Résultats des manipulations des interfaces sans retour d'information	177
6.8	Résultats des manipulations des interfaces avec retours d'informations	179
6.9	Résultats des manipulations des interfaces tactile et gestuelle en conduite	180
6.10	Résultats des manipulations des interfaces tactile et gestuelle en conduite	181
6.11	Distance au-delà de la limite de 40 mètres par type d'interface	183
6.11	ACP réalisée sur les scores de RSME	184
6.12	Mesures de la charge cognitive à différents moments de l'expérience	185
6.13	Évaluation de la maîtrise de l'interface à différents moments de l'expérience	186

6.14	Répartition des réponses sur la facilité d'utilisation des interfaces à l'arrêt	186
6.15	Performances d'utilisation de l'IHM gestuelle sur plusieurs séances de conduite	190
A.1	Représentation graphique des scores obtenus dans l'expérience 1	XXVII
A.2	Scores obtenus pour la stratégie sonore 1	XXVIII
A.3	Scores obtenus pour la stratégie sonore 2	XXVIII
A.4	Scores obtenus pour la stratégie sonore 3	XXVIII
A.5	Scores obtenus pour la stratégie sonore 4	XXVIII
A.6	Scores obtenus pour la stratégie sonore 5	XXIX
A.7	Scores obtenus pour la stratégie sonore 6	XXIX
B.1	Données d'entrée pour la comparaison des sons	XXXI
B.2	Spectrogramme de la stratégie sonore 6 utilisée en expérience 2	XXXII
B.3	Spectrogramme du son de roulement pur	XXXII
B.4	Spectrogramme de la texture sonore choisie pour la stratégie sonore 6	XXXII
B.5	Spectrogramme du son de roulement convolué avec la texture sonore	XXXII
B.6	Spectrogramme de la texture sonore choisie pour la stratégie sonore 1	XXXIII
B.7	Spectrogramme de la stratégie sonore 1 utilisée en expérience 2	XXXIII
B.8	Spectrogramme de la texture sonore proposée avec un décalage fréquentiel	XXXIII
C.1	Vue de l'interface gestuelle lors de l'étape 1	XXXV
C.2	Vue de l'interface gestuelle lors de l'étape 4	XXXVI
C.3	Vue de l'interface gestuelle lors de l'étape 5	XXXVI
D.1	Vue du simulateur pour une distance avec le lièvre est inférieure à 40 m.	XXXIX
D.2	Vue du simulateur pour une distance avec le lièvre est supérieure à 40 m et inférieure à 60 m.	XL
D.3	Vue du simulateur pour une distance avec le lièvre est supérieure à 60 m.	XL
E.1	Échelle RSME utilisée pour l'expérience 5.	XLII
F.1	ACP réalisée sur les scores de confiance	XLIV
F.2	Répartition des réponses sur la facilité d'utilisation des interfaces à l'arrêt	XLV
F.3	Répartition des réponses portant sur la facilité d'utilisation des interfaces en conduite	XLV
F.4	Répartition des réponses portant sur le caractère agréable des interfaces à l'arrêt . .	XLV
F.5	Répartition des réponses portant sur le caractère agréable des interfaces en conduite	XLV
F.6	Répartition des réponses portant sur le caractère sécuritaire des interfaces en conduite	XLVI
G.1	Répartition des réponses portant sur la pertinence du geste pour change de morceau	XLVII
G.2	Répartition des réponses portant sur la pertinence du geste pour avancer ou reculer dans un morceau	XLVII
G.3	Répartition des réponses portant sur la pertinence du geste pour le positionnement de la source de musique	XLVIII
G.4	Répartition des réponses portant sur la pertinence du geste pour la manipulation du volume sonore	XLVIII
H.1	Répartition des réponses portant sur la capacité à imaginer le comportement de l'OV	XLIX
H.2	Répartition des réponses portant sur le nombre de gestes à retenir	XLIX
H.3	Répartition des réponses portant sur la combinaison d'interactions gestuelles et tactiles	L

H.4 Répartition des réponses portant sur l'intérêt des gestes pour manier la ventilation .	L
H.5 Répartition des réponses portant sur l'intérêt des gestes pour manier les médias . . .	L

Liste des tableaux

2.1	Résultats obtenus lors de l'expérience menée dans [Rath et Schleicher, 2008]	56
4.1	Les six stratégies de sonification mises au point	84
4.2	Scores obtenus à l'issue de l'expérience	90
4.3	Les treize stratégies de sonification mises au point	95
4.4	Rappel des treize stratégies de sonification	99
6.1	Résumé des étapes de la phase de découverte de l'OV	171
6.2	Moyennes de paramètres de conduite	182
6.3	Tableau récapitulatif de comparaison des interfaces tactile et gestuelle	187
6.4	Tableau récapitulatif des réponses comparant les fonctions de l'interface de ventilation	188
6.5	Tableau récapitulatif des réponses concernant l'interface des médias	188
6.6	Tableau récapitulatif des réponses concernant l'interface gestuelle	189

Introduction générale

L'interaction est un pilier fondamental dans la construction de la vie. Définie comme la « réaction réciproque de deux phénomènes l'un sur l'autre »¹, l'interaction se trouve dans la dénomination des quatre forces principales régissant les lois physiques que nous connaissons, qui impliquent toutes deux corps agissant mutuellement : l'interaction faible, l'interaction forte, l'interaction électromagnétique et l'interaction gravitationnelle. Ce principe d'échange entre deux corps semble intrinsèquement lié à la physique et à la vie.

Les êtres humains tiennent compte en permanence d'informations provenant de leur entourage pour évoluer dans leur environnement. Ces données extérieures leur permettent de prendre des décisions adéquates à leur situation pour se déplacer, se nourrir ou encore se protéger. En particulier, les humains interagissent socialement entre eux de manière privilégiée par le langage, parfois articulé, pour échanger des informations, des idées ou des points de vue. Après sa création, ce moyen de communication a par exemple permis de chasser de manière coordonnée, conférant un avantage par rapport à une chasse individuelle. L'utilisation du langage semble liée au développement de relations sociales entre les êtres humains, leur permettant de coopérer pour atteindre un but précis [Tylén *et al.*, 2010]. Ces interactions ont pu permettre aux individus de se dépasser au travers d'un collectif. Les animaux utilisent également un « langage » pour interagir entre eux et organiser leurs sociétés basées sur la coopération, ce qui est aussi le cas des végétaux. Les abeilles ou les fourmis emploient par exemple des phéromones pour échanger des informations et organiser leur vie sociale. Dans nos sociétés modernes, l'ensemble des moyens de communication a été développé pour permettre des interactions plus riches et plus nombreuses, entre personnes éloignées par l'intermédiaire du téléphone, d'appels vidéos ou de réseaux sociaux.

L'interaction entre deux personnes implique une action d'un individu 1, assurée par exemple par la vibration des cordes vocales. L'individu 2 doit pouvoir percevoir l'action de l'individu 1, les signaux sonores dans le cas de la parole. Il peut ensuite formuler une réaction ou rétroaction, qu'il peut transmettre également par la parole. La figure 1 présente cet échange d'informations du point de vue de l'individu 1, les flèches étant à inverser si l'on se place du point de vue de l'individu 2. Dans l'exemple que nous avons donné, la conversation peut permettre d'atteindre collectivement un objectif qui n'aurait pas pu l'être individuellement, comme le fait de déplacer un objet lourd.

1. <https://www.larousse.fr/dictionnaires/francais/interaction/43595>

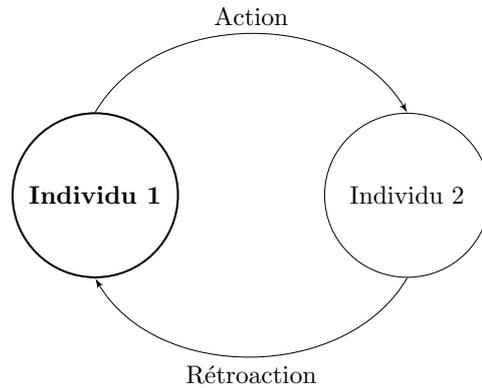


FIGURE 1 – Une boucle d'interaction du point de vue de l'individu 1.

Il est important de noter que l'interaction de l'exemple précédent n'aurait pas pu avoir lieu si les deux individus n'étaient pas dotés d'un système de captation. Les êtres humains assurent une relation avec leur environnement par l'intermédiaire de leurs sens, véritables canaux d'informations permettant de percevoir des messages de différentes natures. Comme nous le verrons dans la suite de ce document, le cerveau humain interprète les données provenant de plusieurs sens simultanément dans le cadre d'une perception dite *multisensorielle*. Ces croisements d'informations permettent d'avoir une perception robuste d'une situation et d'y répondre de manière idoine.

Les humains reçoivent en permanence des signaux de leur environnement au travers de leurs différents sens. La personne concernée peut donc, de manière consciente ou non, allouer des ressources attentionnelles à certains signaux pour en prioriser le traitement. Nous verrons que la perception d'un individu est influencée activement par l'attention [Mack et Rock, 1999]. L'interaction des êtres humains est donc dépendante du système de perception des signaux, mais également de l'attention.

La société humaine s'est donc organisée, à l'image des fourmis ou des abeilles, par des interactions constantes entre individus, relayées par leurs sens. À partir de la révolution industrielle, les machines ont permis aux hommes de faire des progrès techniques et technologiques considérables, en augmentant les capacités de production, de déplacement ou de construction. Il a alors été nécessaire de s'interroger sur la relation entre les humains et les machines. En particulier, un ouvrier doit pouvoir utiliser une machine, en ayant à sa disposition les moyens d'action nécessaires – une porte où mettre le charbon pour une locomotive par exemple. Il est également essentiel qu'il connaisse son état – pour savoir si la locomotive a besoin de charbon ou non. La boucle présentée en figure 1 peut donc se retrouver dans l'exemple de la locomotive, sauf que l'individu 2 est une machine.

Avec l'évolution des technologies et notamment l'arrivée de l'électronique, il a été possible de mettre au point des machines de plus en plus diverses, qui permettent aux usagers des actions sophistiquées, et de donner des retours d'informations variés. Les machines sont actuellement utilisées dans de nombreux secteurs, le micro-ordinateur étant la machine la plus emblématique et probablement la plus répandue.

L'ergonomie des machines a changé selon les usages, pour adapter au mieux les actions et retours d'informations aux utilisateurs. Ces adaptations ont en particulier fortement sollicité la modalité visuelle, qui est particulièrement importante dans l'intégration multisensorielle comme nous le verrons dans ce document. Ainsi, de nombreux écrans sont intégrés aux machines pour servir d'interface entre la machine et les êtres humains, dans le but de donner une multitude d'informations sur l'état présent de la machine ou les actions possibles des utilisateurs.

Contexte automobile

Les automobiles sont des exemples de machines permettant une extension des capacités de déplacement des humains. Elles proposent des moyens d'action aux conducteurs par un volant pour maîtriser la direction, et par des pédales pour la gestion de l'accélération ou décélération longitudinale du véhicule par exemple. De cette manière, le conducteur interagit avec son véhicule pour le maintenir sur la route, et éviter des accidents avec les autres véhicules. D'un point de vue cognitif, la tâche de conduite demande une forte concentration des conducteurs. Ils doivent en effet assurer une coordination de mouvements différents de leurs bras et de leurs jambes, qui peuvent parfois être simultanés. Il est donc nécessaire pour le conducteur d'anticiper les actions qu'il va devoir mener, pour rester maître du véhicule malgré la forte demande attentionnelle liée au pilotage de son véhicule. Pour assurer une conduite sécuritaire, le conducteur doit recueillir des informations provenant des autres usagers de la route, et en tenir compte dans son processus de décision. Conjointement, il a été nécessaire de proposer aux conducteurs des systèmes secondaires à la conduite, actionnant les essuie-glaces, les phares ou la ventilation par exemple. Ils permettent d'adapter le confort de conduite à des conditions particulières. Ces fonctions sont modifiables par des boutons, qu'il est nécessaire de pousser ou de tourner, parfois disposés autour du volant pour faciliter leur utilisation simultanée à la conduite.

L'utilisation conjointe de ces fonctions à la tâche de conduite peut mener à des phénomènes distractifs. Une étude menée par Dingus *et al.* se basant sur des enregistrements vidéos a montré que 78% des accidents en automobiles sont liés à un phénomène d'inattention du conducteur [Dingus *et al.*, 2006]. Ces phénomènes distractifs peuvent provenir d'un manque de sommeil du conducteur, de la manipulation d'un téléphone ou encore de l'utilisation de l'interface du véhicule. Certaines distractions peuvent altérer le temps de réaction du conducteur, en augmentant d'environ 0,5 seconde le temps de freinage par rapport à une situation de conduite normale [Lamble *et al.*, 1999]. La mise en danger est donc bien réelle lorsque le conducteur est distrait.

Cependant le maniement de boutons peut s'effectuer sans une distraction trop importante, grâce à la répartition des informations à traiter selon plusieurs sens. La modalité visuelle est concentrée sur la tâche de conduite, pour scruter les changements de direction de la route et les dangers potentiels à éviter. La modalité auditive peut aider à connaître le régime moteur, pour éventuellement passer un rapport. Les pieds et les mains servent à actionner les pédales et le volant du véhicule. Si le conducteur connaît la position approximative du bouton à actionner, il peut se passer de la modalité visuelle : il peut tendre une main dans la direction concernée, reconnaître le bouton à l'aide de caractéristiques géométriques – forme ou manière d'actionner, et le manipuler de la manière désirée. Les retours d'informations tactiles des boutons perçus par le système sensorimoteur par l'intermédiaire de crans par exemple, peuvent ici prendre la place de retours visuels. En pratique, un regard bref du conducteur est souvent constaté avant et/ou après une manipulation.

Évolution et interfaces tactiles

Des évolutions notables ont récemment eu lieu dans les habitacles des véhicules, avec l'arrivée de nouvelles fonctions comme les aides à la conduite, un système de GPS ou la possibilité de téléphoner. Il a donc été nécessaire d'adapter les interfaces, pour permettre aux utilisateurs d'actionner l'ensemble des fonctions. Dans un premier temps, plusieurs boutons permettant d'actionner ces fonctions ont fait leur apparition, notamment par l'utilisation d'un système de menu affiché sur un écran, placé au centre de l'habitacle. Ce système a permis de centraliser l'ensemble des fonctions. Cependant, l'ergonomie des habitacles a été bouleversée par l'arrivée de la technologie tactile, qui a permis de supprimer une grande quantité de boutons physiques. Ils ont été remplacés par des boutons virtuels sur l'interface, actionnables par des contacts de la main avec l'écran. Ces interfaces permettent d'alléger visuellement les habitacles, où un grand nombre de boutons étaient disposés, et d'en améliorer l'ergonomie.

Surcharge du canal sensoriel visuel

Cette technologie tactile pose cependant problème quand une fonction de confort est manipulée simultanément à la conduite. Les conducteurs ne peuvent plus garder les yeux sur la route, les retours d'informations tactiles ayant disparus : ils doivent donc regarder l'interface pour repérer le bouton virtuel qu'ils veulent actionner, puis diriger leur main vers la zone de l'écran avec laquelle entrer en contact. Pour certaines fonctions, un retour sonore bref permet au conducteur de s'assurer de la modification de la fonction voulue, mais d'autres nécessitent un contrôle visuel. Ces sollicitations additionnelles à la vigilance visuelle nécessaire à une conduite sécuritaire ont un effet distrayant, et peuvent mener à une concentration importante de la part des conducteurs. Si cette charge cognitive est trop importante, il peut devenir impossible d'assurer une conduite sécuritaire pendant la manipulation de l'interface. La manipulation de ces interfaces détourne le regard des conducteurs de la route, ce qui réduit leur capacité d'anticipation et peut donc mener à des situations dangereuses [Dingus *et al.*, 2006] [Lamble *et al.*, 1999].

Approche multisensorielle

Pour faire face à ce problème, nous avons voulu supprimer la nécessité d'utiliser la modalité visuelle pour le maniement de l'interface, en proposant une répartition des informations entre d'autres modalités. De cette manière, il est possible de garantir une faible concentration du conducteur, tout en lui permettant de focaliser son regard sur la tâche de conduite : la conduite peut être plus sécuritaire.

Il est alors nécessaire de choisir une façon de donner des instructions au système central du véhicule – l'action de la figure 1, et un ou plusieurs canaux sensoriels pour permettre le retour d'information. La parole semble être un candidat naturel, grâce à son utilisation pour la communication d'humain à humain. Les technologies progressent dans ce domaine, et des systèmes utilisant la voix sont de plus en plus proposés dans les véhicules. Le Groupe PSA souhaite cependant offrir un panel complet d'interactions possibles aux futurs conducteurs, offrant des technologies tactiles, vocales, ainsi que d'autres technologies à imaginer.

Dans le cadre de cette thèse, nous avons choisi de nous intéresser spécifiquement à des gestes sans contact, qu'il est possible de contrôler sans visuel grâce à la proprioception. Une boucle de rétroaction est alors nécessaire, pour assurer un retour d'information sur les gestes effectués, mais aussi sur les manipulations de l'interface qui ont été prises en compte. Nous avons alors proposé d'utiliser la modalité auditive, qui est relativement peu sollicitée en véhicule. La sonification, définie comme l'emploi de sons non verbaux pour transmettre une information [Kramer *et al.*, 1999], peut permettre de renseigner efficacement les conducteurs.

Les gestes et les sons semblent par ailleurs intrinsèquement liés : des gestes dits *coverbaux* accompagnent parfois la parole, le chef d'orchestre dirige les musiciens par des gestes, et les musiciens eux-mêmes font des gestes pour engendrer des sons grâce à leurs instruments de musique. Cette relation paraît ancienne : les premiers instruments de musique semblent provenir de gestes de frappe d'objets pour exprimer des émotions [Sachs, 2012].

La figure 2 montre que la boucle d'interaction est toujours présente entre l'utilisateur et l'Interface Homme-Machine – ou *IHM* : son action est exprimée par des gestes, et la réaction de l'interface à ces gestes est donnée sous la forme de sons. Il est cependant nécessaire de s'interroger sur le mécanisme permettant de lier les gestes aux sons, ainsi qu'au lien à établir entre gestes et fonctions de l'interface.

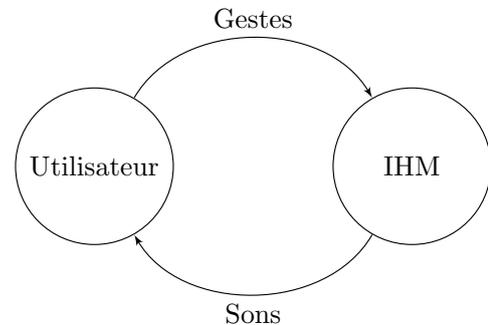


FIGURE 2 – Boucle d'interaction entre l'utilisateur et l'IHM.

Hypothèse d'étude

Pour lier les gestes et les sons, nous nous sommes appuyés sur les travaux de William Gaver, qui avancent que la source d'un son est le résultat d'une interaction entre matériaux [Gaver, 1993a] [Gaver, 1993b]. Ce principe s'applique aux sons du quotidien, lorsqu'un invité frappe à une porte par exemple. La main vient alors interagir avec un objet pour créer un son, comme le montre la figure de gauche de la figure 3. En appliquant cette approche à notre association de gestes et de sons, l'action est accomplie par le geste de l'utilisateur, mais il n'existe aucun objet qui pourrait interagir avec la main, et être la source de l'événement sonore, voir la figure de droite de la figure 3.



FIGURE 3 – Comparaison des situations réelles de production de sons de l'environnement avec notre situation.

Nous faisons alors le choix d'utiliser un objet virtuel – que nous noterons *OV* – qui permet de faire le lien entre les gestes de l'utilisateur, et les retours d'informations sonores. Nous supposons que cet objet peut permettre, en faisant la métaphore de manipulations de véritables objets, d'agréger des gestes de manière cohérente. On peut par exemple imaginer pousser l'objet dans une direction en faisant un balayage de la main, comme peut l'être un objet réel. Nous allons donc devoir déterminer comment lier les gestes de l'utilisateur à l'objet virtuel, dans le but de créer des manipulations faciles à retenir.

Une seconde question posée par l'utilisation d'un objet virtuel est de définir les informations portées par les sons générés. Ces retours sonores doivent par exemple permettre à l'utilisateur de savoir si la fonction qu'il désire modifier l'a bien été, et dans les proportions voulues. Nous pouvons alors imaginer lier l'objet virtuel à des stratégies de sonification donnant des informations continues ou brèves, selon les besoins.

Il est enfin nécessaire que l'OV soit relié à une interface, dans le but de modifier des fonctions de l'IHM. Cette utilisation d'un *médiaireur* entre un *utilisateur* et un *objet d'intérêt* est notamment abordé par Beaudouin-Lafon [Beaudouin-Lafon, 2000]. Nous pensons que des gestes entraînant différentes manipulations de l'objet virtuel peuvent être reliés de manière pertinente à plusieurs fonctions de l'interface. Ainsi, chaque geste est associé de manière unique à une manipulation de l'OV, elle-même liée à une fonction précise de l'interface. Il peut donc être possible de créer une manière simple d'interagir avec l'objet virtuel par des gestes, en reprenant des métaphores de manipulations de véritables objets, et de modifier les fonctions de l'interface en conséquence. La figure 4 résume cette proposition, en prenant l'exemple de trois manipulations modifiant trois fonctions différentes. Un état initial est représenté au centre, à partir duquel trois gestes peuvent être effectués pour modifier les fonctions de l'interface.

Nous pouvons constater la conservation de la boucle d'interaction entre l'utilisateur et l'interface, l'objet virtuel n'étant qu'un moyen de donner un sens au lien entre les gestes et les sons. De plus, s'il existe une cohérence entre les gestes, les manipulations de l'OV et les fonctions manipulées, nous faisons l'hypothèse que l'utilisation de l'interface sera facilitée.

En effet, il est important de garder à l'esprit le contexte de cette étude : les manipulations de l'interface doivent pouvoir être accomplies simultanément à une tâche de conduite. Il faut donc qu'elles ne nécessitent pas une attention du conducteur trop importante, pour que la tâche de conduite puisse être accomplie en toute sécurité. De plus, nous ne désirons pas solliciter le visuel du conducteur : il faut donc que les manipulations de l'objet virtuel soient réalisables en aveugle. Pour que les conducteurs perçoivent l'OV et sa dynamique même sans informations visuelles, nous proposons d'illustrer les déplacements de l'objet virtuel par une stratégie de sonification. Cette stratégie sonore doit permettre de transmettre par le son des informations qui permettent de manipuler un objet, habituellement perçues par le biais des modalités visuelles et tactiles.

L'introduction d'un objet virtuel est donc l'occasion de proposer un nouveau type d'interaction entre l'utilisateur et la machine, en se basant sur des manipulations connues, ce qui peut amener à un nouveau type d'interface. Une autre possibilité aurait été de sonifier directement des paramètres de la main, la vitesse ou la position par exemple, et nous pourrions discuter du caractère intuitif des deux types de sonification au cours d'une expérience présentée dans ce document. Nous faisons cependant l'hypothèse que ce type de sonification conduirait à des manipulations moins naturelles de l'interface, et qui pourraient être plus difficiles à mémoriser. Nous désirons plutôt nous servir de mécanismes de manipulation d'objets déjà présents à l'esprit des utilisateurs, qui peuvent ensuite être adaptés pour l'objet virtuel, ce qui peut rendre le maniement de l'IHM intuitif.

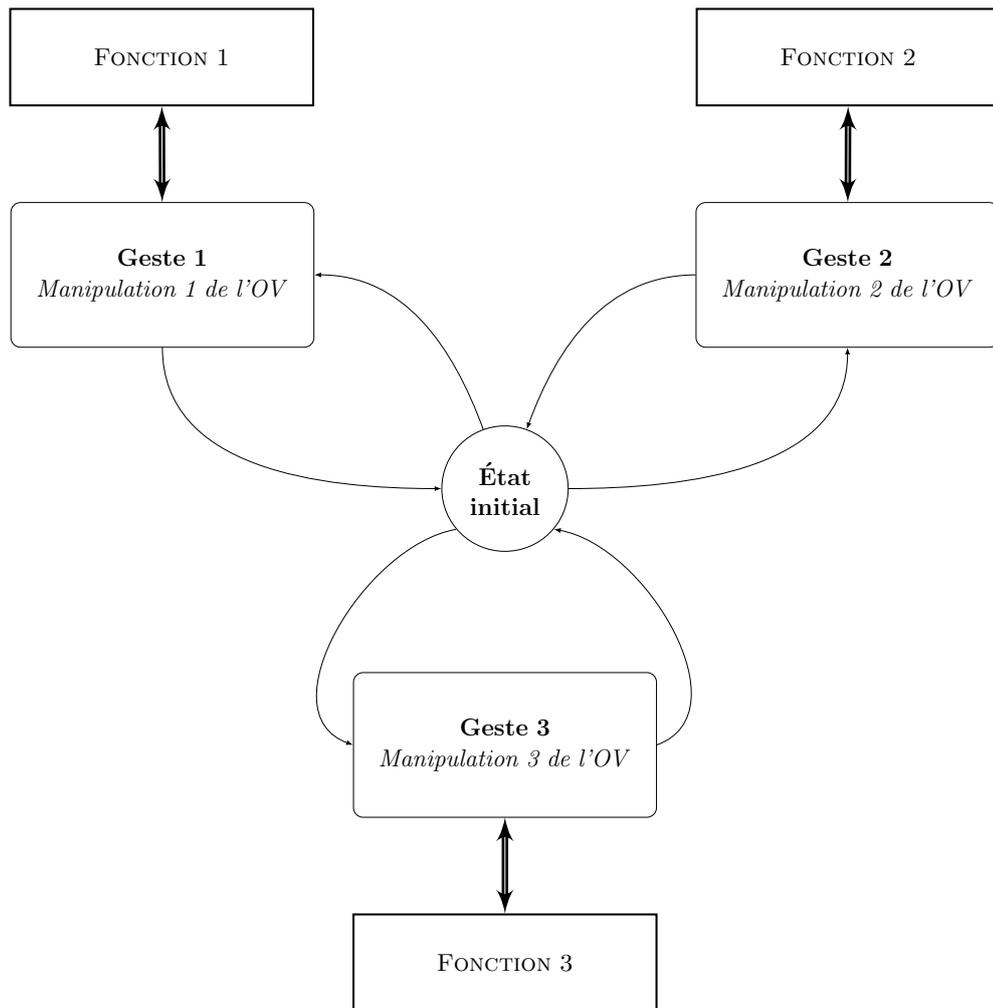


FIGURE 4 – Schéma du dispositif liant les gestes de l'utilisateur, les sons et les fonctions de l'interface.

Problématique

Ce concept d'objet virtuel semble intéressant du point de vue de la répartition des informations selon différentes modalités, et il permet de faire un lien pouvant paraître naturel entre les gestes et les sons. Il soulève néanmoins de nombreuses questions, et nous tâcherons de répondre à certaines d'entre elles au cours de ce document.

Stratégie de sonification

Nous pouvons tout d'abord nous interroger sur la stratégie de sonification permettant de représenter au mieux la dynamique de l'objet virtuel. L'OV doit en effet être entièrement manipulé en aveugle dans l'application que nous prévoyons. Cette stratégie de sonification doit être entièrement mise au point, même si nous verrons que diverses stratégies de sonification ont été proposées dans la littérature. Nous pouvons alors nous poser plusieurs questions.

En particulier, quel type de sonification est-il intéressant de mettre au point, continue ou discrète ? Sur quels sons se baser ? Quels sont les paramètres de l'objet virtuel à sonifier ? Par quels paramètres sonores transmettre ces informations dynamiques ? Comment créer un lien entre ces paramètres qui fasse sens pour l'utilisateur ? La stratégie sonore doit également proposer des sons suffisamment agréables pour que les conducteurs apprécient le maniement de l'interface gestuelle.

En cherchant à répondre à ces interrogations, nous pourrions progresser au fur et à mesure dans la construction de stratégies de sonification. Il sera alors nécessaire de répondre aux questions : quelle stratégie choisir ? Sur quels critères la choisir ?

Intégration multisensorielle

Une fois la stratégie de sonification déterminée, un second aspect de l'utilisation de cet objet virtuel concerne sa prise en main par les futurs conducteurs. La manipulation d'objet virtuel n'est en effet pas courante, et l'utilisation de gestes pour piloter une interface n'est pas encore répandue. Il est donc essentiel de réfléchir à ce processus : comment peut-on imaginer les premières manipulations ? Quels stimuli proposer lors des premières manipulations pour une bonne utilisation du système en aveugle ? Quels stimuli sensoriels proposer pour une prise en main rapide ? Faut-il proposer les premières manipulations directement en conduite ?

Les réponses à ces questions peuvent permettre de bâtir une présentation de l'objet virtuel favorisant la création ou l'adaptation de schémas cognitifs.

Efficacité de l'interface gestuelle

Lorsque nous aurons déterminé comment présenter l'objet virtuel pour avoir des manipulations facilitées, une interface gestuelle devra être mise au point.

Il pourra alors être intéressant d'évaluer les bénéfices d'une utilisation conjointe de gestes et de sons pour manipuler l'interface simultanément à une tâche de conduite. Cette interface gestuelle permet-elle de moins détourner les yeux de la route ? La concentration des conducteurs n'est-elle pas trop importante lors du maniement de l'interface ? La qualité de la conduite est-elle affectée ? Cette évaluation donnera des informations sur la possibilité d'intégrer ce type d'interface gestuelle à de futurs véhicules. En particulier, nous pourrions nous interroger sur la pertinence de cette interface comparée aux interfaces tactiles existantes.

La figure 5 résume la problématique de sollicitation importante de la modalité visuelle lors de la manipulation de l'interface tactile en conduite. Nous proposons donc une meilleure répartition des informations pour permettre l'exécution de réglages sur l'interface tout en maintenant un niveau de sécurité important pour la tâche de conduite. La taille des flèches reflète la quantité d'information allouée à chaque modalité. Le système somatosensoriel est sollicité par l'exécution de gestes, qu'ils soient avec ou sans contact.

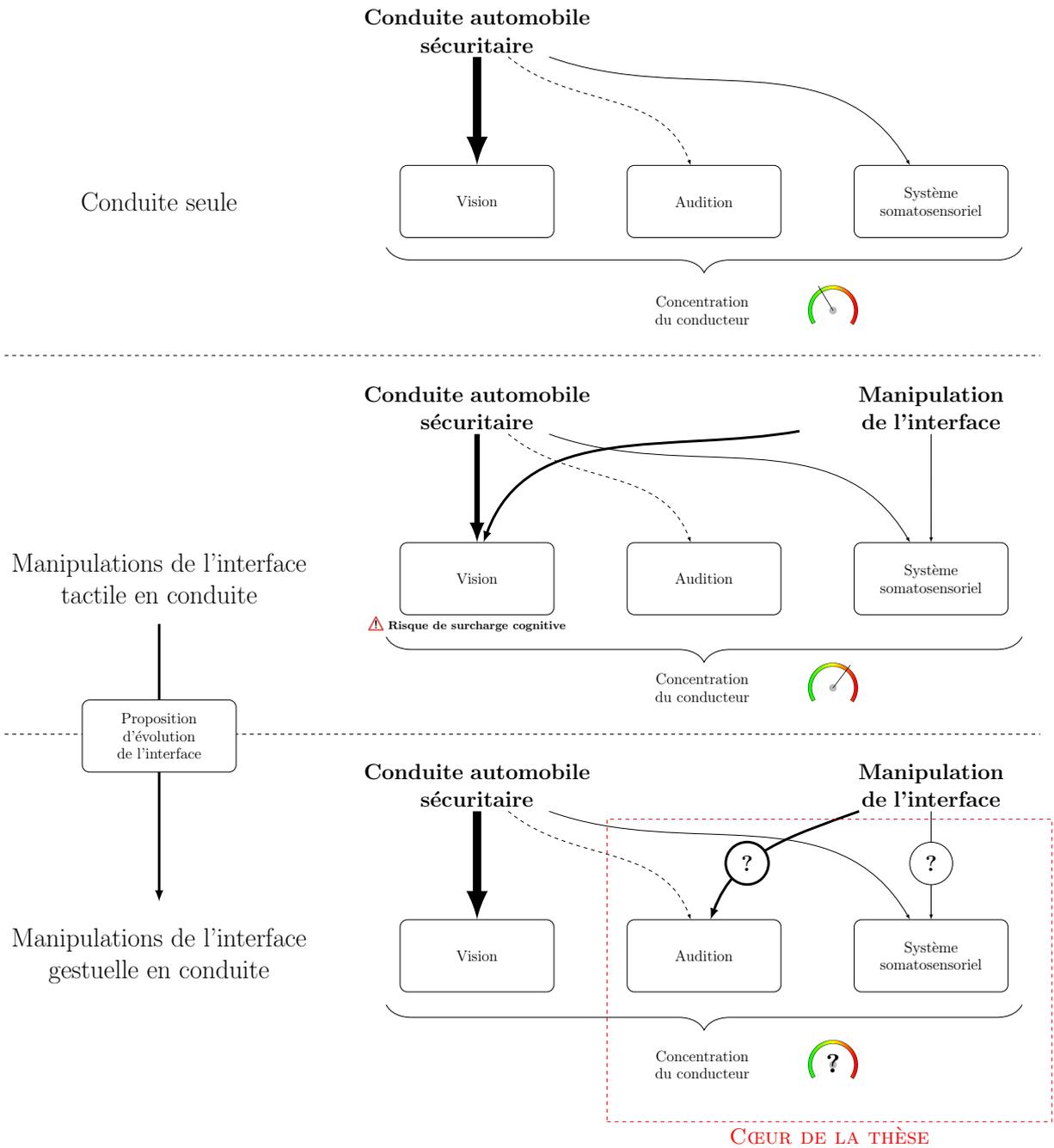


FIGURE 5 – Schéma récapitulatif de la problématique de la thèse.

Projet SoniMove

Ces travaux de thèse s’inscrivent dans le projet ANR-14-CE24-0018 SoniMove², qui vise à étudier les phénomènes de couplage entre les gestes et les sons, et plus particulièrement l’apprentissage de gestes grâce à des sons. Ce projet soulève plusieurs questions fondamentales en lien avec la perception sonore chez l’Homme, et son influence sur la réalisation de gestes. Plusieurs protocoles expérimentaux interactifs ont été mis en place, associant des gestes, des stimuli visuels et sonores. Ces expériences ont permis d’étudier l’intégration multimodale de l’ensemble de ces signaux.

Ce projet est organisé autour de trois tâches complémentaires :

- les questions fondamentales liées aux morphologies sonores et à l’immersion auditive 3D,
- les questions fondamentales liées aux relations entre sons et mouvements dans le contexte de l’ajustement du comportement moteur,
- leurs applications industrielles et sociétales axées sur la sonification des nouvelles IHM et sur l’apprentissage par le son des gestes experts pour le sport et la musique.

Deux acteurs scientifiques académiques prennent part à ces travaux : le laboratoire PRISM³ expert dans le domaine de l’acoustique et de la perception sonore, ainsi que l’institut ISM⁴ spécialiste des sciences du mouvement. La direction de la recherche et de l’innovation du Groupe PSA⁵ est un acteur industriel venant compléter le consortium de ce projet, avec un intérêt particulier pour le développement des nouvelles interfaces gestuelles, que nous allons développer dans ce document.

Les travaux réalisés au cours de ce projet ont conduit à plusieurs publications scientifiques sur l’ensemble des sujets : l’utilisation d’imitations vocales pour la recherche d’invariants sonores [Bordonné *et al.*, 2017], le guidage par le son en une ou deux dimensions [Parseihian *et al.*, 2016] [Parseihian *et al.*, 2017], la mise au point d’un système de synthèse de sons en trois dimensions [Parseihian *et al.*, 2015] [Gandemer *et al.*, 2018], les effets de trajectoires sonores en trois dimensions sur le maintien postural [Gandemer *et al.*, 2017], l’influence des stimuli sonores dans la réalisation de gestes [Thoret *et al.*, 2016a] [Thoret *et al.*, 2016b], les effets d’une contrainte physique sur l’expressivité de musiciens [Rozé *et al.*, 2016], la création d’un outil de sonification interactive [Pruvost *et al.*, 2015], ou la hiérarchisation de stratégies de sonification pour illustrer la dynamique d’un objet virtuel [Bressolette *et al.*, 2017] [Bressolette *et al.*, 2018].

2. <https://sonimove.prism.cnrs.fr>

3. Perception, Représentations, Images, Son, Musique, <https://www.prism.cnrs.fr>. Cette équipe appartenait précédemment au Laboratoire de Mécanique et d’Acoustique.

4. Institut des Sciences du Mouvement, <http://www.ism.univmed.fr>

5. <https://www.groupe-psa.com/fr>

Organisation du document

Une première partie de ce document est consacrée à l'état de l'art, dans lequel nous aborderons les grandes thématiques liées aux problématiques que nous avons énoncées précédemment. Le chapitre 1 est consacré aux interfaces et au dialogue entre l'utilisateur et la machine, que nous voulons mettre en place par l'intermédiaire de gestes et de sons. L'emploi de gestes avec et sans contact est notamment abordé dans une première partie, dans une utilisation classique ou dans le cas particulier automobile. La sonification est au cœur de la seconde partie de ce chapitre, par des icônes sonores ou des sons continus. La perception est au centre du chapitre 2, dans lequel l'intégration multisensorielle est particulièrement développée ainsi que l'approche écologique de la perception du son. Nous nous focalisons également sur des études associant les modalités gestuelles et sonores. Pour clôturer l'état de l'art, plusieurs aspects cognitifs sont abordés en chapitre 3 : le fonctionnement de la mémoire et la construction de schémas cognitifs sont traités dans une première partie, puis les distractions en conduite mobiliseront notre attention dans une seconde partie.

Nous abordons par la suite les travaux accomplis au cours de ces travaux de thèse. La construction de l'objet virtuel est d'abord développée au chapitre 4, ainsi que la mise en place et l'évaluation de stratégies de sonification au cours de deux expériences perceptives. Le chapitre 5 est consacré en premier lieu à des travaux permettant la progression vers une intégration de l'interface gestuelle en véhicule automobile. Deux expériences perceptives sont également abordées, qui concernent les évocations sonores résultant de notre système de sonification des gestes, ainsi que de l'intégration multisensorielle lors de la prise en main de l'objet virtuel. Ces connaissances ont permis la mise au point d'une interface permettant de modifier trois fonctions par trois manipulations gestuelles de l'OV, qui est développée en chapitre 6. Des retours sonores sont mis au point pour autoriser des manipulations en aveugle. Une évaluation en situation de conduite est alors proposée, avec une interface tactile comme point de comparaison. Une cinquième expérience perceptive permet de cerner les avantages et inconvénients de chaque interface, ce qui peut permettre une combinaison intéressante des deux technologies selon les situations. Ce document est clôturé par une conclusion, où une synthèse des principaux résultats est proposée, avant d'aborder les perspectives ouvertes par cette étude, sur les aspects gestuels, sonores ou attentionnels.

Première partie

Contexte de la thèse et état de l'art

Le dialogue Homme – machine

Les interfaces Homme - machine – ou *IHM* – font partie intégrante de notre vie courante. Que ce soit au domicile, en déplacement, ou sur notre lieu de travail, l'utilisation grandissante de machines rend le dialogue avec une interface quotidien. La place occupée par les smartphones rend de plus l'utilisation d'interfaces fréquente pour un grand nombre d'utilisateurs.

Depuis les années 60, avec la sortie de la première interface graphique [Engelbart et English, 1968], la grande majorité des interfaces possède un écran pour transmettre des informations à l'utilisateur, qui peut s'adresser à l'interface de différentes manières, le bouton – à deux états ou interrupteur – étant l'élément de base. Depuis, de nombreuses innovations ont permis de transformer les usages : l'invention de la souris de Ralph Benjamin, intégrée à un ordinateur par Engelbart a lié la main des utilisateurs à la position d'un curseur à l'écran. Des boutons rotatifs, autorisant plus d'états que l'interrupteur, ont facilité l'usage des interfaces. La manipulation des boutons est parfois accompagnée d'un retour haptique par l'intermédiaire de crans : deux crans en profondeur pour les boutons à deux états, ou plusieurs crans pour les boutons rotatifs. Les sons peuvent également donner une confirmation à l'utilisateur de sa manipulation. Ces retours d'informations peuvent être utile dans un véhicule, permettant au conducteur de manipuler son interface sans mobiliser la modalité visuelle sur une durée trop importante : quelques secondes peuvent suffire pour repérer la position du bouton cible, les retours sonores et/ou haptiques confirment ensuite les changements opérés.

Récemment, l'arrivée de la technologie tactile a transformé notre approche des interfaces, en permettant d'appuyer directement sur une zone d'intérêt à l'écran. Ces interfaces se retrouvent également dans les habitacles des véhicules automobiles, avec la récente apparition d'écrans tactiles qui ont remplacé de nombreux boutons. Avec ces IHM, le retour d'information haptique est appauvri : il est possible de savoir que le doigt de l'utilisateur est en contact avec l'écran, mais l'information de manipulation transmise par les crans n'est plus disponible. Les retours par le biais des sons ont alors pris toute leur importance.

Enfin, l'arrivée sur le marché de capteurs à faibles coûts permet de démocratiser de nouvelles interactions, comme l'approche gestuelle sans contact qui va probablement engager un autre changement majeur dans les usages.

Dans ce chapitre, nous nous intéresserons tout d’abord aux moyens d’interagir avec une interface grâce aux technologies tactiles ou gestuelles sans contact. Par la suite, les retours d’informations par le biais du son seront abordés, en détaillant les différentes manières de renseigner l’utilisateur par des sons brefs ou par des sons continus. Nous tacherons de donner des exemples tirés du domaine automobile tout au long de ce chapitre.

Sommaire

1.1	Les gestes comme médiateurs d’interactions	17
1.1.1	Interfaces : cas général	18
1.1.2	Les gestes sans contact en contexte automobile	22
1.2	Le son comme retour d’information	24
1.2.1	Icônes sonores	24
1.2.2	Sonification des données	27
1.2.3	Sonification appliquée à l’automobile	28

1.1 Les gestes comme médiateurs d'interactions

L'interaction gestuelle est une manière naturelle d'interagir avec notre environnement, que ce soit pour manipuler des objets ou accompagner un dialogue entre deux personnes. Le geste a été défini par Kurtenbach et Hulteen comme « un mouvement du corps qui contient de l'information. Faire au revoir de la main est un geste. Appuyer sur la touche d'un clavier n'est pas un geste, car le mouvement du doigt pour aller appuyer sur la touche n'est ni observé ni significatif. Tout ce qui compte est la touche pressée » [Kurtenbach et Hulteen, 1990]. La main, à la fois émetteur et récepteur d'informations, est un organe privilégié pour l'interaction gestuelle. Cadoz [Cadoz, 1994] définit 3 familles de gestes de la main associées à des fonctions : les gestes *ergotiques*, destinés à modifier et transformer l'environnement ; les gestes *épistémiques*, permettant d'obtenir des informations de l'environnement à travers une exploration haptique par exemple ; et les gestes dits *sémiotiques*, émetteurs d'informations à destination de l'environnement. Ce dernier type de gestes est particulièrement intéressant dans le contexte des interfaces ; ils peuvent être divisés en plusieurs catégories [Rimé et Schiaratura, 1991] :

- symboliques : les gestes ayant un sens unique pour une culture donnée, comme les gestes de la langue des signes ;
- déictiques : les gestes de pointage, avec un doigt de la main ou avec un pointeur, et servent à désigner un objet ou un endroit spécifique dans l'espace. Ils sont utilisés en réalité virtuelle par exemple, pour déplacer des objets [Bolt, 1980] ;
- iconiques : gestes utilisés parallèlement à la parole pour donner une représentation d'un objet évoqué ;
- pantomimiques : gestes pour mimer l'utilisation d'un objet ;
- gesticulations ou gestes coverbaux : gestes utilisés conjointement à la parole, pour illustrer les propos tenus.

Il est alors possible de dresser un aspect progressif dans les informations contenues par les gestes :

gesticulation → *geste iconique* → *geste pantomimique* → *geste déictique* → *geste symbolique*

Les gestes qui nous seront utiles doivent à la fois donner une information précise, mais ne pas nécessiter une concentration ou un apprentissage important : les gestes symboliques peuvent parfois être complexes à retenir, avec par exemple la langue des signes. Comme nous le verrons plus précisément aux chapitres 4 et 6, nous nous tournerons préférentiellement vers les gestes pantomimiques qui semblent être un bon compromis.

Après avoir abordé les différents types de gestes d'un point de vue général, nous allons à présent nous pencher sur le cas général des manipulations d'interfaces avec boutons physiques ou virtuels, avant de s'intéresser aux interfaces sans contact, manipulées par des gestes.

1.1.1 Interfaces : cas général

Avant l'arrivée de tablettes tactiles, les interactions avec les interfaces sont quasi exclusivement limitées à l'utilisation de boutons physiques, que ce soit des interrupteurs ou des boutons rotatifs. Chaque bouton est rattaché à une fonction qui lui est propre. Le nombre de boutons d'une interface dépend donc du nombre de fonctions modifiables. Dans ce contexte automobile, des stratégies sont mises en place pour donner la possibilité au conducteur d'interagir avec le maximum de fonctions – clignotants, essuie-glaces, climatisation, radio, musique, warning – par l'utilisation de commandes sous le volant, de boutons sur le volant, proches du levier de vitesse ou au centre de la planche de bord. Un exemple est présenté sur la figure 1.1 avec une Peugeot 308 de première génération.



FIGURE 1.1 – Vue intérieure d'une Peugeot 308 de première génération.

Cependant, il existe plusieurs raisons qui ont amené les constructeurs automobiles à faire évoluer leurs interfaces. Tout d'abord, la transformation des habitudes et usages des conducteurs provoque une forte demande, avec un attrait pour les nouvelles technologies. Ces changements coïncident également avec l'arrivée de nouvelles fonctions – GPS, aides à la conduite, commandes vocales – qui conduiraient à implanter des nouveaux boutons dédiés. Cette nouvelle série de boutons pourrait entraîner une surcharge visuelle au niveau de l'habitacle, avec une ergonomie peu satisfaisante. Enfin, lors de l'utilisation de boutons, il existe une distance entre le lieu de l'interaction et celui du retour d'information. C'est par exemple le cas pour les manipulations de la radio sur la figure 1.1 : les boutons sont situés au centre alors que les retours à l'écran sont situés plus haut. Cette différence de localisation peut entraîner des difficultés de manipulations – comme des allers-retours visuels entre les boutons et l'écran – et un sentiment de distance entre l'utilisateur et l'interface.

Après avoir évoqué les interfaces pilotables par des boutons, nous avons cité certaines limites liées à leur utilisation dans l'habitacle d'un véhicule. Nous allons à présent nous intéresser à deux catégories d'interfaces : les écrans tactiles comme exemple d'une interface se manipulant par des contacts entre l'utilisateur et l'IHM, et les interfaces sans contact dans un second temps. C'est en effet la rupture technologique que les interfaces gestuelles vont amener : l'utilisateur n'interagira plus au contact d'un bouton, d'une souris ou d'un pointeur, mais par des gestes effectués en l'air.

Interfaces avec contact : les écrans tactiles

L'apparition du tactile comme modalité d'interaction a bouleversé les usages des interfaces. Cette technologie a l'avantage de centraliser le lieu d'interaction avec l'IHM. De plus, les lieux d'envoi et de retour des informations sont fusionnés en faisant coïncider la dalle tactile et l'écran. La tablette tactile permet d'interagir avec des boutons virtuels, qu'il est possible de presser de la même manière que des boutons physiques. Cependant, certains gestes porteurs de sens sont apparus, comme par exemple le balayage – aussi appelé *swipe*. Ce mouvement permet par exemple de faire défiler une page d'un document en posant directement le doigt sur la page virtuelle, et en la faisant défiler dans la direction voulue. Ce type d'interaction est parfois utilisé avec des barres de défilement sur les IHM tactiles. Une autre manipulation particulièrement utilisée pour les interfaces tactiles est le pincement à deux doigts, le plus souvent exécuté avec le pouce et l'index. Ce geste, imaginé dès les années 1980 [Krueger *et al.*, 1985], est aujourd'hui passé dans le langage courant des interfaces tactiles, et est associé avec la notion d'agrandissement ou de rétrécissement [Wobbrock *et al.*, 2009]. Il est intéressant de noter que ce geste, qui n'avait pas de réalité physique, est maintenant associé intuitivement à ce type de fonction. Cet appariement semble évident et est utilisé sur la plupart des interfaces tactiles. Ces différentes interactions sont très largement répandues dans l'utilisation des smartphones ou des tablettes tactiles.

Les écrans tactiles se font également leur apparition dans le domaine automobile. La Peugeot 308 de seconde génération est par exemple équipée d'un écran central multifonction, sur lequel sont pilotées des fonctions de climatisation, musique, GPS ou radio. Peu de boutons physiques subsistent au niveau de l'interface centrale, comme le montre la figure 1.2 : le bouton du volume, très couramment utilisé, un bouton de verrouillage du véhicule ou encore le bouton du warning qui doit être accessible de manière permanente et rapide.



FIGURE 1.2 – L'écran central dans une Peugeot 308 de seconde génération.

Les tablettes tactiles semblent cependant limitées dans leur interactivité, puisqu'elles se basent essentiellement sur des boutons virtuels que l'utilisateur peut presser. Enfin les boutons rotatifs à crans permettent, grâce à leurs retours d'informations haptiques, des manipulations sans nécessiter une grande attention visuelle. La technologie tactile manque de ces retours d'informations, même si l'utilisation de retours sonores s'est répandue, et que des travaux actuels visent à ajouter une sensation tactile [Monnoyer *et al.*, 2016] [Bernard *et al.*, 2018]. Leur manipulation peut donc demander une attention visuelle importante, peu sécuritaire en conduite.

L'arrivée sur le marché de nouveaux capteurs gestuels rend possibles de nouvelles interactions, qui peuvent intéresser les acteurs industriels de l'automobile par leur aspect attractif et technologique. Les usages des interfaces sont donc amenés à changer, et il est nécessaire de réfléchir à de nouveaux types d'interactions.

La captation gestuelle sans contact

Même si l'interaction sans contact peut sembler futuriste et encore peu exploitée, elle est utilisée depuis longtemps notamment dans le domaine musical. Le premier instrument se jouant sans contact physique est le Thérémine. Cet instrument de musique inventé en 1919 par le russe Lev Sergueïevitch Termen dispose de deux antennes qui captent les perturbations électromagnétiques induites par les mouvements des mains du musicien. Ces perturbations sont ensuite répercutées sur la fréquence ou l'intensité du son diffusé. La figure 1.3 donne un aperçu de son utilisation. Ces instruments, qui font partie des premiers instruments de musique électronique, sont toujours utilisés actuellement, un exemple est donné figure 1.4.



FIGURE 1.3 – Lev Termen jouant du Thérémine.



FIGURE 1.4 – Exemple de Thérémine actuel vendu dans le commerce.

Depuis cette époque, de nombreux capteurs ont été développés dans l'industrie, se basant sur différentes technologies. Certains nécessitent l'utilisation d'un objet médiateur entre le geste capté et l'ordinateur. C'est par exemple le cas du Myo de *Thalmic Labs*¹, où un bracelet permet de capturer la tension des muscles de l'utilisateur pour remonter au geste effectué. D'autres techniques consistent à porter un gant pour capturer les mouvements de la main, le premier ayant été mis au point par Zimmerman [Zimmerman *et al.*, 1987]. Certaines sociétés commercialisent toujours ce genre de capteur^{2 3}.

1. <https://www.myo.com/>

2. <http://maestroglove.com/>

3. <http://www.cyberglovesystems.com/>

D'autres technologies ne nécessitent pas de médiateur physique. Elles captent des éléments de l'environnement grâce à des caméras. C'est par exemple le cas de la Kinect de *Microsoft*⁴, qui combine un capteur à infrarouges et une caméra classique pour avoir accès à la profondeur et à la couleur d'objets de l'environnement. Concernant spécifiquement la main, l'un des capteurs les plus populaires est la Leapmotion⁵ qui utilise deux caméras infrarouges. Sa popularité vient notamment de la possibilité d'avoir à la fois accès à des données de gestes – centre de la main, positions des doigts – mais également à une reconstruction squelettique de la main grâce aux algorithmes fournis, comme l'illustre la figure 1.5.



FIGURE 1.5 – Le capteur Leapmotion et ses algorithmes permettent de reconstruire schématiquement les mains de l'utilisateur.

D'autres capteurs dits *temps de vol* utilisent un principe similaire au sonar : le temps que met un faisceau lumineux pour faire un aller-retour entre la caméra et l'objet le plus proche est mesuré pour chaque pixel. Une image de la scène est ainsi déduite, où la distance est évaluée avec précision. Plusieurs capteurs utilisent ce principe, c'est par exemple le cas de la caméra SwissRanger de *Heptagon*⁶ ou de la Camboard Pico de *PMD*⁷.

Ces systèmes de captation des gestes sans contact, notamment grâce à des coûts de plus en plus attractifs, se démocratisent. Du côté des smartphones, une technologie sans contact "Air gesture" est sorti en 2013 avec le modèle S4 de *Samsung*, où il est possible d'utiliser des gestes de balayage de la main pour faire défiler des pages. Cette technologie n'a cependant pas été conservée sur les modèles suivants de la marque. *Sony* a également utilisé la technologie gestuelle sur une enceinte connectée MHC-GT4D⁸, où il est possible de contrôler certains paramètres de la musique jouée avec des balayages de la main ou des gestes de rotation de la main.

4. <https://www.xbox.com/fr-FR/xbox-one/accessories/kinect>

5. <https://www.leapmotion.com/>

6. <http://hptg.com/industrial/>

7. <http://pmdtec.com/picofamily/>

8. <https://www.sony.fr/electronics/chaines-hifi/mhc-gt4d>

1.1.2 Les gestes sans contact en contexte automobile

Les interfaces automobiles ont récemment évolué en intégrant des interfaces tactiles, comme le montre la figure 1.2 avec l'exemple de la Peugeot 308. L'utilisation grandissante des smartphones et des tablettes entraîne une envie chez les conducteurs de technologies similaires dans leurs véhicules. Cette demande forte des conducteurs couplée avec l'arrivée de capteurs peu chers sur le marché, entraîne une évolution des habitacles des voitures, équipés de plus en plus de technologies. Le développement d'interfaces gestuelles peut donc permettre de proposer un nouveau type de dialogue avec les IHM, argument d'attractivité pour le constructeur automobile. Mais nous faisons également l'hypothèse que cette technologie peut conduire à des manipulations plus sécuritaires qu'avec les tablettes tactiles dans un contexte de conduite, en sollicitant moins la modalité visuelle des conducteurs. En effet le sens de la proprioception, abordé en chapitre 2, permet de ressentir et contrôler des gestes sans nécessiter un contrôle visuel. De cette manière, les yeux des conducteurs pourraient être concentrés sur la route.

Deux stratégies ont été privilégiées dans le développement de ce nouveau type d'interfaces : l'utilisation de gestes proches du volant, ou de gestes dans la zone de l'interface centrale.

Gestes proches du volant

Les gestes proches du volant, et en particulier les micro-gestes [Wolf *et al.*, 2011] peuvent être particulièrement intéressants dans un contexte automobile [Jacob *et al.*, 2015]. Ces gestes brefs peuvent permettre de faire une tâche secondaire, interagir avec le système central par exemple, tout en gardant une attention suffisante sur la tâche principale de conduite et un contrôle du volant à deux mains. Wolf et Mahr ont par exemple comparé des micro-gestes basés sur des mouvements des doigts : des tapotements sur le volant, des bayages de doigts ou encore des gestes de zoom comme le montre la figure 1.6 [Wolf *et al.*, 2011] [Mahr *et al.*, 2011]. Ces gestes, liés à des fonctions de confort dans [Mahr *et al.*, 2011], peuvent permettre une interaction intéressante en conduite, avec de potentiels gains d'ergonomie et d'attractivité. L'aspect sécuritaire sur les mouvements du volant reste cependant à évaluer dans un contexte de conduite demandeur d'attention, comme le souligne Mahr *et al.*.



FIGURE 1.6 – Exemple de micro-geste de zoom au volant tiré de [Mahr *et al.*, 2011].

Du côté des acteurs industriels, *Continental* propose un système de micro-gestes à proximité du volant basé sur des balayages des doigts, ou un geste de sélection avec un bref geste d'appui⁹. Ce système, basé sur un capteur temps de vol, permet de naviguer dans un menu, répondre à des appels, lire de la musique ou programmer la navigation.

Les micro-gestes proches du volant peuvent être intéressants d'un point de vue sécuritaire, mais sont limités spatialement à cause de la présence du volant dans la main de l'utilisateur. Une seconde solution est de proposer des gestes sans contact proches de l'interface centrale, ce qui est déjà le cas des interfaces tactiles. Ces gestes pourront permettre une plus grande liberté de mouvement ainsi qu'une plus grande diversité gestuelle.

9. <https://www.continental-automotive.com/en-gl/Passenger-Cars/Interior/Comfort-Security/Driver-Status/Gesture-Control>

Gestes dans l'axe central du véhicule, proches de l'interface

Même s'il est possible de proposer de faire des gestes en face du conducteur [Alpern et Minardo, 2003], l'interface centrale semble être le lieu privilégié pour les interactions gestuelles sans contact. Riener a montré que les utilisateurs ont tendance à faire naturellement les gestes dans un triangle entre le volant, le rétroviseur intérieur et le levier de vitesse [Riener *et al.*, 2013]. Les utilisateurs ont de plus acquis cette habitude avec les interactions tactiles que nous avons vu précédemment. C'est également cette localisation qui a été choisie dans de nombreuses études sur les gestes en véhicule [Jaschinski *et al.*, 2016] [Ohn-Bar et Trivedi, 2014] [Ohn-Bar *et al.*, 2012] [Loehmann *et al.*, 2013]. Il est alors possible de proposer des gestes plus complexes que des micro-gestes, pouvant être réalisés tout en allouant une concentration suffisante à la tâche de conduite. Ces interactions sont limitées à une seule main, l'autre étant requise pour la conduite. Plusieurs études mettent essentiellement en évidence des gestes de balayage vers la gauche, la droite, le haut, le bas ainsi que des gestes circulaires ou symboliques comme le poing fermé avec ou sans pouce levé [Ohn-Bar *et al.*, 2012] [Alpern et Minardo, 2003] [Akyol *et al.*, 2000] [Zobl *et al.*, 2003]. Récemment, des systèmes gestuels ont fait leur apparition dans les véhicules. Certains fournisseurs des constructeurs automobiles proposent des interfaces gestuelles proches de l'écran central^{10 11} avec des gestes de balayages ou symboliques. Les premiers véhicules équipés de technologie gestuelle ont été proposés par Volkswagen¹² et BMW¹³. Les gestes principaux de ces systèmes sont les balayages de la main pour faire défiler des menus, chez BMW un geste de rotation du doigt manipule le volume sonore, et un geste de présélection de menu est proposé grâce à un pointage à deux doigts.

Nous avons jusqu'à présent passé en revue plusieurs manières de s'adresser au système central du véhicule pour donner des instructions sur une commande à effectuer. Les interfaces tactiles équipent actuellement les véhicules, et nous avons abordé plusieurs limitations provenant de la disparition de retours haptiques par rapport aux boutons physiques. Leur manipulation nécessite donc une attention visuelle plus soutenue, ce qui peut avoir des conséquences en terme de sécurité en conduite. Les technologies sans contact offrent une nouvelle possibilité d'interagir avec le système central, sans nécessiter un contrôle visuel du conducteur. De plus, cette technologie peut intéresser les constructeurs automobiles par leur attractivité et leur ergonomie. Plusieurs zones semblent intéressantes pour effectuer des gestes. Les micro-gestes permettent par exemple de garder les mains sur le volant. Dans le cadre de cette thèse, nous avons choisi une zone dans l'axe central du véhicule, proches de l'interface pour exécuter des gestes. Cette zone permet de faire des gestes beaucoup moins contraints qu'au niveau du volant, et il semble que ce lieu est bien adapté pour faire les gestes dans un véhicule [Riener *et al.*, 2013].

En complément des retours visuels donnés par des écrans, les interfaces ont intégré depuis plusieurs années des retours sonores qui peuvent être de différentes natures. Ces retours d'informations peuvent être particulièrement nécessaires pour accompagner un retour visuel ou même pour le supplanter lorsqu'il n'est pas disponible. Nous allons donc maintenant nous intéresser aux interactions sonores entre l'interface et l'utilisateur données par le biais de la sonification, avec des icônes sonores ou des sons continus.

10. <http://www.unimax.com.tw/portfolio/in-vehicle-gesture-control/>

11. <http://www.mechaless.com/english/home.html>

12. <http://www.volkswagen.co.uk/technology/comfort-and-convenience/gesture-control#overlay-close>

13. <http://driving.ca/bmw/7-series/auto-news/news/how-it-works-bmw-gesture-control>

1.2 Le son comme retour d'information

Lors de l'interaction avec une interface, l'utilisateur a besoin d'un retour d'information, pour lui indiquer que son action a bien été prise en compte, mais également qu'il a manipulé la fonction voulue et de la manière désirée. Ce retour peut être assuré par une information visuelle donnée par l'écran central du véhicule, comme c'est le cas avec les écrans tactiles. Cependant, lors d'une tâche de conduite, la modalité visuelle doit être concentrée sur la route, l'utilisation de l'audition est donc une alternative. La modalité auditive est un sens relativement peu sollicité pendant la conduite, alors qu'il est en activité permanente et permet un contrôle en continu. De plus, l'audio numérique est un outil qui permet de contrôler précisément le son diffusé via les enceintes de l'automobile, et ainsi transmettre des informations par le son. L'audition a également l'avantage d'être omnidirectionnelle, ce qui peut être particulièrement intéressant dans des situations qui nécessitent une réaction rapide. La sonification, définie comme l'utilisation de sons non-verbaux pour transmettre une information [Kramer *et al.*, 1999], est une solution intéressante pour donner un retour informatif au conducteur. En exploitant les propriétés des sons, qu'ils soient brefs ou durent plusieurs secondes, il est alors possible de véhiculer un grand nombre d'informations qui peuvent compléter le retour visuel d'une interface, ou être la source informationnelle principale.

Le retour sonore est en effet un moyen efficace pour véhiculer de l'information. Un exemple concerne la pratique d'un instrument de musique. Le son produit par un violon dépendant fortement de la position des doigts sur le manche et du maniement de l'archet, il est possible de déterminer des positions et mouvements convenables en écoutant les variations du son. La sonification, en imitant ce processus interactif, peut donc inclure l'utilisateur dans une boucle action - perception, comme nous allons le voir.

Nous allons tout d'abord développer les différents types de retours sonores, en abordant dans un premier temps les sons brefs appelés icônes sonores, puis nous nous intéresserons à différents types de sons continus. Enfin, une partie sera consacrée à la sonification appliquée au domaine automobile, où un certain nombre de systèmes exploitent cette technologie pour informer le conducteur.

1.2.1 Icônes sonores

Les icônes sonores sont des analogues des icônes visuelles. Sur un ordinateur, les fichiers sont représentés visuellement par des icônes indiquant leur contenu : les fichiers de texte par l'image d'une feuille de papier, les fichiers comprenant des photos par l'image d'une photo générique, ou les dossiers par l'image d'une pochette cartonnée. Cette représentation iconique permet de donner une information rapide aux utilisateurs, probablement plus rapide qu'avec des mots [Hemenway, 1982], pour leur permettre de faire facilement et rapidement le tri pour trouver le fichier voulu. De plus, comme ces icônes représentent des objets, elles permettent de franchir les barrières entre les langues [Kolers, 1969].

Le même principe est utilisé par les icônes sonores : l'information est véhiculée par un son bref et la transmission de l'information est rapide et efficace. On distingue deux grandes familles d'icônes sonores, qui se différencient par leur construction.

Les icônes auditives ou *auditory icons*

Les icônes auditives ou *auditory icons* ont la particularité d'être construites à partir de sons de notre environnement, dont elles retiennent les grands traits. Elles sont utilisées par analogie aux événements produisant ces sons [Gaver, 1986], et se réfèrent à l'approche écologique de la perception qui sera abordée en section 2.3. La philosophie de Gaver à propos des *auditory icons* est « d'entendre le monde, pas les sons », les sons n'étant que des vecteurs d'événements. La force de ces sons réside dans la quantité d'information portée par un son bref : les sons de l'environnement étant connus, l'action véhiculée par l'icône est facilement reconnue par l'utilisateur, qui l'associe à l'action effectuée et en déduit une information.

Cette catégorie d'icône sonore est couramment utilisée sur les interfaces, et en particulier sur les ordinateurs. Un exemple courant est le son lu lorsque la corbeille est vidée, qui est le froissement d'un papier. Même si le son ne correspond pas exactement à l'action de vider une corbeille, l'utilisateur comprend directement que l'action a bien été effectuée. Gaver a même mis au point une interface d'ordinateur où sont utilisées les icônes auditives, le SonicFinder¹⁴ [Gaver, 1989]. Sur cette interface, un grand nombre d'actions sont illustrées par un *auditory icon* : la sélection, le déplacement, l'ouverture ou la fermeture d'un fichier, ou encore le changement de taille d'une fenêtre. Les sons peuvent être redondants avec une information visuelle, ou la compléter. Par exemple, une estimation de la taille d'un fichier est donnée par un son d'impact plus ou moins haut en fréquence lors de sa sélection. Gaver souligne l'avantage d'une sensation de manipulation plus directe de l'interface, qui proviendrait de l'association – ou *mapping* – entre l'icône auditive et l'événement qu'il représente. L'utilisation de métaphores communes entre les icônes visuelles et auditives semblent être une solution intéressante pour une prise en main rapide de l'interface.

Cet exemple d'interface avec des retours d'informations sonores riches montre le potentiel informatif des icônes auditives. Pour la construction d'une interface, ces icônes semblent être facilement et rapidement prises en main par les utilisateurs [Dingler *et al.*, 2008] [Garzonis *et al.*, 2009] et relativement intuitives [Garzonis *et al.*, 2009]. Nous allons à présent nous intéresser à une seconde catégorie d'icônes sonores plus abstraites.

Les *earcons*

Un *earcon* est un motif sonore abstrait relativement court, qui est utilisé comme retour d'information pour les interfaces [Blattner *et al.*, 1989]. À partir d'un *earcon*, il est possible de créer des familles d'*earcons*, en modifiant la hauteur du son par exemple. De la même manière, le rythme d'enchaînement de plusieurs *earcons*, la dynamique, le timbre ou la longueur des sons peuvent permettre de créer différents *earcons* familiers. À la différence des *auditory icons*, ils ne sont pas composés de sons de l'environnement, et peuvent être combinés de différentes manières.

Les différents *earcons* sont associés à des manipulations particulières de l'interface. La figure 1.7 donne l'exemple de deux *earcons* similaires associés à des fonctions de création et suppression de fichier.

14. Il est possible de voir et d'entendre le SonicFinder en action à l'adresse suivante : <https://vimeo.com/channels/billgaver/158610127>

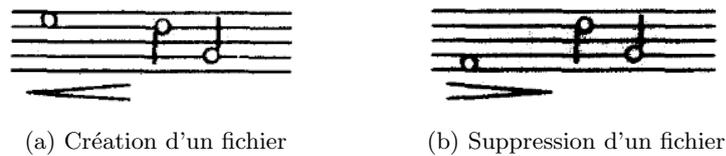


FIGURE 1.7 – Exemple de deux *earcons* tirés de [Blattner *et al.*, 1989].

Sur cet exemple de sons à 3 notes, la seule différence provient de la première note qui est plus aiguë et moins appuyée dans le premier cas. Cette particularité donne l'information aux utilisateurs de deux manipulations ayant des conséquences voisines mais différentes. Il est donc possible de construire de grandes familles d'*earcons* cohérentes pour transmettre des informations complexes [Blattner *et al.*, 1989]. Leur efficacité a été montrée par Brewster [Brewster *et al.*, 1993]. Une limite des *earcons* est leur caractère non intuitif, puisque la relation son - fonction, métaphorique, doit tout d'abord être apprise. Cette durée d'apprentissage peut varier selon la quantité d'*earcons*, et semble être plus importante que pour les *auditory icons* [Dingler *et al.*, 2008].

D'autres types d'icônes sonores ont également été mis au point : les *spearcons* en utilisant de la parole accélérée [Walker *et al.*, 2006], ou encore les *morphocons* [Parseihian et Katz, 2012], mais sont moins utilisés.

Les icônes sonores pour les interfaces automobiles

Comme l'ont souligné Walker *et al.*, les *auditory icons* ont la particularité d'être difficiles à créer pour une grande quantité de menus et sous-menus, puisque ces sons proviennent de métaphore de l'environnement [Walker *et al.*, 2006]. Dans un contexte automobile, il est nécessaire de disposer d'une grande diversité de types de sons. Les sons proposés doivent avoir des propriétés différentes s'ils indiquent la présence d'un clignotant, la bonne prise en compte d'une manipulation ou une alarme. Par exemple, il a été montré que la perception d'urgence d'un son d'alarme est liée aux paramètres d'intensité, de hauteur et de répétition d'un son [Haas et Edworthy, 1996]. Cet exemple montre la nécessité de maîtriser l'ensemble des paramètres des sons proposés. Pour les constructeurs automobiles, le *design* des sons est également un aspect intéressant pour apporter du plaisir à l'usage et pour véhiculer une identité de marque. Ces différentes raisons, combinées à la facilité de création et à la plasticité des *earcons* et des familles d'*earcons*, l'ont donc naturellement amené à être l'icône sonore la plus utilisée dans le domaine automobile.

Il est également possible de penser à un autre type de retour d'information naturel, qui a été abondamment cité par les sujets des expériences perceptives présentées dans ce document : la voix. Ce canal vocal est en effet très utilisé pour l'échange d'informations, cependant ce sujet ne sera pas abordé dans cette thèse. Il sera par contre envisageable pour le Groupe PSA de rajouter un retour vocal en plus de la stratégie de sonification qui sera développée dans la suite de ce manuscrit. En effet, la sonification [Kramer *et al.*, 1999] peut permettre de transporter des informations de manière efficace et ne nécessite pas de traduction selon le pays. Nous supposons donc ici que la sonification peut remplir le rôle de retour d'information pour l'interface.

1.2.2 Sonification des données

La sonification, comme précisé plus haut, consiste à transmettre de l'information par des sons non-vocaux [Kramer *et al.*, 1999]. Nous avons abordé dans le paragraphe précédent le concept d'icône sonore, et leur utilisation dans un contexte automobile avec l'exemple de sons d'alarme, qui nécessitent une réaction rapide [Walker et Nees, 2011].

Même si les icônes sonores transmettent efficacement une information ponctuelle, il peut être intéressant de sonifier un flux de données par un retour sonore continu. Walker et Nees désignent ce procédé par la sonification de statut dans laquelle une évolution dans le retour sonore va renseigner un changement d'état [Walker et Nees, 2011]. C'est par exemple le cas de l'électrocardioscope, qui émet un *earcon* à chaque pulsation du cœur d'un patient. Même si l'information est donnée visuellement, le son permet au personnel hospitalier de suivre de manière continue le rythme cardiaque tout en faisant une autre tâche nécessitant leur attention visuelle. En cas de dégradation du rythme cardiaque du patient, le changement de temps entre deux *earcons* permet de notifier son changement d'état. La sonification des données continues est également intéressante pour un compteur Geiger, inventé par Hans Geiger en 1908, qui permet de mesurer la radioactivité environnante. Dans ce cas, les données ne sont pas disponibles visuellement, mais uniquement à travers le son. Avec cet appareil, plus la radioactivité est importante, et plus l'enchaînement des *earcons* émis est rapide. Le Cosmophone¹⁵, développé notamment par l'équipe de PRISM et le Centre de Physique des Particules de Marseille, est un autre exemple de sonification des données : la présence et la vitesse de particules cosmiques sont sonifiées pour rendre compte de ce phénomène invisible.

Sonification interactive

Alors que la sonification de données peut être une simple prise d'information, la sonification interactive [Hermann et Hunt, 2004] [Hermann et Hunt, 2005] [Hermann, 2008] [Serafin *et al.*, 2011] vise à inclure l'utilisateur dans le processus de sonification. Dans son environnement physique, l'être humain a l'habitude d'entendre des sons comme conséquences de ses actions [Kendall, 1991]. La boucle action - perception ainsi formée, voir figure 1.8, témoigne de l'interactivité homme - environnement naturellement présente. En ajoutant de l'interactivité à la sonification, la possibilité est donnée à l'utilisateur d'influencer le retour sonore. Il peut donc retrouver l'interactivité homme - environnement à travers la relation homme - machine.

En reprenant un exemple précédent, le compteur Geiger peut être un cas de sonification interactive. L'appareil étant portable, il est possible de se déplacer avec, de constater les différences de radioactivité selon la position géographique, et potentiellement de s'éloigner de la zone radioactive dangereuse. Le son du compteur guide alors bien les pas de l'utilisateur.

Il est alors possible, à partir de ce principe de sonification interactive, de proposer des systèmes plus complexes de guidage, en associant des sons soit à des paramètres gestuels, soit à l'erreur par rapport à un geste de référence. Pour illustrer la première stratégie, Schaffert associe par exemple l'accélération d'un bateau muni d'avirons à la hauteur d'un son, ce qui permet de donner un retour d'information pour des rameurs déficients visuels. Il a été constaté une augmentation de la vitesse du bateau par rapport à une condition sans sonification [Schaffert et Mattes, 2012]. La vitesse d'écriture peut également être un paramètre de sonification intéressant [Danna *et al.*, 2013a], qui a permis d'améliorer l'écriture d'enfants atteints de dysgraphie [Danna *et al.*, 2013b]. La sonification de paramètres spatiaux peut également servir de guidage pour l'aide à la navigation [Parseihian, 2012] [Piccolo *et al.*, 2018]. Les informations spatiales, dont la perception sera abordée en début du

15. <http://cosmophone.in2p3.fr>

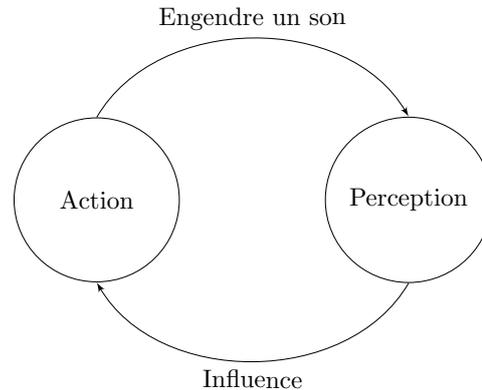


FIGURE 1.8 – Le principe de la boucle action - perception pour la sonification interactive.

chapitre 2, sont donc également à prendre en compte dans notre construction de stratégies sonores. Une seconde stratégie consiste non pas à sonifier un paramètre gestuel, mais l'erreur par rapport à un geste de référence. C'est par exemple le cas de Boyer *et al.* où le geste d'un mouvement angulaire est enregistré [Boyer *et al.*, 2013]. Si un geste exécuté s'éloigne au-delà d'une certaine limite du geste de référence, un son vient alors signaler l'erreur et guide l'utilisateur vers la trajectoire voulue. Ainsi, il est montré que les différences de trajectoires ont tendance à diminuer avec le temps : les sujets de l'expérience menée ont réussi à apprendre le geste désiré. Ce type de sonification peut être appliqué dans le domaine médical, pour guider le geste lors d'une intervention chirurgicale [Wegner, 1998]. Un aspect plus précis concernant la relation sons - gestes sera par ailleurs abordé en section 2.4.2.

Ces différents exemples montrent l'intérêt de la sonification dans différents contextes. Parfois en support de l'information visuelle, le retour sonore par sonification peut également être l'information principale comme pour le compteur Geiger ou les rameurs déficients visuels. Les constructeurs automobiles se sont montrés intéressés par les avantages de la sonification pour donner un retour d'information au conducteur, dans le contexte particulier de la conduite.

1.2.3 Sonification appliquée à l'automobile

Un premier exemple célèbre de la sonification interactive dans le domaine automobile est le radar de recul [Del Signore, 1997]. Ce système a été adopté par l'ensemble des constructeurs grâce à l'avantage des sons sur les informations visuelles dans ce cas précis : il était à l'époque peu pratique de voir les obstacles situés derrière le véhicule tout en conduisant. La transmission de l'information de la distance entre l'obstacle et l'arrière du véhicule est donnée par une série d'*earcons* ce qui permet au conducteur de se garer tout en percevant la marge de manœuvre dont il dispose. L'information est ici contenue dans le rythme de lecture des *earcons*, passant de plusieurs dixièmes de secondes lorsqu'un obstacle est éloigné à un son continu lorsque la distance est très faible. Le son, par son côté omnidirectionnel associé à la perception sonore fine des aspects temporels, a donc été privilégié dans ce cas. Depuis, des caméras ont permis de donner la vue de l'arrière du véhicule par l'écran central, voir figure 1.9, mais le retour d'information auditif a été conservé et amélioré grâce à la spatialisation.



FIGURE 1.9 – La vue arrière du véhicule donnée sur l'écran central d'une Peugeot 308 de seconde génération.

La conduite automobile est une activité qui nécessite une concentration importante de l'utilisateur, notamment au niveau visuel [De Waard, 1996] comme nous le verrons en chapitre 3. Il est essentiel que le conducteur garde les yeux sur la route pour prévenir le risque d'accident : il devient alors difficile de transmettre d'autres informations par le canal visuel. La sonification peut alors être une solution. Le cas particulier des véhicules électriques a notamment été étudié par Denjean [Denjean, 2015], où le retour d'information du bruit moteur lors de l'accélération n'est plus présent. Une stratégie de sonification a alors été mise en place pour assurer ce retour, en utilisant l'illusion auditive de Shepard-Risset [Shepard, 1964] [Risset, 1978], ce qui a permis de transmettre une information dynamique précise. Il a en effet été constaté que les conducteurs de véhicules électriques, par l'absence de bruit moteur, ont tendance à sous-estimer leur vitesse [Denjean *et al.*, 2013]. Cette sonification permet de rééquilibrer cette perception de vitesse en redonnant une information au conducteur sur son régime moteur, sans solliciter son canal visuel.

Certains travaux commencent à porter sur la sonification des interfaces des véhicules automobiles. Tardieu *et al.* ont par exemple étudié l'influence de la sonification d'items dans le menu d'une interface lors de l'accomplissement de deux tâches simultanées [Tardieu *et al.*, 2015]. Les sujets de l'expérience menée doivent prioritairement détecter une cible visuelle, pour simuler une attention visuelle analogue à une tâche de conduite. Conjointement, il leur est demandé de naviguer dans une interface similaire à une IHM automobile grâce à un clavier physique. Les résultats montrent par exemple que le regard des sujets est plus concentré sur la tâche prioritaire dans la condition avec une sonification de l'interface, comparée à une situation de contrôle sans son. Ce type d'étude est amené à se développer avec la volonté de réduire l'attention visuelle du conducteur sur l'interface, pour faire porter aux sons de plus en plus de contenus informatifs. La modalité gestuelle, grâce à l'utilisation conjointe d'informations proprioceptives que nous allons aborder dans le chapitre 2, et d'une sonification efficace peut permettre d'atteindre cet objectif. L'étude menée par Jaschinski *et al.*, où une interface sans contact pilotable par différents gestes est proposée, va dans ce sens [Jaschinski *et al.*, 2016]. Les retours sonores, des *auditory icons* et *earcons*, permettent de naviguer dans un menu sans nécessiter de retour visuel, ce qui peut être intéressant du point de vue du conducteur. L'analyse du regard des conducteurs a notamment permis de montrer que les conducteurs regardent moins l'interface lorsqu'une sonification de l'IHM est proposée.

Conclusion du chapitre

L'introduction d'interfaces pour une utilisation quotidienne amène une évolution dans les modalités d'interactions. L'interactivité connue avec les objets de notre environnement est fortement réduite lors de dialogues avec des IHM. De plus, il est nécessaire d'apprendre un autre type de langage pour pouvoir donner et percevoir des informations d'une interface. La manipulation de boutons est une première étape dans ce dialogue, ainsi que l'utilisation de pointeur en 2 dimensions. Les interfaces automobiles sont également pourvues de nombreux boutons, et l'arrivée de nouvelles fonctions peut poser problème pour l'implantation de boutons dédiés.

Les interfaces tactiles, en fusionnant le lieu d'envoi et de perception visuelle de l'information, augmentent l'interactivité donnée aux utilisateurs. L'utilisation de l'interface peut donc sembler plus intuitive et plaisante. Ces interfaces ont également amené l'utilisation de nouveaux gestes, comme les balayages de la main ou le geste de zoom, qui est passé dans le langage courant des interfaces. L'industrie automobile a implanté ce nouveau type d'interface, avec un gain de place important, et la possibilité de concentrer l'utilisation de fonctions sur l'écran multi-fonction, au centre de la planche de bord.

L'arrivée de nouveaux capteurs gestuels sans contacts permet un nouveau type de dialogue avec les interfaces. Leurs progrès technologiques rendent possible une implantation dans les véhicules automobiles. Il est alors possible d'augmenter l'interactivité entre l'utilisateur et son interface.

Après avoir exécuté une commande sur une interface, l'utilisateur attend un retour d'information de confirmation. Un grand nombre d'interface propose l'association de retours visuels et sonores. En particulier, l'utilisation d'*earcons* sonores permet de donner un retour d'information rapide et efficace à l'utilisateur. Cependant, des stratégies de sonification plus complexes peuvent être mises en place pour donner un retour d'information continu. Il est alors possible d'inclure l'utilisateur dans une boucle action - perception, où il est véritablement acteur d'un processus de sonification interactif.

L'augmentation de l'interactivité des interfaces amène alors la question de la compréhension des signaux par l'utilisateur. Pour avoir des interactions fluides et rapides, et donner les informations aux moments opportuns, il est nécessaire de s'interroger sur la perception de l'utilisateur. Bien qu'il soit nécessaire d'avoir des connaissances sur la perception de chaque sens indépendamment des autres, la perception multisensorielle, en s'intéressant à la combinaison de signaux unimodaux, est à étudier. La relation entre les sens, avec un intérêt particulier pour la combinaison de gestes et de sons sera au centre du prochain chapitre.

Perception et association de modalités

L'Homme évolue en permanence dans son environnement grâce à un système de captation de l'information extérieure : son système sensoriel. Ses sens vont lui permettre de collecter les informations d'une situation à un instant t_i , de les analyser, pour ensuite y réagir et recommencer le processus à un instant t_{i+1} . Dans le cadre de cette thèse, nous avons l'ambition de créer une interface homme-machine qui soit interactive, intuitive et facile d'utilisation. Il est donc nécessaire de s'interroger sur la perception des sens humains, pour permettre à la machine de proposer une interaction à l'utilisateur qu'il comprenne facilement et rapidement. Une connaissance des mécanismes de captation des sens mis en jeux est donc nécessaire, mais il est également essentiel de s'interroger sur l'interprétation de l'ensemble des signaux par le système central. Il s'agit donc d'avoir une approche multisensorielle de la perception humaine, pour tenter de comprendre l'association des sens, qui sera tout d'abord développée.

L'approche écologique de la perception, qui a été introduite en partie 1.2.1 du chapitre précédent avec les *auditory icons*, sera ensuite approfondie. Nous aborderons tout d'abord les travaux de Gaver, mettant notamment en évidence une écoute naturelle, s'attachant aux informations des sources sonores. Nous verrons que l'interaction entre objets lors d'un événement s'appuie sur des études de Gaver, se basant sur des phénomènes physiques. Le son, au travers de paramètres invariants, va véhiculer les informations propres à cet événement et l'auditeur va donc pouvoir percevoir et comprendre une situation par ce biais. Les nombreuses études sur ces structures invariantes ont permis la création d'outils de synthèse, permettant de modéliser et de contrôler des sons porteurs de sens. Ces sons vont donc nous intéresser pour illustrer l'objet virtuel que nous proposons d'utiliser.

L'interface que nous ambitionnons de mettre au point est pilotée par l'intermédiaire de gestes, accompagnés d'un retour sonore informatif. Nous allons donc nous focaliser sur des exemples de combinaison de ces modalités, qui ont déjà été étudiées sous certains aspects. Nous verrons qu'il est notamment possible de véhiculer de l'information de mouvement par les propriétés intrinsèques des sons, et que les sons peuvent permettre de guider les gestes. Ces exemples enrichiront notre connaissance de cette interaction, et nous pourrons nous appuyer sur certaines d'entre eux pour comprendre les phénomènes que nous observerons dans les expériences.

Sommaire

2.1	Systèmes perceptifs gestuel et sonore	33
2.1.1	Système sensorimoteur	33
2.1.2	Système auditif humain et localisation des sons	34
2.2	Cas général de l'intégration et d'interactions multisensorielles	38
2.2.1	Une forte influence de la modalité visuelle	39
2.2.2	Modèles de jugements perceptifs	40
2.2.3	Effets attentionnels sur la perception	41
2.3	Approche écologique de la perception auditive	44
2.3.1	Écoutes quotidienne et musicale	44
2.3.2	L'événement comme origine	45
2.3.3	Les invariants perceptifs comme base de l'écoute écologique	45
2.3.4	Synthèse de sons quotidiens	48
2.4	Relations entre mouvements et sons	51
2.4.1	Évocation du mouvement par les sons	51
2.4.2	Interactions sonores et gestuelles	53

2.1 Systèmes perceptifs gestuel et sonore

Nous allons débiter cette section par une brève introduction sur le système perceptif gestuel et sonore chez l'Homme, qu'il est important d'avoir en tête pour la mise au point de notre interface. La proprioception va par exemple permettre de la piloter sans nécessiter une mobilisation de la modalité visuelle. Nous aborderons le système de perception auditif humain, où les mécanismes de captation seront développés, ainsi que les stratégies mises en place pour percevoir la localisation des sons.

2.1.1 Système sensorimoteur

Nous allons ici donner une brève description du système de perception des gestes chez l'Homme. Riemann et Lephart proposent une description détaillée des différents signaux sensorimoteurs [Riemann et Lephart, 2002], en se basant notamment sur des travaux de Sherrington [Sherrington, 1910]. Le système sensorimoteur se réfère aux sensations de douleur, de température, tactiles ainsi que proprioceptives, qui vont tout particulièrement nous intéresser pour la perception gestuelle, voir figure 2.1.

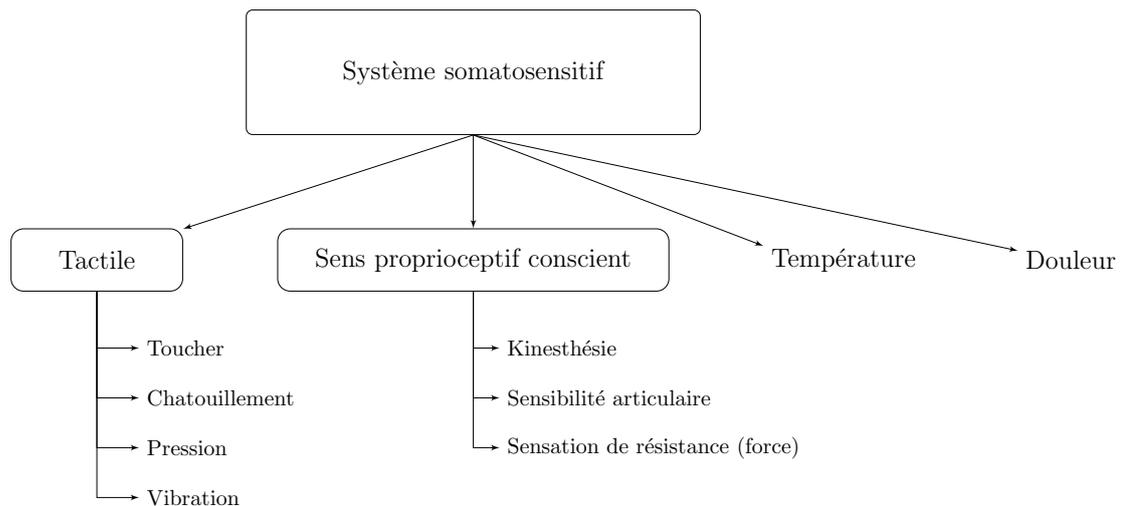


FIGURE 2.1 – Représentation des sensations provenant des sources sensorimotrices, figure adaptée de [Riemann et Lephart, 2002]

La proprioception s'intéresse spécifiquement à la perception de la position relative et des mouvements des différentes parties du corps. Cette sensation provient de mécanorécepteurs disséminés dans tout le corps, et plus particulièrement dans les muscles, les tendons, et les ligaments. Cette sensation est par exemple primordiale pour la reproduction et la coordination de gestes [Gordon *et al.*, 1995] [Sainburg *et al.*, 1993] [Ghez et Sainburg, 1995], bien qu'il soit possible de développer des mécanismes de compensation visuelle pour les personnes dépourvues de proprioception [Sarlegna *et al.*, 2010]. Même si ce système sensoriel est peu connu du grand public, il est particulièrement intéressant dans notre cas puisqu'il permet d'estimer la position des membres, et par exemple de la main, sans en nécessiter un retour visuel.

L'IHM que nous voulons mettre en place sera pilotée par les gestes du conducteur, et les retours sonores viendront confirmer les manipulations effectuées. Il est donc essentiel de se focaliser sur le second système sensoriel qui sera au cœur de cette thèse, permettant la captation de stimuli auditifs, que nous allons maintenant aborder. Nous allons rapidement discuter des mécanismes de captation des sons, avant de passer en revue plusieurs indices caractéristiques sur lesquels se base le système perceptif humain pour déterminer la localisation des sons qu'il perçoit.

2.1.2 Système auditif humain et localisation des sons

L'oreille, représentée sur la figure 2.2, est composée de 3 parties distinctes, ayant chacune un rôle spécifique dans la perception du son. L'oreille externe englobe le pavillon jusqu'au tympan ; l'oreille moyenne est composée du marteau, de l'enclume et de l'étrier ; et l'oreille interne comprend de la cochlée et de l'organe vestibulaire.

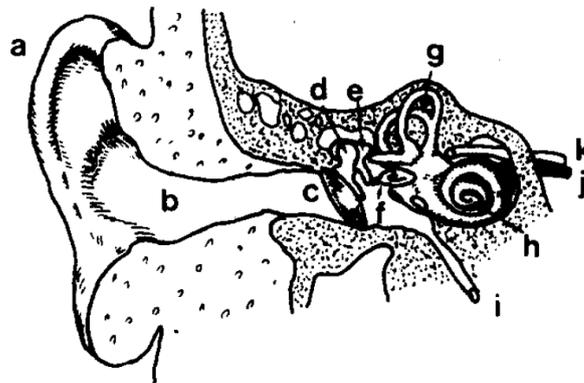


FIGURE 2.2 – Représentation schématique de l'oreille. Figure tirée de [De Boer, 1980]. a : pavillon, b : canal auditif externe, c : membrane tympanique, d : marteau, e : enclume, f : étrier, g : canaux semi-circulaires, h : cochlée, i : tube d'Eustache, j : nerf auditif, k : nerf vestibulaire

L'apport de l'oreille externe dans la perception sonore est double, et provient de sa géométrie particulière. Elle joue le rôle de canal des ondes vers le tympan, grâce à sa forme qui couvre une large zone. Le conduit auditif amplifie également le signal sonore perçu entre 2 et 5 kHz. Le rôle de l'oreille externe est de donner également des informations de localisation – cette partie sera plus développée en section 2.1.2.

L'oreille moyenne fait quant à elle le lien entre le tympan et la fenêtre ovale, frontière avec l'oreille interne. Elle compose la partie mécanique de l'oreille, avec un enchaînement de trois petits os : le marteau, l'enclume et l'étrier. Le rôle majeur de l'oreille moyenne est une adaptation d'impédance entre un milieu aérien et un milieu liquidien.

Enfin, l'oreille interne est composée de deux organes : le système vestibulaire responsable de l'équilibre, et la cochlée. C'est cette dernière qui intervient dans la perception sonore, en convertissant l'information mécanique en sortie de l'oreille moyenne en impulsions électriques qui sont ensuite transmises au cerveau par les nerfs auditifs.

Ces trois parties différentes de l'oreille véhiculent une information captée par l'oreille externe vers le système nerveux central. Le cerveau, en combinant par exemple les signaux provenant des deux oreilles, va pouvoir extraire des informations sur la source des sons, et en particulier sur sa localisation. Nous allons maintenant discuter des différents indices de localisation, qui pourront intervenir pour la mise au point d'un retour sonore de l'interface.

Perception des sons dans l'espace

Les êtres humains sont naturellement immergés dans un environnement en trois dimensions spatiales, y compris d'un point de vue sonore. Au-delà de la nature des signaux sonores captés, il est possible de retirer des informations sur la localisation de la source des sons, qui peuvent par exemple servir au maintien de la posture [Gandemer *et al.*, 2016] [Gandemer, 2016]. Pour une source et un auditeur statique dans une scène, il existe différents indices permettant de localiser la source d'un son. Le premier indice que nous allons aborder concerne la différence interaurale, représentant la différence entre les signaux captés par nos deux oreilles.

Indices interauraux

Il existe deux indices interauraux – ou binauraux : la différence interaurale de niveau – ou ILD pour *Interaural Level Difference* et la différence interaurale de temps – ou ITD pour *Interaural Time Difference*. Ces indices représentent respectivement la différence de niveau et de temps des signaux qui sont captés par les deux oreilles d'un auditeur, comme l'illustre la figure 2.3.

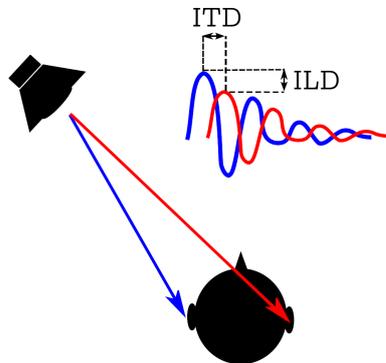


FIGURE 2.3 – Schéma illustrant les différences interaurales de niveau et de temps, figure tirée de [Gandemer, 2016].

Ces indices permettent une première estimation de la localisation d'une source sonore. Cependant, les effets sont variables selon la fréquence des sons. Par exemple, l'ILD est négligeable en dessous de 500 Hz à cause de la trop grande longueur des ondes comparée à la taille de la tête de l'auditeur, provoquant un phénomène d'ombre. En revanche, cet effet est significatif à partir de 1500 Hz et peut aller jusqu'à une différence de 20 dB. En ce qui concerne l'ITD, il existe également deux phénomènes différents selon la plage fréquentielle. En dessous de 1500 Hz, la différence de temps représente la différence de phase des deux signaux, et permet de remonter à un unique azimuth. Cette information est alors fiable. Au-dessus de 1500 Hz, la différence de temps peut correspondre à une ou plusieurs périodes des signaux, il n'est alors plus possible de remonter à un seul azimuth. Dans ce cas, l'ITD correspond à la différence de temps de l'arrivée des deux enveloppes des signaux.

Les deux indices peuvent donc être combinés en fonction de la plage de fréquence du son, pour obtenir une information fiable. Cependant, ces informations ne sont pas suffisantes pour obtenir une unique position de la source dans l'espace. Des informations complémentaires apportées par des indices spectraux vont alors permettre de lever l'ambiguïté.

Indices spectraux

Comme nous venons de l'évoquer, les indices interauraux ne sont pas toujours suffisants pour remonter à la position de la source. Ils permettent en revanche d'isoler une zone où les ILD et ITD sont égaux, zone appelée cône de confusion, voir figure 2.4. Une source peut donc se trouver à plusieurs positions sur ce cône tout en ayant des indices ILD et ITD constants. Il est alors nécessaire d'utiliser d'autres informations pour déterminer l'unique position de la source sonore.

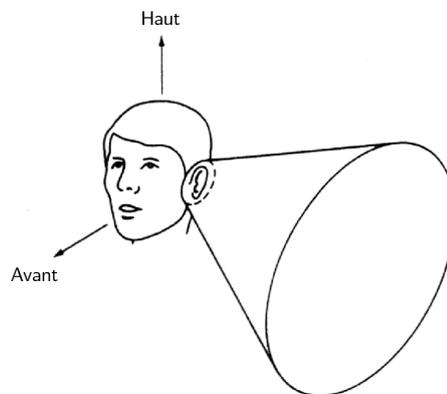


FIGURE 2.4 – Le cône de confusion, figure adaptée de [Blauert, 1997].

Une information supplémentaire vient des interférences constructives et destructives provenant du pavillon de l'oreille, et dans une moindre mesure de la tête et du buste de l'auditeur. Ces interférences sont dépendantes de la direction d'incidence du son, si bien qu'il est possible de remonter à la position de la source sur le cône de confusion. Il est possible de mettre en évidence une fonction de transfert appelée HRTF – pour *Head Related Transfer Function* – entre le signal au niveau du canal auditif externe et celui provenant d'une direction donnée. Les liens entre HRTF et la position de la source dans l'espace ont été appris par les auditeurs : des interférences particulières dans le signal sonore sont naturellement reconnues comme liées à une unique incidence. Il est à noter que ces indices spectraux sont propres à chaque individu, puisqu'ils proviennent directement de la morphologie de l'auditeur.

Perception de la distance

Lorsqu'une onde sonore se propage dans l'air, son intensité décroît avec la distance. En effet, l'équation de propagation donne une amplitude dépendante en $1/d^2$ en notant d la distance entre la source et l'auditeur. L'intensité de niveau sonore est donc diminuée de 6 dB à chaque fois que la distance est doublée.

Il est également possible d'avoir une estimation de la distance à la source en faisant la distinction entre le champ direct, allant directement de la source à l'oreille de l'auditeur ; et le champ réverbéré qui atteint l'auditeur après une ou plus réflexion sur des objets de l'environnement. Le rapport entre ces deux champs est décroissant avec la distance, ce qui peut renseigner sur ce paramètre.

Enfin, le spectre d'un son est modifié avec la distance : les hautes fréquences ont tendance à être atténuée par l'absorption de l'air pour une faible distance relativement aux basses fréquences. Ainsi, un auditeur peut estimer sa distance à une source sonore par le contenu spectral d'un son : s'il ne perçoit que des basses fréquences, la source est probablement éloignée. Plus il se rapprochera de la source, et plus le son s'enrichira en hautes fréquences.

Effet Doppler

Nous nous sommes pour l'instant concentrés sur la localisation de sources fixes. L'effet Doppler concerne en revanche les sources mobiles, et peut même être perçu avec une seule oreille. Il provoque un décalage fréquentiel du son selon un facteur :

$$\frac{1 + v/c}{1 - v/c}$$

où v représente la vitesse de la source et c la vitesse du son. Le décalage est donc croissant avec la vitesse de la source de son. Cet effet bien connu semble en revanche ne donner que peu d'informations concernant la perception de la distance [Rosenblum *et al.*, 1987]. La direction du décalage – vers les fréquences hautes ou basses – donne en revanche le sens du mouvement de la source.

Nous avons abordé ici les différentes stratégies mises en place pour localiser la source d'un son, où les indices binauraux et spectraux sont déterminants. De plus, il est possible d'obtenir une meilleure estimation de la position de la source grâce à des mouvements de tête de l'auditeur, même si la source sonore est immobile [Wallach, 1939] [Thurlow et Runge, 1967]. Il est important d'avoir ces connaissances à l'esprit pour la suite de la thèse, qui peuvent être un bon moyen pour véhiculer des informations spatiales.

Après avoir abordé les systèmes perceptifs des gestes et des sons individuellement, nous allons nous intéresser à l'interprétation de signaux provenant simultanément de plusieurs modalités dans un cas général. Nous pourrions ainsi essayer de comprendre les mécanismes du système perceptif humain, où nous verrons que la modalité visuelle semble particulièrement importante. L'influence de l'attention sur la perception sera également discutée, dans le but de comprendre comment les signaux liés aux manipulations de notre interface seront perçus lorsque l'utilisateur devra conjointement accomplir une tâche de conduite.

2.2 Cas général de l'intégration et d'interactions multisensorielles

Nous utilisons en permanence des informations provenant de nos différents sens, pour nous renseigner sur l'état de l'environnement extérieur. Il nous est alors possible d'apprécier certains événements de plusieurs manières : une balle qui roule sur une table peut être perçue par la vue, où des caractéristiques comme sa position, sa vitesse, sa taille ou sa couleur sont accessibles ; ou par l'audition via le son du roulement. Si l'objectif est d'attraper cette balle, ces informations peuvent être combinées, en prenant également en compte la sensation tactile qui indique le moment où la balle atteint notre main. Il est alors possible de réagir rapidement en fermant la main pour attraper la balle. Cet exemple montre comment les sens agissent en coopération pour percevoir l'environnement, afin d'accomplir une tâche précise.

L'apport de la redondance d'informations provenant de plusieurs sens a été mesuré dans la littérature à travers l'exécution de tâches. Il a par exemple été montré à plusieurs reprises qu'une sollicitation conjointe de la vue et de l'audition permet de réduire un temps de réaction, comparé aux cas unimodaux [Hershenson, 1962] [Doyle et Snowden, 2001]. De même, dans une étude menée par Giard et Peronnet, les sujets devaient reconnaître deux objets de formes différentes, parfois accompagnés de sons brefs de hauteurs différentes. Il a alors été constaté que le temps d'identification était significativement plus faible dans le cas bimodal comparé aux cas visuels ou auditifs seuls [Giard et Peronnet, 1999]. Cet apport de la combinaison de modalités est également observable avec des stimuli auditifs-tactiles [Murray *et al.*, 2005] [Kjær *et al.*, 2007] ou olfactifs-visuels [Gottfried et Dolan, 2003]. L'utilisation conjointe de la vue, de l'ouïe et du toucher permet également de supplanter les cas bimodaux [Diederich et Colonius, 2004].

Cette redondance de signaux permet donc une perception plus robuste, en combinant des indices provenant de différentes modalités. Elle permet également la résolution d'ambiguïtés [Merfeld *et al.*, 1999]. Green et Angelaki abordent par exemple le problème de la perception de l'accélération [Green et Angelaki, 2010]. Les organes dits « otolithiques » situés dans l'oreille interne sont chargés de capter cette information. La situation du point de vue de ces organes peut être ambiguë : l'accélération perçue peut être égale lors d'une translation latérale, ou d'une rotation de la tête. Pour lever cette ambiguïté, l'apport d'informations provenant des canaux semi-circulaires [Green et Angelaki, 2003] ou visuelles [Zupan et Merfeld, 2003] est alors nécessaire. Un exemple bien connu concerne le départ d'un train alors que nous sommes à l'intérieur : la vue du train de la voie d'en face indique un mouvement, sans préciser quel train se déplace. Un coup d'œil au quai permet alors de lever cette ambiguïté.

L'information provenant d'une modalité peut donc confirmer d'autres sensations et permettre une perception robuste, ou lever l'ambiguïté d'une situation. Cependant lorsqu'une perception multisensorielle entraîne des signaux contradictoires, certains sens semblent privilégiés. Ce phénomène témoigne de poids différents appliqués aux informations portées par chaque modalité. Par exemple, plusieurs études montrent une mise en avant particulière de la modalité visuelle, qui peut prendre le pas sur les autres sens.

2.2.1 Une forte influence de la modalité visuelle

Un premier exemple qui illustre bien l'influence du visuel sur la modalité sonore est l'effet « ventriloque », où un ventriloque parle en faisant bouger le moins possible ses lèvres. Il utilise conjointement une marionnette avec des mouvements de bouches synchronisés à ses paroles. L'audience du ventriloque a alors la sensation que la marionnette parle, par la combinaison d'informations visuelles et sonores résultant en une mauvaise localisation de la source sonore [Bermant et Welch, 1976]. Soto-Faraco *et al.* ont également étudié cet effet « ventriloque », mais d'un point de vue dynamique, et ils ont montré que des signaux visuels avaient tendance à perturber la perception de sources sonores en mouvement [Soto-Faraco *et al.*, 2002]. L'interaction entre les modalités sonore et visuelle est également soulignée avec l'effet McGurk, où il a été montré qu'une confusion dans la perception de la parole peut avoir lieu [MacDonald et McGurk, 1978]. Par exemple, les sujets ayant un stimulus visuel correspondant au son /ta/ présenté conjointement à un son /ba/ vont avoir tendance à comprendre la syllabe /da/ lorsqu'il leur est demandé de renseigner ce qu'ils ont entendu. De Gelder a également montré que le visuel pouvait perturber la perception de l'émotion dans la parole, dans un effet surnommé « McGurk émotionnel » [De Gelder et Vroomen, 2000]. Ces exemples démontrent la prédominance du visuel – appelée « capture visuelle » – sur la modalité sonore, mais il a également été constaté des effets du visuel sur le toucher [Pavani *et al.*, 2000], le goût [DuBose *et al.*, 1980] ou la proprioception [Hay *et al.*, 1965].

Cette domination n'est cependant pas systématique : Shams a montré que le son pouvait significativement influencer la modalité visuelle. Dans son expérience, un unique flash lumineux est envoyé aux sujets, conjointement à un ou plusieurs sons brefs. Le nombre de flashes perçus a alors tendance à augmenter avec le nombre de sons qui l'accompagne [Shams *et al.*, 2000] [Shams *et al.*, 2001] [Shams *et al.*, 2002]. La perception d'une situation visuelle ambiguë peut également être influencée par le son : dans leur expérience, Watanabe et Shimojo sont partis d'une situation où deux cibles visuelles se rapprochant peuvent être interprétées comme se traversant ou rebondissant [Watanabe et Shimojo, 2001]. L'ajout d'un son lorsque les deux cibles coïncident a alors permis d'orienter la perception des participants vers un effet de rebond. Plus récemment, Thoret a montré qu'il existe une forte interaction entre les modalités sensorimotrices, visuelles et auditives, où des retours sonores peuvent être aussi perturbateurs que des stimuli visuels [Thoret, 2014]. Cette étude montrant des interactions entre gestes et sons sera détaillée plus précisément en section 2.4.

Plusieurs signaux sensoriels sont donc intégrés pour produire une perception de la situation robuste et cohérente, même si plusieurs effets de capture peuvent avoir lieu. Comme nous l'avons vu précédemment, la prédominance d'une modalité sur une autre dépend fortement des conditions de présentation : dans plusieurs cas la modalité auditive est capturée par la vue, mais Shams a montré que des sons pouvaient perturber la perception visuelle [Shams *et al.*, 2000]. Dans le cas des manipulations de l'interface gestuelle, il sera nécessaire de s'interroger sur l'intégration de stimuli sonores s'ils sont présentés simultanément à des signaux visuels.

Plusieurs tentatives ont alors cherché à modéliser la perception globale d'une situation, pour expliquer l'influence de ces conditions.

2.2.2 Modèles de jugements perceptifs

Plusieurs modèles ont été mis au point pour expliquer les jugements perceptifs à partir des signaux potentiellement conflictuels mis en jeu. Une première explication provenant de Welch et Warren énonce que la modalité « la plus précise » dans une situation donnée va prendre le pas sur toutes les autres [Welch *et al.*, 1986] [Welch et Warren, 1986]. Ce modèle se base sur une hypothèse selon laquelle une modalité a un domaine de prédilection par rapport aux autres. Par exemple, l'audition a tendance à prendre le pas sur le système visuel dans les tâches de précision temporelle [Welch *et al.*, 1986], alors que la vision excelle dans les tâches de localisation dans l'espace. Cette hypothèse sous-entend que la hiérarchie utilisée pour estimer la perception la plus robuste n'est pas fixe, et change selon la situation.

Une seconde explication propose que le jugement se base sur une multitude d'informations provenant de différentes modalités. Selon ce modèle, pour obtenir la perception la plus fiable possible en intégrant de manière efficace l'ensemble des sources d'informations sensorielles, notre système perceptif doit tendre à minimiser la variance de l'estimation perceptive globale. Pour obtenir ce résultat, tout en supposant que le bruit provenant des signaux perceptifs est gaussien, le modèle approprié est celui du maximum de vraisemblance – ou MLE pour *Maximum Likelihood Estimation* [Blake *et al.*, 1993] [Ernst et Bühlhoff, 2004] [Clark et Yuille, 2013]. Cette modélisation suppose que l'estimation de l'intégration \hat{s} est une somme pondérée des estimations dans les différentes modalités \hat{s}_i , avec des poids ω_i proportionnels à l'inverse de leur variance au carré σ_i^2 :

$$\hat{s} = \sum_i \omega_i \hat{s}_i \quad \text{avec} \quad \sum_i \omega_i = 1 \quad \text{et} \quad \omega_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$$

Avec ce modèle, les exemples précédents peuvent être expliqués par une pondération selon la fiabilité des informations sensorielles grâce aux ω_i . Plusieurs études ont appuyé la pertinence de ce modèle [Young *et al.*, 1993] [Johnston *et al.*, 1994] [Wuerger *et al.*, 2003]. Le premier modèle peut alors être considéré comme un cas particulier où le poids des informations d'une modalité est prédominant, celle-ci étant jugée plus fiable. Les deux modélisations ont été jugées partiellement correctes dans une étude sur la localisation spatiale grâce à des signaux visuels ou auditifs menée par Battaglia *et al.* [Battaglia *et al.*, 2003]. Plus particulièrement lorsque les signaux visuels sont dégradés, les sujets ont tendance à donner plus de poids aux informations auditives. Le modèle MLE prédit correctement cette tendance même si le poids donné au visuel est sous-estimé. Un hybride entre les deux modèles proposés pourrait alors expliquer au mieux les résultats. Il est alors calculé un modèle d'intégration bayésien modifiant le modèle MLE pour tenir compte des expériences postérieures, qui explique mieux les résultats de l'expérience que le modèle MLE, en donnant un poids particulier au visuel [Battaglia *et al.*, 2003].

Il est possible de résumer les grandes étapes des traitements des informations perceptives conduisant à une action par la figure 2.5, inspirée par des figures de Rao *et al.* [Rao *et al.*, 2002] et Ernst et Bühlhoff [Ernst et Bühlhoff, 2004]. Une fonction de gain est ajoutée à la chaîne présentée précédemment, pour prendre en compte le but recherché par l'utilisateur.

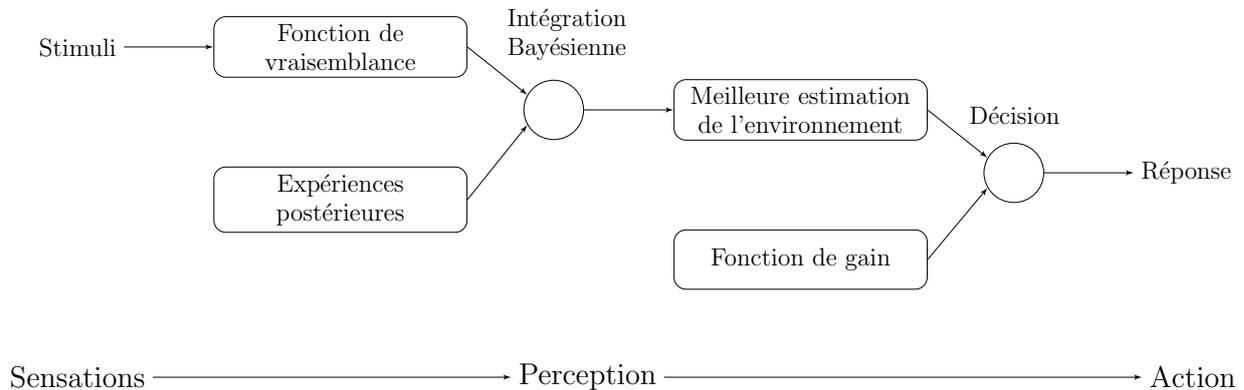


FIGURE 2.5 – Grandes étapes du traitement des informations perceptives menant à une action, selon Ernst et Bühlhoff [Ernst et Bühlhoff, 2004] et Rao *et al.* [Rao *et al.*, 2002].

Nous avons donc abordé l'intégration des stimuli pour une perception robuste et stable, pouvant mener à une prise de décision. Dans ce processus, la perception des différentes modalités sensorielles est primordiale, et il est nécessaire de prendre en compte les différents effets qui peuvent alors se produire dans l'interprétation des signaux sensoriels. Plusieurs modèles ont été proposés pour simuler ces phénomènes, et le modèle d'intégration bayésienne semble expliquer correctement l'intégration multisensorielle.

Cependant, le contexte automobile de cette thèse amène à prendre en compte un autre aspect de la perception. L'interface gestuelle que nous voulons mettre en place doit être utilisable alors que l'utilisateur effectue une tâche de conduite, qui va lui demander une certaine concentration. Il est alors important de prendre en compte l'effet de l'attention sur la perception, pour que le conducteur puisse percevoir convenablement les signaux émis par l'interface.

2.2.3 Effets attentionnels sur la perception

Nous avons tous eu l'expérience de réfléchir intensément à quelque chose alors qu'une personne est en train de nous parler. Même si les ondes sonores sont bien captées par nos tympans, nous ne saisissons pas le sens des phrases prononcées. À l'inverse, il est fréquent d'avoir l'impression de percevoir des signaux qui n'existent pas : les pas d'une personne que nous attendons impatiemment par exemple, où le moindre son sera interprété comme un bruit de pas. Ces deux exemples semblent bien indiquer un effet de l'attention sur la perception. Le premier cas illustre un manque d'attention alors le second se rapproche d'illusions. Il semble donc que l'attention est liée à un ralentissement voire à un blocage du traitement de l'information dans certains cas, et à une accélération dans d'autres. Cette section traitera plus particulièrement de la dégradation de la perception liée à des effets attentionnels, alors que la partie 3.1 abordera les questions de concentration.

La « cécité attentionnelle » est un phénomène qui a lieu lorsque qu'un stimulus n'est pas perçu par un individu lors de l'accomplissement d'une tâche, alors qu'il est tout à fait détectable. Cette découverte provoque des questionnements par rapport à l'implication de l'attention dans la perception, Mack et Rock allant jusqu'à estimer qu'il « n'y a pas de perception sans attention » [Mack et Rock, 1999]. En 1979, Neisser étudie par exemple cet effet au sein de la modalité visuelle. Dans l'une de ses expériences, une vidéo est diffusée et il propose à des sujets de compter le nombre de fois que deux équipes se passent une balle. Après avoir donné les résultats, seuls 21% des sujets ont

remarqué qu'une femme est passée au milieu de la scène avec un parapluie ouvert, étant néanmoins clairement visible [Neisser, 1979]. Simons et Chabris confirment ces résultats dans une expérience similaire, où deux équipes habillées en blanc ou en noir, se passent une balle. Cette fois, c'est un homme portant un costume noir représentant un gorille qui traverse la scène [Simons et Chabris, 1999]. Les résultats montrent que ce phénomène de cécité attentionnel semble dépendant de la tâche demandée – compter les passes de l'équipe en noir ou en blanc. Les sujets ont alors focalisé leur attention sur une couleur, et les autres stimuli visuels semblent ignorés. Mack et Rock ont également étudié l'attention visuelle, et sont arrivés à un constat similaire de cécité attentionnelle : lorsque des stimuli visuels inattendus sont présentés à des sujets concentrés sur une autre tâche visuelle, ils sont alors ignorés [Mack et Rock, 1999]. Dans le domaine auditif, Cherry propose des expériences similaires avec des écoutes dichotiques, où les sujets doivent répéter le message diffusé dans une oreille, et ignorer le message restitué dans l'autre [Cherry, 1953]. Les performances sont excellentes, mais les sujets ne sont pas capables de donner le contenu de l'autre message diffusé. Moray a poursuivi les travaux de Cherry par des signaux perturbateurs très simples, avec des mots répétés plusieurs fois. Les résultats montrent que les seuls mots « passant la barrière attentionnelle », c'est-à-dire qui attirent l'attention des sujets, sont leurs propres prénoms [Moray, 1959].

Ces effets attentionnels se retrouvent également lors de la présentation de signaux multisensoriels. Plusieurs études se sont intéressées à l'attention en se basant sur des temps de réaction. Spence et Driver ont par exemple montré qu'il était possible de diriger l'attention vers certaines modalités : des temps de réponses sont systématiquement plus faibles lorsqu'un stimulus est présenté dans la modalité indiquée, par rapport à une présentation dans une modalité inattendue [Spence et Driver, 1997]. Les sujets ont donc élaboré une stratégie visant à prioriser une modalité par rapport aux autres. Dans une étude publiée en 1976, Posner suppose que la modalité visuelle mobilise une attention supérieure à l'audition, à cause d'un manque de réactivité de la modalité visuelle [Posner *et al.*, 1976]. Cette supposition est appuyée par une série d'expériences montrant la tendance à ne pas percevoir occasionnellement les stimuli auditifs lors de sollicitations audio-visuelles simultanées, phénomène appelé « effet Colavita » [Colavita, 1974] [Spence, 2009] [Koppen et Spence, 2007] [Koppen *et al.*, 2009]. L'interaction de plusieurs modalités et de l'attention est également au centre d'une expérience de Pizzighello et Bressan, qui étudie l'exécution de tâches visuelles ou auditives sur la détection d'un stimulus visuel [Pizzighello et Bressan, 2008]. Les résultats montrent qu'il existe un phénomène de « désengagement » lors d'une tâche purement auditive : les sujets ne tiennent pas compte des signaux visuels dans ce cas, en se concentrant uniquement sur le son. Une conclusion principale est à tirer de cette étude selon leurs auteurs : pour faire attention à un stimulus sonore, les sujets retirent automatiquement de l'attention à une scène visuelle. Pizzighello et Bressan rapprochent ces résultats d'une étude menée par McEvoy *et al.*, liant l'utilisation d'un téléphone au volant avec des accidents, même avec des kits mains-libres [McEvoy *et al.*, 2005].

Ces études montrent un effet important de l'attention sur la perception qu'elle soit unimodale ou multimodale. Pour la mise au point de notre interface gestuelle, il est important de garder en tête que la concentration mobilisée pour exécuter les gestes et percevoir les sons ne doit pas se répercuter sur la conduite. En proposant des modalités d'interaction différentes, notre interface peut alors avoir un avantage par rapport aux interfaces tactiles. En effet, il semble qu'il soit possible de diviser efficacement son attention entre les modalités visuelles et auditives. Une étude de Duncan propose de détecter deux stimuli auditifs et/ou visuels [Duncan *et al.*, 1997]. Des flux de sons et/ou de stimuli visuels contenant des cibles visuelles ou auditives à détecter visibles en figure 2.6 sont présentés à des sujets.

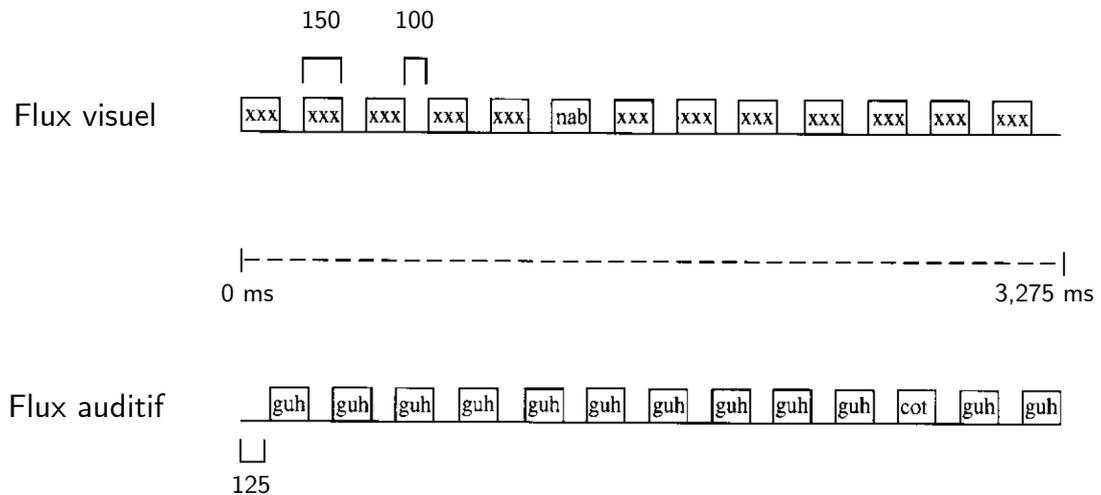


FIGURE 2.6 – Stimuli visuels et auditifs présentés dans le cas audio-visuel de l'étude de Duncan [Duncan *et al.*, 1997], dont la figure est adaptée. Visuellement les sujets doivent reconnaître les lettres « nab » parmi une suite de « xxx », alors qu'ils doivent détecter le son de la syllabe « cot » parmi une suite de son de la syllabe « guh » au niveau des stimuli auditifs.

Dans les cas auditifs ou visuels purs, une baisse de performance est observée lorsque les deux cibles sont présentées simultanément. Cet effet n'a en revanche pas lieu pour une présentation audio-visuelle simultanée des cibles. Diviser l'attention d'un conducteur entre le visuel pour la conduite, et le gestuel lié au sonore pour l'interaction avec l'interface peut donc être bénéfique par rapport à une situation de concentration des sollicitations sur la modalité visuelle.

Comme nous l'avons constaté, l'attention est un facteur déterminant de la perception, et son intégration au sein du système perceptif est encore débattue. Ses effets dépendent de la tâche proposée, des modalités sollicitées, mais également de la concentration des utilisateurs. Certains aspects seront développés en partie 3.1, dont par exemple l'exécution de tâches simultanées que nous avons brièvement abordée au travers de plusieurs exemples. La division de l'attention entre plusieurs modalités semble être une piste intéressante que nous explorerons dans la suite de la thèse.

Après avoir abordé la perception et l'intégration multisensorielle, ponctuées par des modèles et exemples tirés de différentes études, nous allons à présent nous intéresser à une théorie d'interprétation des sons, développée par Gaver et que nous avons évoquée précédemment. Cette théorie propose un angle de vue particulier, centré sur l'interprétation des sons, et elle introduit un type de perception qualifiée de « directe » par Michaels et Carello [Michaels et Carello, 1981]. Les concepts liés à cette approche écologique vont donc être développés, ce qui va nous conduire aux mécanismes de construction de synthétiseurs de sons naturels.

2.3 Approche écologique de la perception auditive

Imaginons que vous vous trouvez aux abords d'une route. Lorsqu'un véhicule s'approche, il vous est possible d'avoir un certain nombre d'informations simplement en écoutant le son émis. Par exemple, il est possible de dire si le véhicule est plutôt un camion ou une voiture, s'il est d'une petite ou une grosse cylindrée, et s'il approche à vive allure ou à vitesse plus modérée. Cet exemple tiré d'un article de Gaver [Gaver, 1993b] montre bien que les auditeurs se basent en premier lieu sur une écoute du son s'attachant aux sens véhiculés par les sons, plutôt qu'aux caractéristiques acoustiques. Cette écoute est directe, ne nécessitant pas d'analyse des paramètres sonores, et est tirée de l'approche écologique de la perception proposée par Gibson pour la modalité visuelle [Gibson, 1961] [Gibson, 1979]. Gaver distingue donc deux manières d'écouter les sons [Gaver, 1993b].

2.3.1 Écoutes quotidienne et musicale

L'écoute quotidienne ou écologique – *everyday listening* – concerne une écoute du son directe, permettant de retirer de l'information d'un son sur l'événement qui l'engendre. Ce type d'écoute s'intéresse au pouvoir évocatoire du son. Dans l'exemple précédent, le bruit de moteur d'un véhicule peut évoquer une petite voiture, une grosse cylindrée ou encore un camion. C'est ce lien entre son et information concrète qui est caractéristique de l'écoute écologique.

Le deuxième type d'écoute mis en avant par Gaver est l'écoute analytique ou musicale – *musical listening*, elle s'intéresse aux propriétés acoustiques d'un son : sa fréquence, son intensité, son attaque, ou sa durée par exemple. Cette écoute est analytique, et elle est notamment utilisée en création musicale où les paramètres acoustiques sont manipulés, ce qui en justifie le nom.

Ces deux types d'écoute peuvent être mis en application sur un même son. En reprenant l'exemple précédent du son émis par un véhicule circulant sur une route, il est possible d'avoir une estimation de la cylindrée d'une part ; mais l'écoute musicale permet de percevoir les changements de la fréquence fondamentale du bruit moteur d'autre part. Une expérience intéressante a été mise en œuvre par Vanderveer, où il est demandé à des sujets de décrire brièvement les sons qui leur étaient diffusés, représentant des sons du quotidien [VanDerveer, 1979]. Les résultats indiquent que le langage utilisé correspond à une description de la source ou de l'événement générateur du son, sauf dans les cas où la source n'est pas identifiée. Dans ces cas là, les sujets se focalisent alors sur les propriétés du son. Cette étude semble indiquer que l'écoute écologique est véritablement une écoute naturelle, utilisée en premier lieu, alors que l'écoute musicale vient dans un second temps lorsque la source n'est pas explicable par un événement. Lemaitre *et al.* ont également étudié l'emploi de ces deux écoutes pour classifier des sons [Lemaitre *et al.*, 2010]. Il est intéressant de noter que le degré d'expertise semble jouer un rôle dans le choix de l'écoute : les experts ont tendance à employer le vocabulaire de l'écoute musicale, alors que les non-experts se tournent instinctivement vers l'écoute écologique.

2.3.2 L'événement comme origine

L'écoute écologique de Gaver s'intéresse à la source du son et donc aux événements et aux phénomènes physiques qui y sont liés, plutôt qu'aux propriétés du signal sonore. Pour en revenir à notre exemple, les explosions qui interviennent dans les cylindres sont les événements à l'origine du bruit de moteur, dont une partie de l'énergie est rayonnée sous forme sonore. Il peut alors être intéressant pour les constructeurs automobiles de s'interroger sur les caractéristiques de ces explosions pour qu'un moteur soit perçu de la façon désirée par exemple.

Gaver insiste également sur le fait que chaque événement à l'origine d'un son de l'environnement provient d'une « interaction de matériaux » [Gaver, 1993b]. Les variations de pression provenant de cette interaction, et donc le son produit, dépendent fortement des caractéristiques physiques des deux éléments : taille, forme, ou encore leur composition. De la même manière, l'interaction va influencer les caractéristiques du son produit. Le son résultant peut donc informer l'auditeur sur l'ensemble des caractéristiques d'une interaction. Une étude vient également appuyer la théorie de Gaver plaçant l'événement au centre de la perception : Kohler *et al.* ont découvert des neurones chez le singe qui semblent liés à une action. En effet, ils se déchargent indifféremment si une action est accomplie par le singe, s'il entend le son résultant de cette action, ou encore s'il voit l'action se dérouler [Kohler *et al.*, 2002]. Ces neurones miroirs « audio-visuels » semblent donc coder le contenu d'une action, ce qui semble bien être un concept important de la perception. Ces résultats montrent une imbrication entre la réalisation d'une action, et sa perception.

D'autre part, le son produit va se propager dans un milieu avant d'atteindre les oreilles de l'auditeur. Ce milieu va également influencer le son sur certaines caractéristiques, la plus évidente pouvant être la réverbération. En effet, faire rouler une bille en métal sur un plateau en bois dans une petite salle confinée ou dans une cathédrale ne va pas modifier l'interaction et le son produit. En revanche, la propagation du son dans le deuxième cas de figure va s'effectuer dans un lieu beaucoup plus grand et réverbérant que dans le premier cas.

Selon Gaver, cette théorie ouvre un champ d'étude en acoustique basée sur un contrôle haut-niveau de processus de synthèse sonore à partir de descriptifs d'une source. Nous pouvons alors formuler plusieurs questions : comment caractériser un son généré par l'interaction d'objets en bois ou en métal ? De la même façon, quelles sont les caractéristiques d'un son identifié comme du roulement ou du frottement ? Nous allons à présent nous intéresser aux informations permettant d'identifier ces propriétés des sources, les invariants perceptifs.

2.3.3 Les invariants perceptifs comme base de l'écoute écologique

Gibson, qui s'intéresse à la modalité visuelle, qualifie certaines propriétés du flux visuel de *variantes* et d'autres d'*invariantes* [Gibson, 1961] [Gibson, 1979]. Pour illustrer ce propos, imaginons le flux provenant d'une observation d'un crocodile. Le déplacement de l'observateur dans la scène va entraîner des changements dans le flux capté par la rétine, qui va faire varier certaines propriétés. En revanche, le crocodile sera toujours reconnu car certaines propriétés restent invariantes selon le point de vue – la forme de l'animal, sa couleur par exemple. Ce terme d'invariant peut être transposé aux signaux sonores, pour désigner les propriétés des sources sonores perçues même après une transformation du signal. Par exemple, il est toujours possible d'identifier une voix connue même si elle est enrôlée ou dégradée par une transmission par un téléphone.

Dans le domaine sonore, Michaels et Carello, repris ensuite par McAdams, introduisent les termes d'invariants *structuraux* et d'invariants *transformationnels* [Michaels et Carello, 1981] [McAdams, 1993]. Les premiers se réfèrent aux propriétés physiques de l'objet source du son – sa forme, sa taille ou son matériau par exemple, alors que les seconds concernent l'action exercée sur l'objet. McAdams prend un exemple parlant pour illustrer ces deux catégories d'invariants : le son produit par un violon peut être reconnu qu'il soit un Stradivarius ou d'une marque quelconque, que le musicien joue dans un registre grave ou aigu car il véhicule des invariants *structuraux*. De la même manière, il est aisé de reconnaître une corde pincée, que ce soit sur un clavecin, une guitare ou un violon, cette propriété est véhiculée par un invariant *transformationnel* [McAdams, 1993].

Les invariants structuraux propres aux objets

De nombreuses études portant sur les invariants structuraux se sont concentrées sur l'étude des sons d'impacts d'un objet sur un autre, permettant de se concentrer sur les propriétés de l'objet impacté [Gaver et Norman, 1988] [Freed, 1990] [Klatzky *et al.*, 2000] [Avanzini et Rocchesso, 2001] [Aramaki et Kronland-Martinet, 2006] [Lutfi et Liu, 2007] [Giordano *et al.*, 2010], ce qui a permis d'étudier les bruits de fermeture de portes dans le domaine automobile [Bezât *et al.*, 2014]. Gaver a commencé à étudier les propriétés transmises par les sons de barres de différentes longueurs et différents matériaux [Gaver et Norman, 1988]. Il a été observé que le jugement du matériau entre des barres en bois ou en métal était simple pour les sujets, contrairement au jugement de la longueur qui s'est révélé beaucoup plus problématique. Depuis, des liens ont pu être tissés entre des propriétés acoustiques et des objets. Par exemple, la forme d'objets impactés peut être discriminée par certaines composantes spectrales [Lakatos *et al.*, 1997], et la perception du matériau est principalement influencée par la décroissance temporelle du son, qui dépend de la fréquence [Roussarie, 1999] [Klatzky *et al.*, 2000] [Avanzini et Rocchesso, 2001] [Giordano et McAdams, 2006] [Avanzini et Rocchesso, 2001] [Giordano et McAdams, 2006].

Les invariants transformationnels propres aux actions

Les invariants transformationnels concernent quant à eux les actions propres aux interactions entre objets [McAdams, 1993]. Warren et Verbrugge ont par exemple étudié le rythme d'une série d'impulsions sonores ayant un même contenu spectral, et ont prouvé qu'il était possible de savoir si une bouteille se brise ou rebondit uniquement par cette information temporelle [Warren et Verbrugge, 1984]. Un autre exemple connu concerne le remplissage d'un récipient cylindrique par de l'eau [Cabe et Pittenger, 2000]. La fréquence de résonance du récipient augmentant avec le volume d'eau versé à l'intérieur, il est possible d'anticiper le remplissage total du récipient simplement avec l'information auditive. Une étude de Repp montre qu'il est également possible d'avoir une information sur la position de la main lors d'un applaudissement uniquement grâce à certaines composantes spectrales du son [Repp, 1987]. Plus récemment, Thoret a étudié le lien qu'il pouvait exister entre mouvements biologiques et perception auditive [Thoret, 2014]. Il a été montré que la vitesse tangentielle lors d'un geste d'écriture est l'invariant transformationnel qui permet de distinguer des formes géométriques. Il est donc possible d'associer facilement des sons provenant d'un tracé à sa forme géométrique. Conan a quant à lui mis en exergue un invariant de la perception de roulement, indépendant des propriétés des objets [Conan, 2014]. Dans le domaine automobile, Bezât a montré le lien entre le geste de fermeture d'une portière et le son engendré lors du claquement de la porte [Bezât, 2007].

Il est également intéressant de noter le lien entre invariants structuraux et transformationnels. Ainsi Houben s'est intéressé aux informations véhiculées par le son de billes roulant sur un plan [Houben, 2002]. Dans un expérience de comparaisons de sons par paires, Houben montre qu'il est possible de discriminer la vitesse ou la taille des billes au travers de sons de roulement [Houben *et al.*, 2004]. Les sujets semblent se baser sur des indices spectraux, et une correspondance a été mise en avant entre un « barycentre d'intensité spécifique » d'une part – ou *centroid of specific loudness* – et la taille ou la vitesse de billes de l'autre, voir figure 2.7.

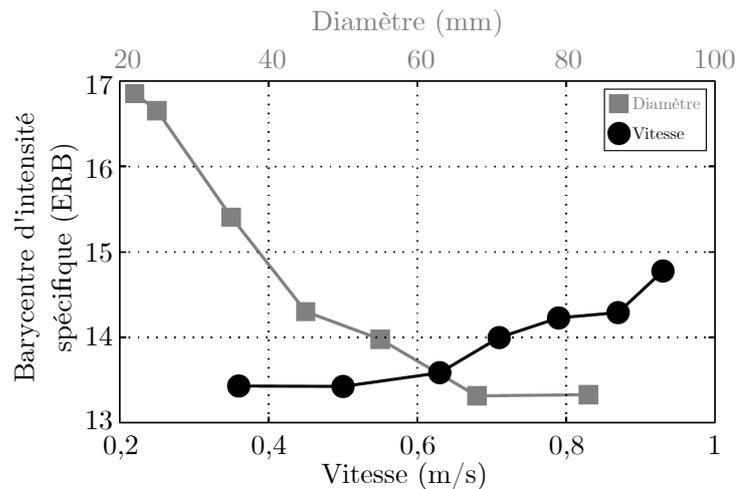


FIGURE 2.7 – Liens entre le barycentre d'intensité spécifique et la taille ou la vitesse de billes roulantes, figure tirée de [Houben *et al.*, 2004].

Houben a même cherché à expliquer d'où venaient les indices auditifs pour discriminer les sons de billes roulantes, en combinant les caractéristiques spectrales d'un premier son avec les caractéristiques temporelles d'un second. Ces sons ont ensuite été utilisés dans un test perceptif de comparaison par paires. Les résultats montrent que les sujets ont tendance à privilégier les indices spectraux pour juger de la vitesse et de la taille des billes.

Les invariants sont donc des morphologies sonores permettant de véhiculer les informations nécessaires pour reconnaître les événements à l'origine des sons. Ces morphologies peuvent être particulièrement intéressantes dans le cadre de ce travail de thèse, puisque nous voulons illustrer un objet virtuel invisible par un retour sonore informatif. La synthèse de sons quotidiens est donc un outil intéressant pour cette étude. Il faut cependant veiller à bien transmettre les invariants pour provoquer l'évocation voulue. La méthodologie de l'analyse – synthèse est un outil important dans la synthèse de sons de l'environnement, et permet de s'assurer de la conservation des invariants. De plus, cette méthodologie a progressivement permis de relier des paramètres bas-niveau, aux paramètres de contrôle qui décrivent la source sonore. Cette technique a été utilisée à de nombreuses reprises pour créer des sons synthétiques, et pour arriver à isoler certains attributs des signaux sonores. Nous allons donc nous intéresser à cette technique de création de sons du quotidien, pour ensuite présenter un outil de synthèse.

2.3.4 Synthèse de sons quotidiens

La méthode d'analyse – synthèse est évoquée par Risset et Wessel avant Gaver, où elle est appliquée à la synthèse de sons imitant différents timbres [Risset et Wessel, 1999]. Dans ce cas, l'objectif est que les sons recréés ressemblent au timbre du son étudié. La figure 2.8, adaptée de [Risset et Wessel, 1999], illustre ce principe sur une itération. Cette technique consiste en un processus basé tout d'abord sur une analyse d'un signal sonore – point ① de la figure 2.8. Selon Gaver, l'analyse doit se faire à 3 niveaux différents : la physique de l'événement, la manière dont elle impacte le son, et comment elle influence la perception de l'événement [Gaver, 1993a]. Il n'est cependant pas systématiquement possible de se baser sur la physique de l'événement pour synthétiser un son de l'environnement, celle-ci étant parfois trop complexe, par exemple pour recréer le son d'un feu ou des turbulences du vent par exemple. Dans ces cas, un modèle de signaux basé sur l'événement peut être proposé. Dans le cas du feu, trois sons principaux sont par exemple isolés puis modélisés : les crépitements, les sifflements et les bruits de combustion [Farnell, 2013] [Verron, 2010]. Cette étape d'analyse permet d'estimer les paramètres d'un modèle de synthèse à partir duquel pourra être généré un son – points ② et ③. L'oreille est ensuite utilisée comme juge entre le son original et le son synthétisé – point ④. Si une différence est détectée par rapport au critère perceptif étudié – le timbre dans l'exemple de Risset et Wessel – alors le modèle de synthèse peut être retravaillé. Une nouvelle itération pourra permettre d'améliorer le modèle et de potentiellement réduire la différence perceptive. Dans le cas contraire, la différence entre les deux signaux est jugée nulle par rapport à l'attribut perceptif à l'étude.

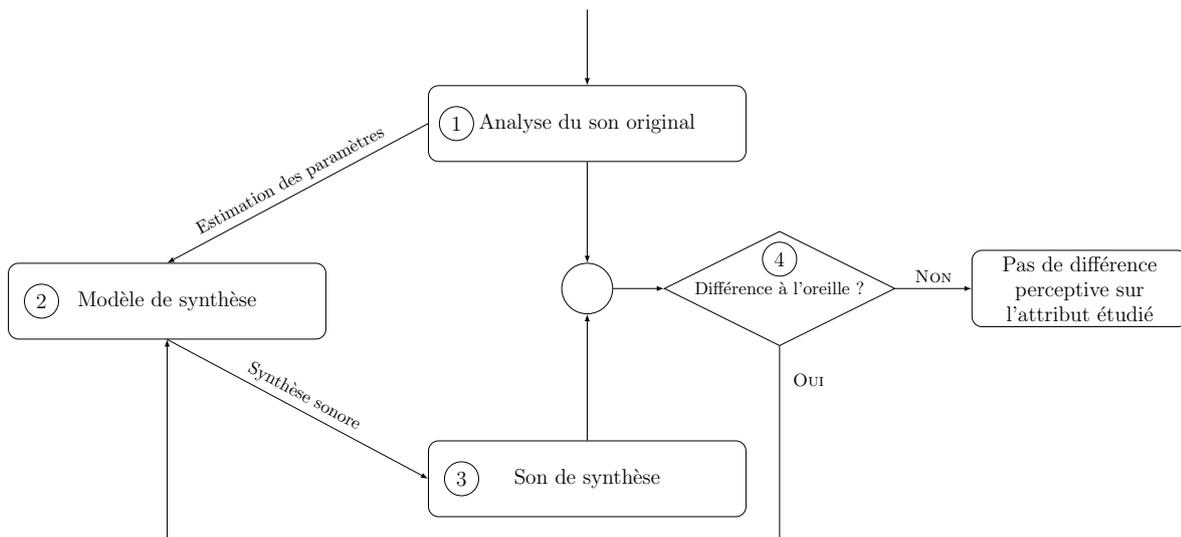


FIGURE 2.8 – Illustration de l'analyse – synthèse, figure adaptée de [Risset et Wessel, 1999]. L'entrée du système est indiquée par la flèche en haut du schéma.

Cette méthodologie est très utile pour isoler certains attributs du son, d'où sa popularité pour étudier les sons quotidiens et les invariants perceptifs, ce qui a permis d'accumuler des connaissances sur le lien entre paramètres bas et haut-niveau, pour considérablement améliorer la synthèse de sons écologiques.

Le paradigme action - objet

Comme nous l'avons vu précédemment, il est possible de distinguer dans un son naturel des invariants transformationnels liés aux actions, et des invariants structuraux propres aux objets. Nous pouvons alors lier ces paramètres de contrôle haut-niveau au sein du paradigme *action - objet*, qui énonce qu'un événement sonore est la conséquence d'une action sur des objets. Ce paradigme va structurer les paramètres de contrôle du synthétiseur que nous allons présenter.

Ce modèle pose également les bases de la synthèse des sons du quotidien. En prenant l'exemple d'un impact sur une plaque de métal, l'action de frappe va venir faire vibrer la plaque, qui va émettre un son caractéristique à cause de ses résonances particulières. Il est alors possible de mettre en évidence une source excitatrice – l'objet qui frappe – et un résonateur – ici la plaque de métal. L'utilisation d'un modèle de synthèse source – filtre semble alors particulièrement adapté à notre paradigme.

Un synthétiseur de sons naturels

Un synthétiseur de sons d'impacts a été tout d'abord mis au point par Aramaki, Ystad et Kronland-Martinet [Aramaki et Kronland-Martinet, 2006] [Aramaki *et al.*, 2009] [Aramaki *et al.*, 2010], en s'appuyant sur un modèle source – filtre basé sur le paradigme action – objet que nous venons de présenter. L'étude des invariants structuraux a permis de faire le lien entre paramètres bas-niveau et le contrôle haut-niveau d'évocations d'objets de différentes natures. Il a donc été possible de proposer des sons d'impacts avec des transitions continues entre des paramètres évocatoires : les matériaux – métal, bois, ou verre par exemple ; la taille ou encore la forme de l'objet impacté. Par la suite, plusieurs études ont porté sur la prolongation de ces travaux pour intégrer des excitations plus complexes liées aux invariants transformationnels comme le frottement [Thoret *et al.*, 2013], le grattement [Van Den Doel *et al.*, 2001] [Conan *et al.*, 2012], ou le roulement [Stoelinga et Chaigne, 2007] [Conan *et al.*, 2014], et permettre de proposer un large ensemble de possibilités pour synthétiser des sons quotidiens [Conan *et al.*, 2013], comme le montre la figure 2.9. Il est alors possible de faire des transitions continues entre les actions, entre plusieurs matériaux ou entre différentes formes grâce aux études précédemment citées. Des études peuvent alors utiliser ce synthétiseur pour créer des sons de l'environnement en maîtrisant les différents paramètres, pour travailler sur les caractéristiques principales perçues lors de l'écoute d'un son par exemple [Bordonné *et al.*, 2017]. Ce synthétiseur permet également de créer des sons de l'environnement n'ayant pas de réalité physique, comme un son de corde en verre.



FIGURE 2.9 – Interface proposée du synthétiseur de sons naturels, figure tirée de [Conan *et al.*, 2013].

Gaver s'est également intéressé à d'autres types de sons, comme les sons de liquides ou de machines [Gaver, 1993a]. Des études ont depuis été menées pour explorer ces différents types de sons [Doel, 2005] [Verron, 2010] [Forsan, 2016], ainsi que des sons de vent [Dobashi *et al.*, 2003] ou de feu [Dobashi *et al.*, 2004]. Verron a d'ailleurs pu mettre également au point un synthétiseur de sons environnementaux, restitués en trois dimensions [Verron *et al.*, 2010]. Ce synthétiseur ne se base pas sur les phénomènes physiques comme le fait Gaver, mais sur les morphologies des signaux pour les événements trop compliqués à décrire par un modèle physique.

Au cours de cette section, nous avons donc abordé la théorie écologique, qui a apporté un changement de perspective intéressant pour l'étude de sons de l'environnement. Selon le paradigme action – objet, les sons de l'environnement sont la conséquence de l'interaction de plusieurs objets. Les caractéristiques des objets seront alors transmises par les invariants structuraux tandis que les caractéristiques de l'action seront transmises par les invariants transformationnels. L'onde sonore va ensuite se propager et être influencée par le milieu environnant, avant d'être captée par un éventuel auditeur. Un aperçu des informations qu'il est possible de percevoir au cours de ce processus est résumé sur la figure 2.10.

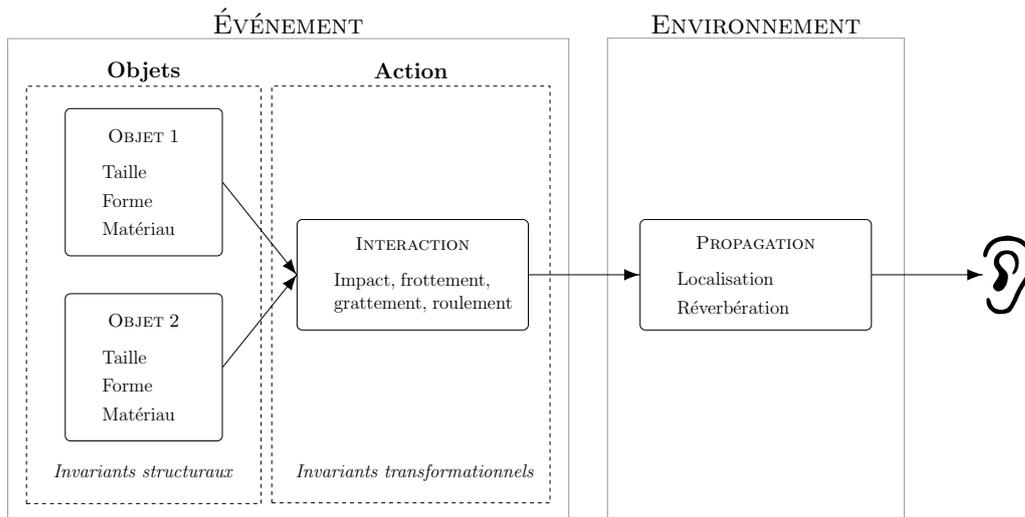


FIGURE 2.10 – Schéma listant différentes informations évoquées par le son.

Nous avons également pu voir comment la technique d'analyse – synthèse a permis d'étudier des invariants de sons de l'environnement. Ces connaissances ont permis de mettre au point plusieurs synthétiseurs de sons de l'environnement, permettant aux utilisateurs non experts de créer des signaux correspondant à leurs attentes à partir de mots du langage courant. Au-delà de la mise au point de sons réalistes, la synthèse permet également de créer des sons métaphoriques, qui n'existent pas naturellement mais qui portent tout de même un sens.

La volonté de transmettre des informations de manière efficace par le biais du son a naturellement orienté cette thèse vers l'étude de sons de l'environnement. La génération des sons fortement évocatoires grâce à un synthétiseur peut être une solution intéressante dans la construction d'une métaphore, et sera abordée en section 5.1.2. Nous allons à présent nous intéresser aux interactions qui peuvent exister entre modalités sonores et gestuelles, thème central du sujet de cette thèse, ainsi qu'aux informations de mouvements contenus par les sons.

2.4 Relations entre mouvements et sons

L'interface que nous voulons créer se pilote grâce à des gestes de la main de l'utilisateur, et donne un retour informatif par le biais de sons. Le but est d'inclure l'utilisateur dans une boucle action – perception qui doit être la plus intuitive possible, pour qu'il s'approprie l'IHM. Le concept d'objet virtuel a pour objectif de rappeler les interactions naturelles que l'utilisateur a avec les objets de son environnement. Nous allons à présent nous intéresser aux études où des interactions entre mouvements et sons ont été proposées, pour en étudier les impacts et limites.

Nous examinerons tout d'abord comment certains sons peuvent contenir une notion de mouvement, et le lien qu'il peut exister entre des trajectoires et des paramètres acoustiques. Le cas particulier de la relation entre des mouvements de la main et des sons sera ensuite étudié, qui est un sujet central de ces travaux de thèse. Plusieurs exemples de recherches dans le domaine pourront nous guider dans la construction de l'interface gestuelle.

2.4.1 Évocation du mouvement par les sons

Nous allons ici nous intéresser aux évocations de mouvements qu'il est possible de créer grâce à un contenu sonore. Comme nous l'avons vu en section 2.1.2, plusieurs indices permettent de percevoir la localisation d'une source, et donc par extension un mouvement. L'effet Doppler est par exemple caractéristique d'une voiture de course circulant à vive allure. Il est donc question ici d'évocations en excluant ce type d'indice, pour analyser l'évocation qui peut être créée par le contenu même du son. Les stimuli sont par exemple monophoniques, pour supprimer les indices de localisation. De la même manière, les sons de l'environnement sont exclus à cause de leur fort pouvoir évocatoire, qui peut varier selon le contexte ou la personne. Nous nous intéressons donc aux évocations de mouvements véhiculées par des sons abstraits ou acousmatiques [Schaeffer, 1966].

Plusieurs études ont porté sur les évocations de mouvements liées à la musique [Shove et Repp, 1995] [Eitan et Rothschild, 2011]. Le vocabulaire musical semble en effet rattaché à des mouvements – on parle de *hauteur* pour la fréquence d'une note ; de *mouvement* pour une partie d'une œuvre musicale ; de *ballade* ou *berceuse* pour des genres musicaux ; ou encore de *glissando* pour une transition entre plusieurs notes. La production musicale en elle-même requiert le plus souvent des gestes, comme nous le verrons dans la prochaine section.

Merer s'est intéressé à l'évocation de mouvements induits par les propriétés intrinsèques des sons [Merer, 2011]. Ses études ont tout d'abord consisté en une catégorisation libre de sons abstraits, et il a été possible de distinguer 5 catégories de sons associées à des mouvements : « tourne », « descend », « approche », « passe », et « monte » [Merer *et al.*, 2008]. Un second test a alors été mené avec les mêmes sons abstraits pour observer ou non un consensus de catégorisation selon les catégories définies. À partir des sons catégorisés, il a été possible d'étudier d'éventuels descripteurs liés à chaque catégorie de son. La fréquence de modulation a par exemple pu être reliée à un son qui évoque une rotation, et des sons d'impacts caractérisés par un temps d'attaque court et une longueur faible sont souvent associés aux sons qui évoquent une chute.

Merer a observé que l'écoute de sons évoquant des mouvements était souvent décrite avec des dessins. Il s'est alors intéressé à l'appariement de représentations visuelles avec des sons évocatoires de déplacements. Une première étape a consisté à apparier des sons avec des formes de trajectoires dessinées sur une interface grâce à un jeu de paramètres pré-déterminés [Merer *et al.*, 2013]. Cette IHM était nécessaire pour diminuer les différences inter-individuelles provenant des capacités de dessin. Un exemple de l'interface est donné figure 2.11.

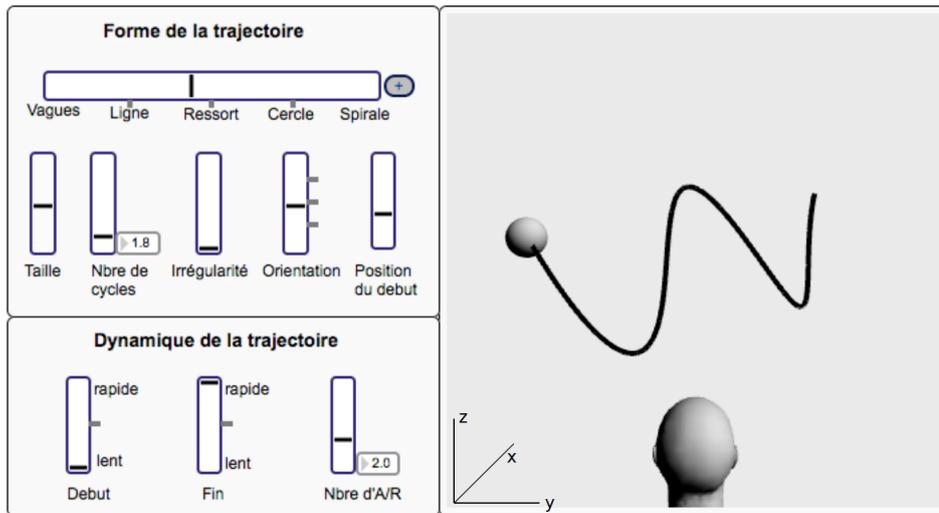


FIGURE 2.11 – Exemple de l’interface permettant des dessins de trajectoires, figure tirée de [Merer, 2011].

Lors de la seconde étape, Merer a présenté une série de sons abstraits conjointement à des trajectoires. Les mouvements pouvaient être décomposés en différentes formes, directions et fréquences. La tâche des sujets consistait à donner une note de cohérence sur une échelle continue. Les résultats montrent un bon appariement entre les sons et les vidéos de dessins, que ce soit pour les formes – notamment avec les formes *creux*, *dôme* et *ligne*, les fréquences ou les directions.

À la suite de ces études, Merer propose un outil liant des trajectoires à des sons, par un processus de sonification. Les données d’entrée peuvent provenir de différents supports – vidéos, dessins ou gestes par exemple – d’où sont extraites des trajectoires. S’en suit une phase de traitement des trajectoires, qui sont découpées en sous-éléments et analysées selon leur forme, taille, direction, ou encore durée. Certaines caractéristiques des trajectoires sont ensuite liées à des paramètres de contrôle de sonification. Cet outil est intéressant quant à la démarche adoptée, même si nous ne pourrions pas l’utiliser tel quel. En effet, nous n’utiliserons pas directement des trajectoires, mais certains paramètres plus haut-niveau liés à des gestes précis.

Ces études tendent à montrer qu’il est possible de véhiculer une notion de mouvement à travers le son, et Merer a pu isoler des paramètres sonores permettant d’évoquer certaines formes, fréquences ou directions. L’utilisation de dessins est une originalité des protocoles mis en place, mais ils reflètent en réalité un appariement naturel entre mouvements et sons. Effectivement, certains sujets ont spontanément dessiné des trajectoires pour classer des sons [Merer, 2011]. Cet appariement amène alors naturellement à la notion de gestes, et à l’association qui peut exister entre les gestes et les sons.

2.4.2 Interactions sonores et gestuelles

La relation entre les sons et les gestes date de plusieurs milliers d'années, au travers des instruments de musique. Dans son livre *The History of Musical Instruments*, Sachs suppose que les prémices des instruments de musique viennent de l'expression d'émotions au travers de gestes, en frappant le sol, ou en tapant des mains [Sachs, 2012]. Le son ne devait alors être que secondaire, et ce n'est qu'ensuite que les instruments de musique ont dû apparaître, lorsque la production sonore était au centre de l'intérêt. Il est intéressant de souligner ici le lien de cause à effet entre gestes et sons.

Nombre d'instruments acoustiques produisent du son grâce à des gestes – au sens large, on ne restreint pas ici aux gestes définis par Kurtenbach et Hulteen [Kurtenbach et Hulteen, 1990] – d'une part pour la production sonore, que ce soient les percussions, les claviers ou encore les cordes. Le retour sonore des instruments de musique va alors influencer les gestes des musiciens, en reproduisant la boucle interactive action – perception de la figure 1.8. Cette boucle est particulièrement intéressante pour nos travaux, et nous l'utiliserons dans la mise en place de notre stratégie de sonification. Les gestes entrent également en compte dans l'expressivité musicale, par exemple en ce qui concerne les violoncellistes [Rozé, 2017]. Rozé a en effet mené une étude concernant les mouvements de postures, où les sons produits avec des gestes libres ou contraints sont analysés [Rozé *et al.*, 2016]. Il est alors observé des variations sonores importantes entre les deux conditions, mettant en évidence le lien entre expressivité musicale et mouvements posturaux. Les gestes sont également utilisés avec les instruments électroniques, que ce soit avec contacts comme avec le *Radio Drum* de Mathews [Mathews, 1989] ou sans contact comme nous l'avons vu avec l'exemple du Theremine en section 1.1.1.

D'un point de vue fondamental, plusieurs études se sont penchées sur l'association gestes – sons et deux utilisations sont alors à distinguer selon Boyer *et al.*, qui ont particulièrement étudié cette appariement [Boyer *et al.*, 2013] [Boyer, 2015]. La première a pour but la création sonore – une *tâche orientée son* [Boyer *et al.*, 2013], où le geste est un intermédiaire d'expression entre l'homme et la machine, comme c'est le cas pour le Theremine. Plusieurs domaines ont été explorés, comme l'étude et l'analyse de gestes musicaux [Cadoz, 1994] [Cadoz et Wanderley, 2000] [Godøy *et al.*, 2006] [Jensenius *et al.*, 2009], les stratégies d'association entre gestes et sons dans un contexte musical [Caramiaux *et al.*, 2010] [Barrett et Nymoen, 2015], ou la création de nouveaux moyens de composition [Wanderley et Depalle, 2004] [Goïna et Polotti, 2008]. Le projet SkAT-VG¹ propose des outils pour créer l'ébauche d'un son à partir de la voix et de gestes, ce qui peut être intéressant dans le cadre de la composition musicale [Rocchesso *et al.*, 2015]. Dans le cadre de ce projet, Lemaitre *et al.* ont par exemple étudié la combinaison de gestes et de vocalisations non-verbales pour l'imitation de sons écologiques ou non [Lemaitre *et al.*, 2017]. Lors des expériences qui ont été menées, les vocalisations servent à imiter le plus fidèlement possible le son, au travers de l'ensemble de ses caractéristiques ; alors que les gestes illustrent une seule caractéristique du son, supposée la plus importante aux yeux des sujets. Les résultats montrent l'utilisation de métaphores audio-visuelles, dont deux sont partagées par les sujets : un son ayant une hauteur ascendante est accompagné d'un geste vertical de la main vers le haut, et un son caractérisé par des variations aléatoires est illustré par des gestes rapide de tremblements de la main et des doigts. Ces enseignements peuvent nous inspirer pour la création de notre interface gestuelle, que nous voulons également basée sur des métaphores.

1. <http://skatvg.iuav.it/>

Le second volet des utilisations possibles de l'interaction son – geste, qui va nous intéresser plus particulièrement pour cette thèse, a pour but la production d'un geste particulier où un retour sonore est présent pour l'accompagner – une *tâche orientée geste* [Boyer *et al.*, 2013]. Plusieurs études ont utilisé des gestes conjointement à des stratégies de sonification interactive – voir section 1.2.2 – pour donner un retour en temps réel à l'utilisateur sur ses gestes. Le son est par exemple utilisé comme retour d'information par Kane pour permettre à des utilisateurs aveugles d'utiliser une interface tactile [Kane *et al.*, 2013], ou pour permettre une amélioration de l'écriture pour des enfants dysgraphiques comme nous l'avons déjà vu [Danna *et al.*, 2013a]. La sonification associée à des gestes permet d'augmenter les performances de tâches motrices, par exemple dans une tâche de régularité [Van Vugt et Tillmann, 2015]. Dans une étude portant sur la sonification de trajectoire, Dyer a montré que le retour sonore d'un geste à la trajectoire parfaite pouvait permettre de « rafraîchir » un apprentissage [Dyer *et al.*, 2016]. Les gestes sportifs associés à des sons peuvent également permettre l'augmentation de performances comme nous l'avons vu avec Schaffert [Schaffert et Mattes, 2012], ou comme l'ont montré Sors *et al.* [Sors *et al.*, 2015]. Effenberg a par exemple montré que la perception et la reproduction d'un mouvement sportif sont meilleures dans une condition audio-visuelle, comparée aux conditions unimodales [Effenberg, 2004].

Rosati *et al.* présentent également une série d'expériences intéressantes, ayant pour but d'évaluer différents types de stratégies sonores sur une tâche motrice de suivi de cible [Rosati *et al.*, 2012]. Une stratégie de sonification de la vitesse de la cible mène alors à de meilleures performances, comparée à une sonification de l'erreur positionnelle. Une seconde expérience montre ensuite qu'un retour sonore est plus efficace qu'un retour visuel sur la tâche demandée. Dans les conditions d'utilisation du visuel seul, la vue des sujets doit à la fois se concentrer sur le suivi de la trajectoire, et sur la couleur de la cible donnant une indication de leurs performances. Dans les conditions audio-visuelles, le retour d'information sur les performances est donné cette fois par le son. Ces résultats sont à rapprocher des hypothèses que nous avons formulées concernant l'attention : la concentration d'informations accessibles par la modalité visuelle mène à de moins bons résultats comparée à une utilisation conjointe de différentes modalités, même si les informations visuelles sont concentrées au même endroit. De plus, l'association de gestes et de sons a permis d'arriver à de surprenant effets, comme l'a montré Thoret [Thoret, 2014] [Thoret *et al.*, 2016a]. Dans cette étude, Thoret a associé les gestes aux modalités visuelles et auditives. Il est alors demandé à des sujets d'utiliser un stylet pour reproduire le mouvement d'un point lumineux présenté sur un écran, tout en écoutant des sons au casque. La vue de leur main est masquée, et les stimuli visuels et sonores correspondent dans certains cas à des cercles et dans d'autres à des ellipses, comme le montre la figure 2.13a. Il a alors été montré que les stimuli sonores et visuels influencent les gestes avec des dessins représentant des ellipses, avec un effet conjoint comme le montre la figure 2.13b.



FIGURE 2.12 – Sujet au cours d'une expérience de Rosati *et al.*, consistant à suivre une cible visuelle sur une tablette graphique. Figure tirée de [Rosati *et al.*, 2012].

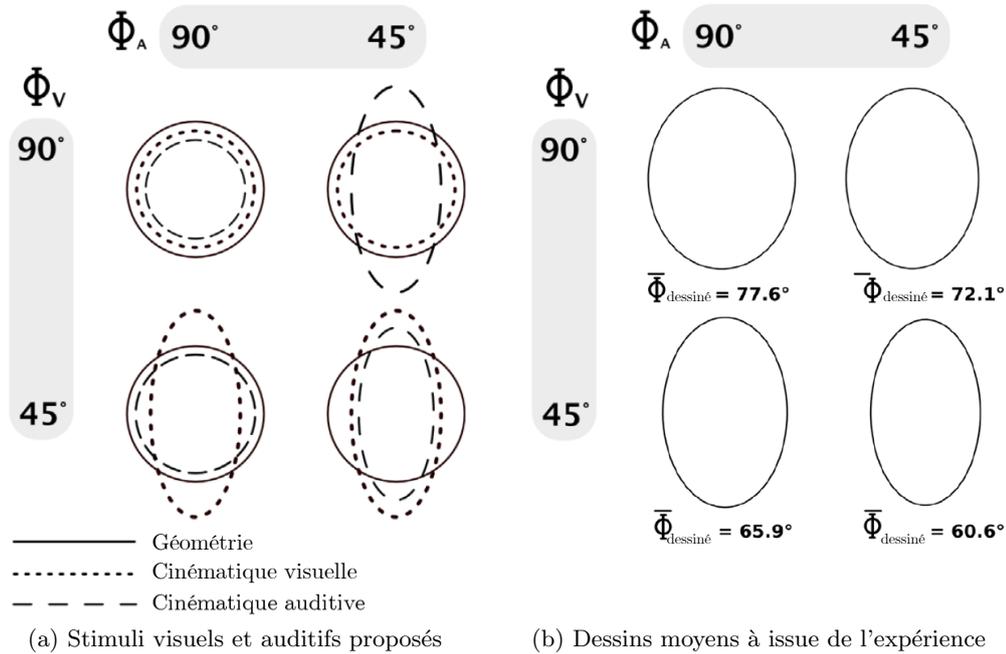


FIGURE 2.13 – Stimuli et résultats de l'expérience menée par Thoret [Thoret *et al.*, 2016a], dont les figures sont adaptées. Φ désigne la phase relative des ellipses et les indices Φ_V et Φ_A désignent respectivement les phases des modalités visuelles et auditives.

Il a été possible de mettre en évidence un effet de la cinématique visuelle – $F(1,16) = 147,22$; $p < 0,001$ – sur l'excentricité des ellipses, qui semble un peu plus prononcé que l'effet de la cinématique auditive – $F(1,16) = 12,83$; $p < 0,01$. Cette différence peut être expliquée par le rôle prépondérant que joue le visuel dans l'intégration multi-sensorielle, comme nous l'avons vu précédemment. Il est intéressant de noter que les distorsions observées lorsqu'une seule modalité est modifiée – $11,7^\circ$ pour le visuel, $5,5^\circ$ pour le son – sont pratiquement additives avec une distorsion de 17° . Il n'y a alors pas d'effet audio-visuel combiné – $F(1,16) = 1,01$; $p = 0,96$. Ces résultats tendent à montrer la grande influence des retours d'informations sur les gestes, et des effets des retours sonores et visuels, chacune de ces modalités ayant provoqué une variation sur la forme dessinée.

Une autre étude menée par Rath et Schleicher va particulièrement nous intéresser dans le cadre de cette thèse. Ces travaux portent sur une interface visuelle, auditive et gestuelle basée sur la métaphore d'une balle virtuelle roulante appelée le « Ballancer » [Rath et Rocchesso, 2005]. Ce dispositif est composé d'une barre qu'il est possible d'incliner, et d'une balle virtuelle, qui roule selon l'angle de la barre sous l'effet de la gravitation. La vitesse de la balle est sonifiée, et il est proposé d'atteindre une zone de la barre avec la balle dans 3 conditions sonores différentes : sans son, avec un son abstrait, et avec un son de roulement [Rath et Schleicher, 2008]. Il est également possible de voir la balle sur un écran, comme le montre la figure 2.14.

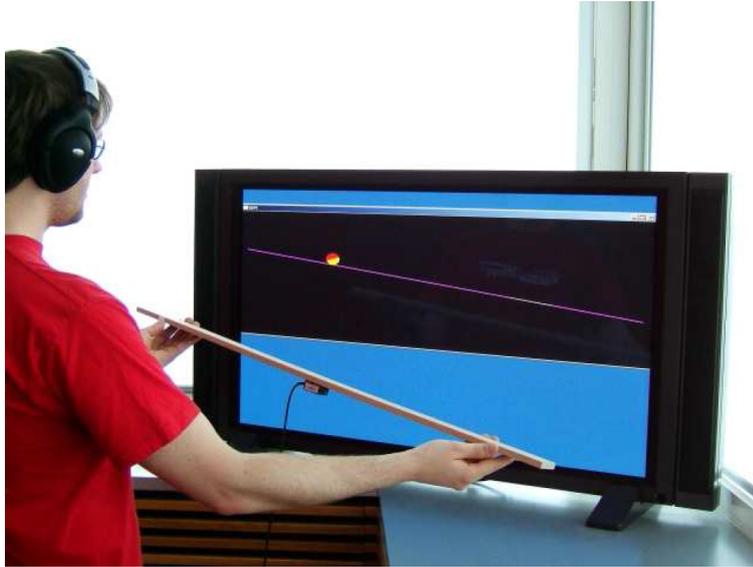


FIGURE 2.14 – Le dispositif du « Ballancer », proposé par Rath, figure extraite de [Rath et Schleicher, 2008].

La zone à atteindre est indiquée à l'écran par une couleur de la barre différente, mais également au travers du son par un retour sonore plus rugueux. Les résultats portant sur le temps d'accomplissement de la tâche, renseignés dans le tableau 2.1, montrent un effet intéressant : le son de roulement conduit à de meilleurs résultats dans un premier temps, de manière significative avec la condition sans son. En revanche, après un entraînement c'est le son abstrait qui semble être le meilleur retour sonore, avec une différence significative avec la condition sans son. Il n'y a pas d'effets significatifs entre le son de roulement et le son abstrait, probablement à cause de la participation de 6 sujets à cette expérience, ce qui est relativement faible. Nous ne pouvons donc pas conclure sur la « meilleure » stratégie sonore avant ou après entraînement, mais cette étude met en avant une tendance intéressante. La sonification semble conduire à des comportements différents selon le type de son choisi, ce qui peut être intéressant dans la suite de la thèse.

	Sujets non entraînés		Sujets entraînés	
	<i>as</i>	<i>no</i>	<i>as</i>	<i>no</i>
<i>rs</i>				
$\delta(\%)$	-9,1	-17,9	8,1	-3
<i>p</i>	0,1345	0,0052	0,1414	0,5977
<i>as</i>				
$\delta(\%)$		-9,7		-10,2
<i>p</i>		0,1825		0,0206

TABLEAU 2.1 – Résultats obtenus lors de l'expérience menée par Rath et Schleicher, tableau reproduit depuis [Rath et Schleicher, 2008]. Les différences de temps δ sont renseignées en pourcentages, et la significativité *p* est également indiquée avec les effets significatifs en gras.

rs = son de roulement, *as* = son abstrait et *no* = condition sans son.

Dans cette partie, nous avons abordé différents aspects de la relation entre les mouvements et les sons. Des exemples d'évocations de trajectoires par les sons ont été développés, avant d'aborder plus particulièrement la relation gestes – sons. Dans cette dernière section, il est possible de distinguer deux visions de cette association, qui diffèrent dans le but à atteindre. Dans les *tâches orientées son*, le son est l'objectif et le geste n'est qu'une modalité pour y parvenir ; alors que le geste est le but des *tâches orientées geste*, le son jouant le rôle de guide pour y parvenir. Nous avons abordé plusieurs études concernant ce deuxième cas, notamment celles menées par Rosati, Thoret ou Rath qui ont conduit à des résultats intéressants, et qu'il est important de garder en tête dans l'optique de la construction de notre interface gestuelle. Selon l'application que nous choisirons, nous pourrions nous placer dans une tâche *orientée geste* ou *orientée son*.

Conclusion du chapitre

Pour aborder ce chapitre centré sur la perception, nous avons tout d'abord étudié les systèmes mis en jeux dans la perception, en restant focalisé sur les aspects sonores et gestuels. La proprioception est notamment un retour d'information indispensable, et il sera utile pour l'utilisation de notre interface par des gestes sans nécessiter de retour visuel. Nous nous sommes ensuite plongés dans l'intégration multisensorielle, en observant comment les modalités utilisées en coopération pouvaient permettre une perception robuste et précise, ou résoudre des ambiguïtés. La modalité visuelle semble être privilégiée par rapport aux autres sens, même si elle peut être perturbée dans certains cas. Le facteur de l'attention a ensuite été discuté pour souligner son implication dans la perception multisensorielle. La mise au point de l'interface gestuelle, qui vise à combiner des informations multisensorielles, devra prendre en compte ces effets attentionnels pour une bonne utilisation en conduite.

La théorie écologique appliquée au son, prolongement des travaux de Gibson, a été décrite pour mettre en lumière une perception sonore basée sur une écoute naturelle, avec l'événement comme origine. Les sons découlent d'interactions de plusieurs objets, et il est possible de remonter à ces informations grâce à la présence d'invariants dans le son. Ils peuvent être propres à l'objet ou à l'interaction, et sont respectivement nommés invariants structuraux ou transformationnels. Après de multiples études sur ces invariants, l'équipe du laboratoire PRISM a mis au point un synthétiseur permettant de générer des sons d'événements, avec des paramètres de contrôle haut-niveau.

Nous nous sommes ensuite plus particulièrement intéressés à la relation entre les mouvements et les sons. Les moyens d'évoquer le mouvement grâce aux sons ont été développés, en se focalisant sur le contenu des sons. Nous avons ensuite évoqué différentes études mettant en relation des mouvements de la main et les sons, où nous avons distingué l'utilisation du geste comme modalité d'expression musicale, et l'utilisation du son comme retour d'information sur les gestes. L'intérêt d'un retour sonore sur une tâche gestuelle a notamment été montré par rapport à un retour visuel [Rosati *et al.*, 2012]. Il a également été montré que le son peut modifier les gestes [Thoret, 2014]. Enfin, la nature du retour sonore semble également avoir une influence sur l'évolution des performances gestuelles au cours du temps [Rath et Schleicher, 2008].

Les études et exemples qui ont été abordés dans ce chapitre montrent la complexité de la perception humaine, qui dépend des modalités sollicitées, de la tâche ciblée, ou de l'attention du sujet. Les sons de l'environnement, grâce à leur perception directe et le sens qu'ils véhiculent, peuvent être intéressants dans la suite de nos travaux. Le développement de synthétiseurs de sons naturels va nous permettre de les utiliser facilement, et d'en contrôler les paramètres. La métaphore de l'objet virtuel peut alors être augmentée par l'utilisation de ce type de sons.

Aspects cognitifs et distractions en conduite

Les innovations technologiques font continuellement évoluer les interfaces avec des fonctions qui permettent de nouveaux usages. Lors de la première confrontation avec une interface, un apprentissage est nécessaire. Ce temps de prise en main permet une familiarisation avec l'IHM, et est nécessaire pour faciliter une manipulation ultérieure. La concentration de l'utilisateur, élevée au début de la manipulation de l'interface, va ensuite diminuer au fur et à mesure grâce à l'automatisation de certains procédés. Après un usage fréquent, il peut être possible de proposer une seconde tâche sans mener à une surcharge cognitive.

Dans un véhicule automobile, des interfaces ont été proposées depuis de nombreuses années pour manipuler des fonctions de confort, comme la climatisation ou les médias. Ces IHM, tout d'abord proposées sous la forme de boutons, ont ensuite évolué en intégrant de nouvelles fonctions – GPS ou systèmes d'aide à la conduite par exemple – vers des écrans multifonctions. Leur utilisation en conduite est fréquente et il est nécessaire de maintenir la concentration du conducteur à un niveau peu élevé, prioritairement pour que la conduite reste maîtrisée, ainsi que la manipulation de l'interface. Cependant, ces IHM mobilisent des sens parfois requis pour la conduite : la manipulation d'un écran tactile nécessite par exemple la modalité visuelle du conducteur, et son maniement peut dégrader la vigilance du conducteur sur la tâche conduite. Il est aisé d'imaginer que si l'attention du conducteur est portée principalement sur une manipulation de l'IHM, le temps de réaction à un danger sera probablement augmenté, ce qui peut mettre en danger les piétons et les occupants du véhicule. Cette supposition est appuyée par l'étude de Duncan que nous avons évoqué en section 2.2.3 sur des détections auditives ou visuelles simultanées [Duncan *et al.*, 1997]. C'est l'une des problématiques des interfaces implantées en véhicule actuellement, qui nous a poussé à envisager une alternative. L'objectif de notre étude consiste à proposer un contournement de ce problème par l'utilisation de modalités plus disponibles que la vision dans un contexte de conduite, ce qui pourrait permettre une meilleure réactivité chez les conducteurs.

Lors de la présentation de notre interface gestuelle à de nouveaux utilisateurs, un temps d'adaptation leur est nécessaire pour qu'ils intègrent son fonctionnement, et qu'ils puissent s'en servir convenablement. Ces mécanismes de création de schémas cognitifs seront abordés dans une première partie, où différentes définitions concernant la mémoire et la charge cognitive seront précisées. Nous verrons notamment plusieurs moyens d'estimer la charge cognitive, ce qui sera un outil important lors du test de l'interface gestuelle en conduite.

Nous aborderons ensuite plusieurs recherches concernant les effets distractifs en conduite, provenant de manipulations de systèmes annexes à la conduite, comme le téléphone ou les écrans tactiles. L'apport de l'outil du simulateur de conduite sera discuté, avec notamment les critères de validité et l'apport sécuritaire pour étudier les phénomènes liés à l'attention. Les conséquences des différents types de distractions sur la conduite seront ensuite présentées, ce qui permettra de connaître les paramètres de conduite qu'il sera intéressant d'observer lors du test de notre interface gestuelle.

Sommaire

3.1	La construction de schémas cognitifs	61
3.1.1	Mémoire et charge cognitive	61
3.1.2	Mécanismes d'apprentissage	62
3.1.3	Techniques de mesure de la charge cognitive	65
3.2	Les phénomènes distractifs en conduite	68
3.2.1	Le simulateur de conduite : avantages et limites	69
3.2.2	Distractions motrices	72
3.2.3	Distractions visuelles	72
3.2.4	Distractions cognitives	73
3.2.5	Comparaison et conséquences sur la conduite	74

3.1 La construction de schémas cognitifs

Cette partie présente différentes notions selon la théorie de la charge cognitive mise au point notamment par Paas, Van Merriënboer et Sweller [Paas et Van Merriënboer, 1994a] [Sweller *et al.*, 1998]. La notion de mémoire est fondamentale pour comprendre l'élaboration de schémas cognitifs, ce qui peut permettre une baisse de la charge cognitive globale.

3.1.1 Mémoire et charge cognitive

Nous débutons cette section par quelques notions simples des sciences cognitives concernant la mémoire. La charge cognitive sera également abordée conjointement, puisqu'elle y est fortement liée. Ces définitions seront utiles pour la mesure de la charge cognitive d'une part, et pour la compréhension des mécanismes d'apprentissage d'autre part.

La mémoire de travail

La mémoire de travail peut être définie comme une capacité de stockage et de manipulation des informations temporaires pour l'accomplissement de tâches complexes [Baddeley, 1992]. Sweller *et al.* relie même la mémoire de travail à la conscience [Sweller *et al.*, 1998]. Cette mémoire est limitée [Sweller *et al.*, 1998], il est donc possible d'interagir simultanément avec un nombre limité d'éléments de notre environnement – un élément étant défini par l'information qui peut être traitée par un apprenant comme une seule unité dans la mémoire de travail [Pollock *et al.*, 2002]. Miller a dénombré à environ sept le nombre d'éléments pouvant être retenus simultanément en mémoire de travail [Miller, 1956].

Baddeley divise cette mémoire en trois sous-systèmes spécifiques dont un système central exécutif, dédié à l'attention et deux systèmes esclaves : une boucle phonologique – ou *phonological loop*, et une ardoise visuo-spatiale – ou *visual-spatial scratch pad* [Baddeley, 1992]. Le premier concerne une mémoire de travail auditive, alors que le second se réfère au traitement des images par une mémoire de travail visuelle. Plusieurs expériences comportant des doubles tâches ont mis en évidence la séparabilité de ces deux composants de la mémoire de travail. Plusieurs études tendent à prouver que l'utilisation commune de ces sous-systèmes est optimale par rapport à une utilisation unimodale [Penney, 1989] [Mousavi *et al.*, 1995] [Tindall-Ford *et al.*, 1997], ce qui confirme nos observations concernant l'intégration multisensorielle et le traitement de l'attention.

La charge cognitive peut être définie comme la concentration qui est mobilisée par le système cognitif de l'apprenant pour l'accomplissement d'une tâche particulière [Paas et Van Merriënboer, 1994a]. Elle est associée à la mémoire de travail et il est possible de la diviser en deux catégories :

- La charge cognitive intrinsèque : c'est la charge intrinsèque à la nature de la tâche demandée, elle est fixe pour un élément donné. Il est par exemple beaucoup plus difficile de conduire une voiture que d'écrire un mot, la différence de charge cognitive provient de la nature même de la tâche.
- La charge cognitive extrinsèque : elle concerne le mode de présentation d'une tâche, qu'il est possible d'optimiser. En reprenant l'exemple de conduire une voiture, la charge cognitive concernant la conduite avec une seule main sera plus grande que si les deux mains sont autorisées. La différence provient ici de la charge extrinsèque, puisque la tâche de conduire la voiture reste identique.

Les charges cognitives intrinsèque et extrinsèque sont additives. Pour une tâche donnée, la baisse de charge cognitive globale d'une personne peut donc se faire uniquement par la diminution de la charge cognitive extrinsèque, en ciblant la méthode de présentation de la tâche. Il est alors possible d'employer différentes stratégies selon la nature de la tâche. Si la tâche consiste à expliquer un concept théorique, dans le cadre d'un cours de mathématique par exemple, l'utilisation d'exemples d'applications peut permettre de faciliter la compréhension des élèves. Dans notre cas d'utilisation d'une interface en conduite, l'exploitation de plusieurs modalités au lieu d'une seule peut également être une solution pour diminuer la charge cognitive extrinsèque [Sweller *et al.*, 1998].

La mémoire long terme

Ce type de mémoire a quant à lui une capacité supposée illimitée. Les nouvelles informations doivent être traitées par la mémoire de travail pour construire des schémas qui sont catégorisés dans la mémoire long terme [Sweller *et al.*, 1998], qui sont ensuite traités comme un seul élément. La mémoire long terme semble par exemple être un facteur à l'origine des grandes différences entre des joueurs d'échecs moyens et les grands maîtres [De Groot, 1966] [Chase et Simon, 1973].

Ces deux types de mémoire peuvent être utilisés conjointement, en sollicitant des connaissances de la mémoire long terme, et en y associant de nouvelles données traitées par la mémoire de travail. Dans notre application, les conducteurs devront pouvoir créer un schéma concernant les manipulations de l'objet virtuel dans un premier temps. C'est seulement après avoir acquis un tel schéma qu'il peut être possible de faire une autre tâche conjointement sans mobiliser une concentration trop importante. La conduite peut donc être possible seulement après l'acquisition de ce schéma. Nous allons à présent aborder les mécanismes d'apprentissage, qui sollicitent d'abord la mémoire de travail pour ensuite stocker des informations dans la mémoire long terme.

3.1.2 Mécanismes d'apprentissage

Comme nous l'avons abordé rapidement précédemment, il est possible de construire des schémas grâce à la mémoire de travail, qui sont ensuite stockés dans la mémoire long terme. C'est la construction et l'utilisation de tels schémas, qui vont ensuite permettre l'accomplissement de tâches complexes. Nous allons ici prendre l'exemple de la lecture, comme l'a fait Van Merriënboer [Van Merriënboer et Sweller, 2005]. Un lecteur, en lisant un texte, ne va pas déchiffrer chaque symbole, lui associer ensuite une lettre, et enfin prendre en compte tous les symboles dessinés côte à côte pour lire un mot. Ces processus demanderaient une charge cognitive importante. En réalité, le lecteur possède un schéma qui permet de reconnaître l'écriture d'une lettre, malgré l'infinité de formes qu'elle peut avoir – provenant de l'écriture manuscrite ou des différentes polices. De la même façon, un schéma relie les lettres pour former un mot, ce qui nous permet de traiter chaque mot comme une seule entité. Ce schéma est alors stocké en mémoire long terme, et peut être utilisé conjointement à la mémoire court terme. L'utilisation fréquente de certains schémas peut alors mener à un phénomène d'automatisation, ce qui permet de libérer de la mémoire de travail pour s'occuper d'autres éléments [Sweller *et al.*, 1998] [Van Merriënboer *et al.*, 2002] [Pollock *et al.*, 2002]. Dans le cadre de cette thèse, il sera important de s'interroger sur la construction de schémas pour la manipulation d'une interface par des gestes, modalité qui n'est pas utilisée de manière fréquente sur les interfaces. Il est alors intéressant de se pencher sur les méthodes de présentation de nouveaux concepts, pour permettre un apprentissage optimisé.

Interactivité

La mémoire de travail est capable de manipuler plusieurs éléments simultanément, comme nous l'avons vu précédemment. Au-delà du nombre d'éléments manipulables, il est nécessaire de s'interroger sur l'interactivité qu'il existe entre ces éléments [Van Merriënboer et Sweller, 2005]. Si les éléments ne sont pas liés entre eux, il est alors possible de les étudier en série, engendrant une charge cognitive moyenne. Au contraire, une forte interactivité nécessite de traiter l'ensemble des éléments simultanément. Si l'apprenant n'a pas développé les schémas cognitifs correspondants, cette grande interactivité peut mener à une charge cognitive élevée, et donc à des difficultés à comprendre et à apprendre. Par contre, si les schémas ont déjà été acquis et sont stockés en mémoire long terme, la charge cognitive globale sera alors moins élevée.

Il est ici intéressant de noter que l'expertise de l'apprenant entre en jeu, à travers la présence ou non de schémas cognitifs concernant une tâche spécifique. Il est alors essentiel de proposer les instructions adaptées à cet apprenant pour optimiser le processus. Les instructions qui fonctionnent pour des novices peuvent conduire à un effet neutre voire négatif sur les personnes expertes : c'est l'effet d'inversion d'expertise – ou *expertise reversal effect* [Kalyuga *et al.*, 2003]. Après avoir testé notre interface gestuelle, il peut être intéressant d'imaginer une évolution des retours sensoriels, probablement par des sons allégés, pour éviter cet effet et ne pas détourner des utilisateurs experts du système.

Un paradoxe

La théorie de la charge cognitive indique qu'il est important, pour des éléments avec une forte interactivité, de traiter l'ensemble simultanément dans la mémoire de travail. Comme nous l'avons vu, l'acquisition de schémas permet de diminuer la charge cognitive induite par ce processus.

En revanche, lors d'une première exposition à un problème complexe ayant de nombreux éléments interconnectés, la capacité de l'apprenant peut être trop faible par rapport à la charge cognitive à mobiliser pour comprendre le problème dans sa globalité. La marche à suivre pour diminuer sa charge cognitive est alors d'utiliser un schéma cognitif, qu'il ne possède pas et qu'il ne peut pas créer à cause d'une surcharge cognitive : il y a alors paradoxe.

Pollock *et al.* ont étudié ce paradoxe dans une série d'expérience comparant deux groupes d'apprenant : un premier groupe dispose dans premier temps de fragments d'information isolés, et leur interactivité est montrée dans une seconde phase ; alors que l'ensemble de l'information est disponible directement pour le second groupe [Pollock *et al.*, 2002]. Ces expériences ont montré que dans certains cas, le mécanisme d'apprentissage est facilité par la séparation d'information en parties peu interactives. Dans d'autres cas, l'interactivité est essentielle à la compréhension, et le partage de l'information nuit à la compréhension générale. Pour contourner le problème du paradoxe, Pollock *et al.* proposent de supprimer la nécessité de prendre en compte tous les éléments en interaction. Les apprenants ont alors le temps pour développer les schémas nécessaires à la poursuite de leur étude.

Dans le cas de notre interface gestuelle, les éléments qui seront mis en jeux présenteront une forte interactivité, ce qui est le but recherché avec la sonification d'un objet virtuel manipulé par des gestes. Pour éviter une surcharge cognitive lors de la première prise en main de l'interface, il sera alors intéressant de la présenter à l'arrêt, pour expliquer les différentes manipulations. Après une période d'entraînement, nous supposons que les sujets auront pu développer des schémas cognitifs permettant d'ajouter la tâche de conduite sans atteindre une surcharge cognitive.

La charge cognitive essentielle

La charge cognitive *essentielle* – ou *germane cognitive load* – réfère à l’effort mobilisé pour la construction de schémas, qui permettent de transférer des éléments de la mémoire de travail à la mémoire long terme. Il est donc primordial que l’apprenant y consacre une part importante de sa charge cognitive. La charge cognitive essentielle est additive avec les charges intrinsèques et extrinsèques. Il est possible de résumer les différents types de charges cognitives lors de l’accomplissement d’une tâche par la figure 3.1.

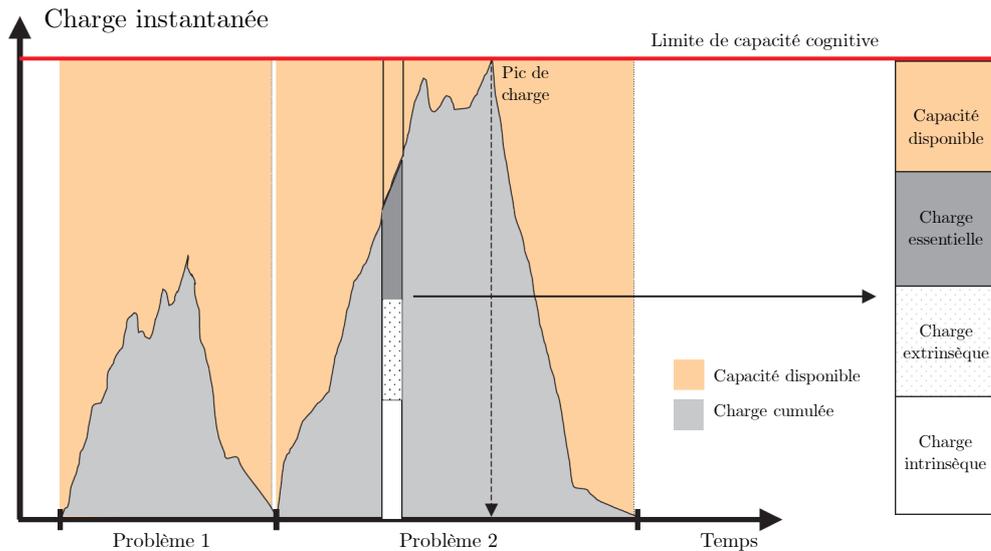


FIGURE 3.1 – Résumé des différentes catégories de charges cognitives évoluant avec le temps, figure adaptée de [Paas *et al.*, 2003].

Lors d’une tâche, la charge cognitive cumulée doit donc être inférieure à la limite de capacité cognitive, tout en réduisant au maximum la charge cognitive extrinsèque, et en laissant à la charge cognitive essentielle la capacité de construire les schémas cognitifs indispensables à l’apprentissage. Les processus d’acquisition et d’utilisation de schémas cognitifs sont résumés en figure 3.2.

Dans le cadre de cette thèse, il sera tout d’abord nécessaire de laisser suffisamment de temps aux sujets novices pour comprendre le fonctionnement de notre interface et s’y habituer. Cette première phase permettra la construction d’un schéma cognitif grâce à la charge cognitive essentielle, comme illustré figure 3.2a. L’utilisation de l’objet virtuel comme métaphore d’un véritable objet peut alors permettre d’accélérer la création de ce schéma, en adaptant le schéma cognitif de la manipulation réelle des objets. C’est l’une des raisons qui nous a poussé à introduire ce concept d’objet virtuel. Par la suite, ce schéma cognitif pourra être utilisé lors de la tâche de conduite pour réduire la complexité d’un problème comme le montre la figure 3.2b, ce qui permettra de minimiser la charge cognitive liée à l’interface, et de permettre une conduite sécuritaire. La charge cognitive totale pourra alors être maintenue sous la limite de la surcharge cognitive.

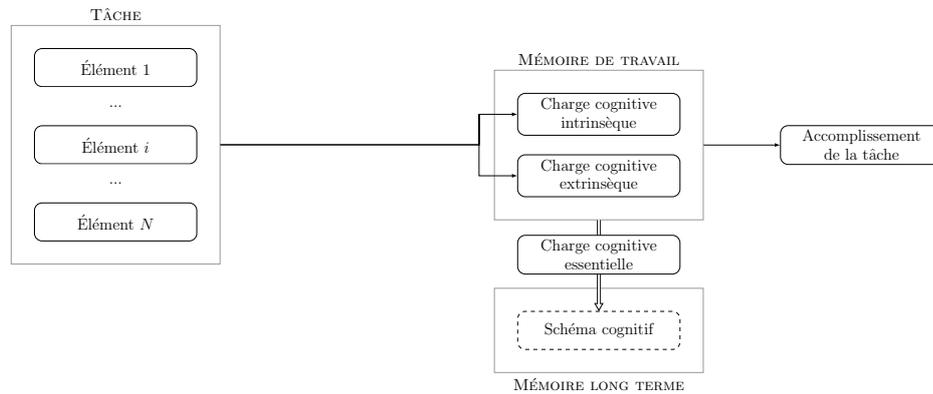
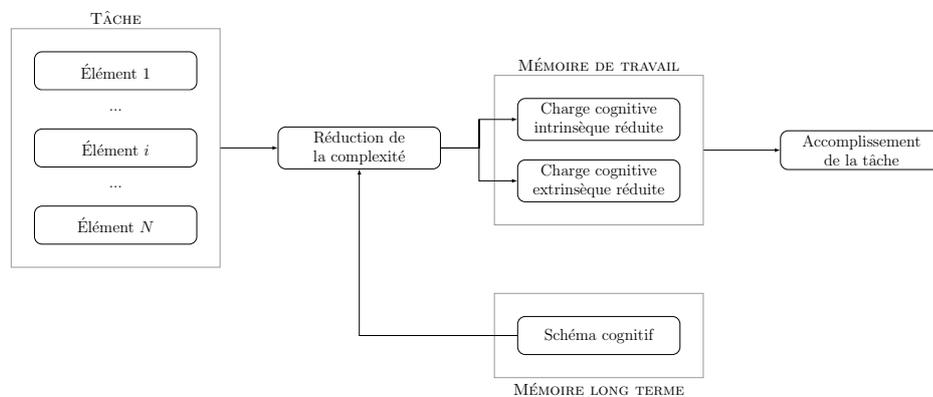
(a) Pour une tâche inconnue avec N éléments : création d'un schéma cognitif(b) Pour une tâche connue avec N éléments : utilisation d'un schéma cognitif

FIGURE 3.2 – Mécanismes intervenant lors de la présentation d'une tâche.

Pour vérifier ces hypothèses lors du test de l'interface gestuelle, il faut alors arriver à estimer la charge cognitive globale des sujets. Nous allons donc maintenant nous intéresser aux méthodes de mesures existantes, pour choisir celles qui conviendront le mieux pour l'installation que nous mettrons en place.

3.1.3 Techniques de mesure de la charge cognitive

Comme nous l'avons vu précédemment, la charge cognitive est un facteur déterminant dans l'acquisition de nouvelles connaissances, par la constitution de schémas cognitifs grâce à la charge cognitive essentielle. La charge cognitive extrinsèque, liée à la présentation de la tâche, est souvent une donnée que l'on cherche à minimiser alors que la charge cognitive totale doit être maintenue sous un certain seuil.

Même s'il n'existe pas de mesure absolue, plusieurs indicateurs peuvent permettre une estimation de la charge cognitive totale d'un utilisateur en se basant sur des données physiologiques, sur des performances ou sur une estimation subjective par l'utilisation d'échelles.

Les données physiologiques

Dans le cas des mesures physiologiques pour l'estimation de la charge cognitive, l'hypothèse est qu'une forte concentration d'un utilisateur va entraîner un état de stress, qui va ensuite affecter ses variables physiologiques. Plusieurs données peuvent être mesurées : le rythme cardiaque, l'activité cérébrale ou la dilatation de la pupille par exemple. Les mesures des données physiologiques semblent conduire à des résultats variables, parfois peu convaincants pour le rythme cardiaque [Paas et Van Merriënboer, 1994b], d'autres fois plus significatifs [Beatty, 1982] [Beatty et Lucero-Wagoner, 2000]. Une revue de différentes techniques de mesures physiologiques a été bâtie par De Waard, notamment dans un contexte automobile [De Waard, 1996].

L'utilisation de ces données nécessite d'équiper les sujets par différents capteurs, pouvant être gênants dans l'exécution de gestes notamment. Cette gêne potentielle, combinée à des mesures parfois peu sensibles à la charge cognitive, nous ont poussé à envisager d'autres techniques d'estimation de la charge cognitive, qui sont décrites dans les paragraphes suivants.

Les performances

Il est également possible d'évaluer la charge cognitive par les performances d'accomplissement d'une tâche. La mesure n'est alors pas continue comme le pouvaient être les données physiologiques, mais reflète la charge cognitive globale à un instant donné.

Pour une tâche unique, l'hypothèse est que la performance de manipulation décroît avec l'augmentation de la charge mentale, comme le montre la figure 3.3. Il est alors possible de distinguer 3 zones : dans la zone A, l'augmentation de la charge cognitive n'entraîne pas de baisse de performances ; la zone B correspond à une dégradation des performances avec la demande ; dans la zone C, les performances ne baissent plus et sont arrivées à un niveau plancher.

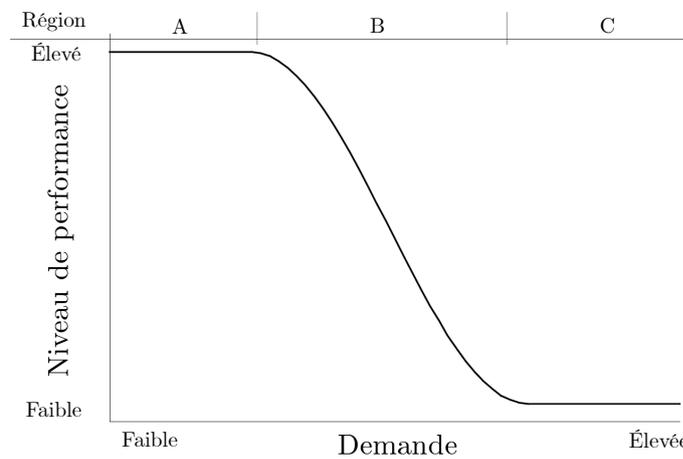


FIGURE 3.3 – Performances en fonction de la demande, figure tirée de [Meister, 1976].

Dans ce cas de figure, la littérature parle alors de mesure de tâche principale – ou *primary task measurement*. La performance est traditionnellement mesurée par une précision ou une note. Cependant, il n'est pas possible dans ce cas de dissocier un utilisateur avec une charge cognitive moyenne d'un autre à la limite de la surcharge cognitive.

Avec l'hypothèse d'une charge cognitive globale limitée [Yeh et Wickens, 1984], une solution pour les distinguer est de proposer une deuxième tâche, et d'en mesurer la performance. On parle alors de mesure de tâche secondaire – ou *secondary task measurement*. Ces tâches sont relativement simples, comme la détection de signaux. Cette combinaison de tâches peut être utilisée de deux façons différentes, qui dépendent des instructions qui sont données lors d'un test [De Waard, 1996]. Dans un cas, les performances de la tâche secondaire sont supposées représenter la charge cognitive imposée par la tâche principale et sont donc les données scrutées [Paas *et al.*, 2003]. Dans l'autre cas, ce sont les performances de la tâche principale qui sont étudiées, qui doivent diminuer lors de l'ajout d'une deuxième tâche comparées à la situation avec une seule tâche [Brunken *et al.*, 2003]. Lors de l'utilisation de notre interface gestuelle en conduite, nous allons combiner ces deux approches. En effet la tâche de conduite ne peut pas être considérée comme une tâche distractive pour mesurer la concentration nécessaire aux manipulations de l'IHM, elles sont toutes deux des tâches principales – même si l'aspect sécuritaire peut faire penser le contraire. Notre but est ici d'assurer leur bon accomplissement simultanément. Comme affirmé par Papantoniou *et al.* « la charge cognitive peut être estimée par une mesure de la dégradation des performances des tâches primaire ou secondaire » [Papantoniou *et al.*, 2017]. Nous allons donc à la fois scruter les performances de manipulations de l'IHM pour voir si les conducteurs arrivent à l'utiliser convenablement, mais nous allons aussi observer les tâches liées à la maîtrise du véhicule et au trafic routier pour estimer si la sécurité est garantie lors de la conduite.

Les échelles subjectives

La charge cognitive peut également être évaluée par l'utilisateur lui-même. Il est en effet le mieux placé pour évaluer sa concentration totale, et plusieurs études ont tendance à prouver la fiabilité de cet indicateur [Gopher et Braune, 1984] [Paas, 1992]. C'est de son côté subjectif que réside l'intérêt de cette source d'informations [Muckler et Seven, 1992]. Comme pour les performances, cette estimation n'est pas continue et donne une estimation globale pour une tâche donnée.

Il existe plusieurs types d'échelles, qui peuvent être multidimensionnelles en prenant en compte plusieurs facteurs – fatigue, effort mental, frustration, par exemple. L'indice NASA-TLX – pour *Task Load indeX* – est très utilisé, tout comme l'indicateur SWAT – pour *Subjective Workload Assessment Technique* ou encore l'indice MCH – pour *Modified Cooper-Harper scale* [Hart et Staveland, 1988] [Reid *et al.*, 1981] [Wierwille et Casali, 1983]. Dans le milieu automobile, un indice a également été mis au point par le Groupe PSA, le PSA-TLX, qui a pour but d'évaluer la concentration du conducteur avec une tâche secondaire ou des aides à la conduite [Chin *et al.*, 2006]. Plusieurs échelles sont au contraire unidimensionnelles permettant une évaluation plus rapide, comme le RSME – pour *Rating Scale of Mental Effort* – proposée par Zijlstra, ou l'échelle d'activation citée par De Waard [Zijlstra, 1985] [Zijlstra, 1993] [De Waard, 1996]. Dans un contexte de conduite, ces échelles semblent être particulièrement sensibles aux variations de l'état du conducteur [De Waard, 1996]. Le choix entre les échelles unidimensionnelles ou multidimensionnelles repose essentiellement sur le but de l'étude. S'il est nécessaire de déterminer la charge cognitive selon plusieurs critères, alors les premières conviennent tout à fait. Si au contraire une représentation globale de la charge cognitive est nécessaire, alors les échelles multidimensionnelles sont à privilégier.

Dans l'application que nous visons, l'estimation de la charge cognitive globale lors d'un parcours de conduite semble adapté. De plus, la procédure de mesures basées sur des échelles multidimensionnelles comme le NASA-TLX est relativement longue, ce qui peut augmenter la fatigue de sujets fortement sollicités. Dans le cadre de notre étude, nous nous orienterons donc vers les échelles unidimensionnelles, permettant une estimation simple et rapide de la charge cognitive globale.

Malgré le nombre de mesures possibles de la charge cognitive, il faut garder à l'esprit qu'elles ne sont que des indicateurs de cette charge, et il n'est pas possible d'avoir une mesure parfaite. La combinaison de plusieurs de ces indices peut être un moyen d'estimer plus précisément la charge cognitive d'un conducteur [Paas et Van Merriënboer, 1993].

Les différentes notions que nous avons abordées dans cette section permettent de connaître les mécanismes d'apprentissages qui peuvent se répercuter sur la concentration des utilisateurs. Après avoir permis la création de schémas cognitifs, il sera intéressant d'appliquer certaines techniques de mesures de la charge cognitive lors du test de notre interface gestuelle.

La manipulation d'interfaces conjointe à la conduite peut cependant perturber les conducteurs, et diminuer leur vigilance comme nous allons le voir. Ces phénomènes distractifs ne se répercutent pas de la même manière selon les modalités sollicitées, et leur étude va permettre de sélectionner des variables de conduite d'intérêt.

3.2 Les phénomènes distractifs en conduite

La conduite est une activité complexe, demandant une forte concentration. Le conducteur doit en effet effectuer des mouvements coordonnés de ses bras et de ses jambes pour à la fois maîtriser la direction prise par le véhicule, son régime moteur ou son accélération. Il doit de plus s'assurer en permanence qu'il respecte le code de la route : circuler au centre de la voie, dans le bon sens de circulation, à une vitesse autorisée, et tenir compte des panneaux présents sur le bord de la route. Il est également impératif de prendre en compte les autres usagers de la route, et notamment ceux présents devant lui pour maintenir une distance de sécurité permettant un freinage d'urgence sans accident. L'attention des automobilistes est aussi nécessaire pour pouvoir réagir à certains événements imprévus : un piéton qui traverse sans regarder, ou un freinage brusque du conducteur du véhicule situé devant par exemple. Une grande partie des informations de contrôle de l'environnement extérieur au véhicule est capté par la modalité visuelle, qui est fortement sollicitée.

Pour pouvoir accomplir l'ensemble de ces tâches dans de bonnes conditions, les conducteurs anticipent certains événements, pour répartir temporellement la charge cognitive globale qu'il est nécessaire de mobiliser. De plus, l'expérience de conduite leur permet de développer des schémas cognitifs. Ce procédé permet d'« automatiser » certaines tâches, ce qui peut permettre aux conducteurs expérimentés de garantir une plus grande sécurité comme l'ont montré plusieurs études : les conducteurs novices ont tendance à regarder moins loin [Quenault et Parker, 1973], à avoir une distance inter-véhiculaire plus faible [Evans et Wasielewski, 1983] et peuvent mettre plus de temps à réagir à un danger que des conducteurs aguerris [Quimby et Watts, 1981]. Ces résultats sont cependant à modérer, puisqu'une étude de Duncan *et al.* a semblé montrer un effet contraire, faisant possiblement ressortir l'effet d'une confiance surestimée [Duncan *et al.*, 1991].

Comme nous l'avons vu dans la section 2.2.3, l'attention est un facteur important à prendre en compte dans la perception d'une situation. C'est donc également le cas dans un véhicule, où le conducteur n'a pas son attention focalisée en permanence sur la tâche de conduite. Selon Regan *et al.*, « l'inattention a lieu dans un grand nombre de situations dans lesquelles le conducteur n'arrive pas à atteindre la concentration nécessaire à la conduite » [Regan *et al.*, 2008]. Le conducteur peut en effet être amené à être distrait par un grand nombre d'éléments : une particularité du paysage peut l'amener à détourner les yeux ; une conversation avec un passager ; ou encore la manipulation de l'interface centrale du véhicule. Ce manque d'attention peut avoir des répercussions importantes

sur la conduite. Une étude datant de 2006, qui se base sur plus de 42000 heures de vidéos enregistrées sur 12 mois auprès de plus de 100 conducteurs, montre que 78% des accidents sont liés à un phénomène d'inattention du conducteur [Dingus *et al.*, 2006]. Quatre types d'inattentions pour les conducteurs ont pu être distingués :

- l'accomplissement de tâches secondaires ;
- la distraction liée à la tâche de conduite, mais qui détourne le champ de vision du conducteur, surveiller sa vitesse ou regarder dans un angle mort par exemple ;
- la somnolence ;
- les mouvements des yeux non reliés à l'activité de conduite.

Parmi les accidents impliquant un manque d'attention, environ 45% sont liés à l'accomplissement d'une tâche secondaire, source la plus importante. Ces données montrent bien l'importance que revêt l'étude de l'inattention impliquée par des tâches secondaires en conduite. Ces tâches sont par exemple liées à des manipulations d'interfaces de médias ou de ventilation, ainsi qu'à l'utilisation de téléphones. Le nouveau type d'IHM que nous voulons proposer doit donc prendre le critère de l'attention des conducteur en compte, l'utilisation conjointe de gestes et de sons vise précisément à alléger la concentration nécessaire pour manipuler l'interface.

De nombreuses études ont pu être menées dans des situations de conduite réelle. Cependant, les études portant sur l'attention des conducteurs peuvent mener à des situations dangereuses pour les occupants du véhicule, et sont souvent lourdes à mettre en place avec des automobiles équipées. L'utilisation de simulateurs de conduite peut être une alternative intéressante, qui présente des garanties sécuritaires. La reproduction de conditions de conduite identiques pour chaque sujet est par exemple aisée en simulateur, paramètre particulièrement important pour les études perceptives. Il sera cependant primordial de garantir la validité de l'utilisation d'un simulateur, pour que les résultats puissent en partie être extrapolés pour une situation de conduite réelle. Nous allons tout d'abord nous intéresser aux simulateurs de conduite, en se focalisant sur leurs avantages notamment pour mettre en place des tests perceptifs, mais aussi à leurs inconvénients. Nous aborderons ensuite trois types de distractions qu'il convient de distinguer, selon les modalités mises en jeu. Ces situations vont montrer des effets différents sur l'attention du conducteur, que nous aborderons au travers d'études réalisées en partie en simulateur de conduite. Plusieurs paramètres de conduite sensibles aux effets distractifs seront identifiés ; il sera intéressant de scruter les variations de ces paramètres lors du test de l'interface gestuelle.

3.2.1 Le simulateur de conduite : avantages et limites

Les simulateurs de conduite sont utilisés par les constructeurs automobiles pour des études portant sur les comportements du conducteur, les aides à la conduite, ou encore l'ergonomie d'interfaces. Ils possèdent un grand nombre d'avantages pour les études théoriques notamment, mais aussi certaines limites comme le relève Fisher *et al.* [Fisher *et al.*, 2011]. Leur utilisation permet par exemple une reproductibilité parfaite des scénarios de conduite, paramètre difficilement maîtrisable sur une route avec trafic. Cet atout est particulièrement utile pour les études perceptives, visant à comprendre les réactions de sujets à la présentation de stimuli identiques. Les simulateurs permettent également un gain de temps dans l'avancement des projets, une réduction des coûts, et l'enregistrement d'un grand nombre de variables de conduite – angle du volant, écart latéral ou vitesse par exemple – de manière aisée. Enfin, ils autorisent des manipulations en conduite qui sont

dangereuses sur route, sans risquer la vie du conducteur et de ses passagers. Ce dernier point est particulièrement intéressant, et une raison importante justifiant son utilisation dans notre cas pour tester l'interface gestuelle. L'ensemble de ces avantages font des simulateurs des outils utiles pour des études en contexte automobile, utilisés par exemple dans de nombreuses recherches menées au Groupe PSA [Stratulat, 2011] [Sciabica, 2011] [Denjean, 2015] [Carbajo *et al.*, 2015] [Savona, 2016] [Jacobé de Naurois *et al.*, 2017].

Cependant, les simulateurs de conduite ne permettent qu'une reproduction simplifiée de la réalité, et « même les simulateurs permettant des retours dynamiques [...] ne peuvent pas reproduire les conditions réelles de conduite » selon Shechtman *et al.* [Shechtman *et al.*, 2009]. En prenant en compte ces limites, il est alors nécessaire de s'interroger sur la fidélité d'un simulateur. Mudd propose deux critères de jugement concernant les simulateurs de vol, qui ont ensuite été repris par Blaauw pour les simulateurs automobiles : la correspondance du comportement psychologique de l'utilisateur entre les situations simulées et réelles, et la correspondance du simulateur avec un véhicule roulant d'un point de vue physique, permettant de reproduire les mêmes stimuli [Mudd, 1968] [Blaauw, 1982]. Ces deux versions de la fidélité d'un simulateur amènent naturellement vers la notion de validation : comment juger qu'un simulateur est fidèle à la réalité ? L'enjeu ici est de déterminer les critères permettant d'affirmer que des comportements observés sur un simulateur sont extrapolables à une situation de conduite réelle.

Le critère de validité

Les deux critères proposés précédemment interrogent sur la validité d'utilisation d'un simulateur : le simulateur doit-il reproduire physiquement les comportements d'un véhicule pour être valide ? Dans notre cas d'études de facteurs humains, où les processus perceptifs et cognitifs des utilisateurs sont au centre de nos préoccupations, la correspondance dans les comportements psychologiques des utilisateurs sera plus importante pour la validité de l'utilisation d'un simulateur ; la validité physique n'est donc pas nécessaire dans notre cas. Rolfe *et al.* affirment que « la valeur d'un simulateur dépend de sa capacité à provoquer le même type de réponse chez l'opérateur que dans une situation réelle » [Rolfe *et al.*, 1970]. Un bon exemple est la reproduction du sentiment d'accélération dans un simulateur de conduite dynamique, en utilisant la perception ambiguë des organes otolithiques que nous avons abordée en section 2.2. Il est alors possible de proposer une combinaison des accélérations latérale et angulaire pour recréer une sensation réaliste d'un point de vue perceptif, sans pour autant recréer réellement cette accélération [Stratulat *et al.*, 2011]. Pour valider d'un point de vue perceptif l'utilisation d'un simulateur, il faut donc que les conducteurs aient des retours sensoriels comparables entre les situations réelles et simulées : le réalisme est alors une donnée importante lors de l'utilisation d'un simulateur de conduite. Il est donc nécessaire de donner des retours multisensoriels au conducteur, pour proposer des situations les plus écologiques possibles. Cependant, les stimuli présentés aux sujets sont rarement les mêmes que dans un véhicule : au niveau de la vision par exemple, il est difficile de proposer une définition suffisante en trois dimensions pour que l'utilisateur ait l'impression d'être dans une situation de conduite réelle.

Plusieurs études ont également introduit les termes de validité *absolue* ou *relative* [Blaauw, 1982] [Kaptein *et al.*, 1996]. La validité d'un simulateur peut être étudiée en comparant des paramètres de la condition réelle de conduite avec ces mêmes paramètres en simulateurs. Il est également possible de comparer des différences de performances entre deux situations en conduite avec deux situations similaires en simulateur [Blaauw, 1982]. Un simulateur a une validité relative si les valeurs de ces paramètres sont du même ordre et suivent la même évolution qu'en situation réelle. Dans ce cas, le simulateur permet d'observer une tendance globale. Si additionnellement à cette tendance, les

valeurs observées en simulateur sont approximativement les mêmes qu'en situation réelle, alors le simulateur possède une validité absolue. Un simulateur ayant une validité absolue semble être la condition la plus satisfaisante pour une étude. Cependant une validité relative est suffisante dans le cadre de recherches, puisque sont étudiés presque exclusivement des effets entre variables indépendantes [Törnros, 1998]. Après une première observation en simulateur, une confirmation peut ensuite être apportée en situation réelle de conduite, avec une estimation plus précise.

Kaptein *et al.* évoquent également l'influence de la tâche dans la validité de l'utilisation d'un simulateur [Kaptein *et al.*, 1996]. Un simulateur de conduite peut être fidèle dans la reproduction de certains stimuli perceptifs sans qu'il ait une validité globale. Il est donc nécessaire de se poser la question de la validité d'un simulateur par rapport aux questions posées par l'étude, pour se concentrer sur les stimuli les plus importants. La manipulation de l'interface gestuelle et sonore tout en conduisant entraîne des questionnements par rapport à l'attention du conducteur, et ne nécessite donc pas impérativement de système de restitution des effets dynamiques de l'habitacle.

Cas particulier de l'attention et de la perception du danger

Étudier l'attention du conducteur en situation de conduite réelle présente des risques de provoquer des accidents, et donc d'engendrer des dégâts humains et matériels. L'utilisation d'un simulateur de conduite semble alors particulièrement adaptée.

Wang *et al.* ont comparé des mesures effectuées en simulateur et avec véhicule instrumenté lors de trois types de manipulations d'une interface en conduite [Wang *et al.*, 2010]. Des effets comparables sont observés concernant les temps de manipulation de l'interface, la fréquence et le temps de regard, mais les paramètres de conduite comme la vitesse ou l'écart type de l'écart latéral se sont révélés différents entre le simulateur et la condition réelle : les sujets ont tendance à aller moins vite en simulateur – mais cet effet peut venir de la vitesse constante des autres véhicules du trafic – et à faire des écarts latéraux de plus grande amplitude. Cette étude semble montrer qu'une comparaison effectuée en simulateur entre les interfaces actuelles et l'interface gestuelle que nous voulons mettre au point peut se retrouver en condition réelle, même si les paramètres de conduite sont à prendre avec précaution. Il peut également être intéressant de prendre en compte une situation de contrôle sans manipulation d'interface, pour observer si l'introduction d'une tâche distractive entraîne des effets comparables en simulateur ou en situation réelle. Une étude menée par Santos *et al.* s'intéresse à cette question, dans trois conditions de simulation différentes : en laboratoire, en simulateur de conduite ou en condition réelle [Santos *et al.*, 2005]. La condition laboratoire propose un simulateur simple, avec un volant et des pédales reliées à un ordinateur, où le retour visuel est donné par un unique moniteur. Le simulateur utilisé pour cette étude est statique, et il est bâti à partir d'une Rover 216GTi. Enfin, la condition réelle utilise un véhicule instrumenté, enregistrant toutes les informations nécessaires à l'étude. Quatre conditions de manipulations sont proposées : sans tâche secondaire, et avec trois niveaux de difficultés de manipulations distractives. Des effets similaires sont perçus dans les trois simulations proposées entre les situations de conduite normale et avec tâche secondaire : une diminution de la vitesse – cet effet compensatoire sera abordé en section 3.2.5 – ou une perception des performances de conduite dégradée. Les résultats concernant les paramètres d'écart latéral ou les performances sur l'interface secondaire ne sont en revanche pas disponibles pour comparer les simulations à la conduite en véhicule instrumenté.

Les simulateurs de conduite semblent être une solution intéressante pour l'étude des effets attentionnels pendant la conduite [Mullen *et al.*, 2011] [Fisher *et al.*, 2011]. Ils permettent de mener des recherches sur les manipulations d'interfaces, tout en garantissant la sécurité des utilisateurs, ce qui nous motivera à faire ce choix, comme nous le verrons dans le chapitre 6. Comme l'étude de Santos *et al.* semble le montrer, même un simulateur simple peut permettre d'étudier l'attention liée à une tâche secondaire [Santos *et al.*, 2005]. Cependant, les effets liés à la perception du danger peuvent être diminués en simulateur, ce qui peut avoir des effets sur les études menées : les conducteurs peuvent avoir tendance à consacrer plus de concentration sur la tâche secondaire en simulateur qu'ils ne le feraient en conduite. Il est raisonnable de penser que l'augmentation du réalisme et de l'effet immersif des simulateurs, dont de nombreux points peuvent encore être améliorés [Green, 2005], peut réduire cette différence de perception du danger.

Le simulateur de conduite peut donc permettre une première étude des effets distractifs, dont les conclusions seront à confronter avec des études en situation de conduite réelle. Comme nous l'avons évoqué précédemment, les effets attentionnels sur la conduite peuvent être différents selon les modalités sollicitées. L'utilisation du téléphone en conduite requiert par exemple l'attention du conducteur et parfois une de ses mains lorsqu'il n'est pas utilisé avec un kit mains-libres : un appel introduit donc des distractions cognitives, ou motrices et cognitives. L'utilisation d'un écran tactile introduit quant à lui des distractions visuo-motrices. Ces effets attentionnels ont des répercussions différentes selon les modalités sollicitées, que nous allons à présent détailler.

3.2.2 Distractions motrices

La première catégorie concerne de manière évidente les manipulations nécessitant une main, qui n'est plus disponible pour la tâche de conduite. Ce type de distraction est rarement étudié de manière isolée, mais est plutôt combiné avec des distractions visuelles ou cognitives, comme nous le verrons dans les deux parties suivantes. En ce qui concerne les manipulations d'interfaces, cette distraction est présente que l'IHM soit basée sur des boutons, des technologies tactiles ou gestuelles. Certains effets peuvent cependant être ajoutés avec l'utilisation de gestes plus complexes que le simple appui sur un bouton physique ou virtuel.

3.2.3 Distractions visuelles

Les distractions visuelles dans les véhicules actuels proviennent par exemple de l'utilisation grandissante des interfaces tactiles. Comme nous l'avons évoqué précédemment, la tâche de conduite nécessite de nombreuses informations accessibles par le visuel. L'utilisation conjointe de la modalité visuelle pour la conduite et pour une tâche secondaire peut donc provoquer des interférences dans l'accomplissement simultané de ces deux tâches dans de bonnes conditions. Par exemple, lorsque la difficulté de la tâche secondaire augmente, les conducteurs regardent moins la route, car concentrent leur vue plus longtemps et plus souvent sur la tâche [Green, 1999]. Ce comportement peut alors dégrader leur capacité d'anticipation d'événements pouvant être dangereux. En revanche lorsque la tâche de conduite devient plus difficile, les conducteurs ont tendance à regarder la route plus longtemps ou à ralentir [Senders *et al.*, 1967] [Wierwille, 1993] [Green, 1999]. Tsimhoni et Green étudient l'influence d'une tâche secondaire visuelle par rapport à une situation de contrôle sans visuel, et ils ont montré que l'écart-type de l'écart latéral sur la route est significativement plus élevé avec une tâche secondaire [Tsimhoni et Green, 2001]. Les conducteurs ont donc tendance à avoir de plus grandes variations d'écart latéral en présence de tâche secondaire visuelle.

Tsimhoni *et al.* se sont également intéressés aux manipulations d'interfaces tactiles à l'arrêt ou en conduite [Tsimhoni *et al.*, 2004]. L'écriture d'une adresse sur un écran tactile grâce à un clavier virtuel prend par exemple 32,5 secondes en moyenne à l'arrêt et 83,9 secondes en conduite, soit une multiplication par plus de 2,5 entre les deux conditions. Des écarts significatifs sont constatés entre la situation de conduite de contrôle – sans tâche secondaire – et lors de l'écriture de l'adresse tout en conduisant. L'écart-type de l'écart latéral, ou de l'angle au volant sont par exemple significativement différents. De plus, les sujets de cette étude ont tendance à ressentir des difficultés par cette sollicitation importante du visuel, traduites dans des notes subjectives de difficulté de la tâche. Cette étude témoigne bien des difficultés éprouvées par des manipulations d'un écran tactile en conduite, provenant de la sur-sollicitation de la modalité visuelle du conducteur. Dans une autre étude, Crandall et Chaparro ont comparé les scénarios de composition d'un message avec un clavier physique ou virtuel [Crandall et Chaparro, 2012]. Le retour haptique des boutons physiques permet de diminuer significativement la déviation latérale du véhicule, même si la condition de contrôle sans distraction obtient les meilleures performances. De la même manière, le nombre d'erreurs de frappe est significativement diminué, et les sujets perçoivent des sollicitations mentales et physiques significativement plus importantes avec le clavier virtuel de l'écran tactile qu'avec un clavier physique. Le retrait du retour d'information haptique du clavier physique est donc à l'origine d'une dégradation de performances dans l'accomplissement de tâches simultanées. Grâce aux retours sonores de notre interface gestuelle, nous désirons rétablir ce canal d'information pour permettre une conduite sécuritaire, et des manipulations de l'interface aisée.

Cette dernière étude permet de souligner l'importance du retour haptique dans la manipulation d'une interface en conduite. L'introduction d'écrans tactiles dans les véhicules pour la manipulation d'interfaces supprime ce retour sensoriel, au profit d'un retour visuel. La tâche de conduite nécessite cependant une attention visuelle soutenue, ce qui peut conduire à une sur-sollicitation du visuel des conducteurs et donc à une conduite ou des manipulations d'interface dégradées. L'interface gestuelle et sonore que nous voulons proposer doit permettre, par l'utilisation pertinente de sons, de rétablir un retour d'information provenant de l'IHM, et ainsi corriger ces problèmes de surcharge de la modalité visuelle. Le conducteur peut accomplir plus facilement des tâches simultanées, ce qui peut permettre une conduite sécuritaire, et des manipulations de l'interface aisée. Cependant notre interface peut amener une distraction cognitive, liée au phénomène de « cécité attentionnelle » que nous avons abordé en section 2.2.3.

3.2.4 Distractions cognitives

Les distractions cognitives – qui englobent également les distractions auditives – concernent des actions faisant appel à l'attention du conducteur sans pour autant nécessiter de détours du regard ou l'utilisation d'une main. Un exemple classique est l'utilisation du téléphone portable au volant avec un kit mains-libres. Dans ces situations, le conducteur est concentré sur sa conversation, et sa perception des événements liés à la conduite est alors dégradée. Il est alors observé une concentration spatiale du regard, ainsi qu'une réduction de la fréquence des regards vers le compteur de vitesse ou les rétroviseurs [Recarte et Nunes, 2000] [Harbluk *et al.*, 2002]. Une augmentation du temps de réaction est également mise en avant par Horrey et Wickens, dans les cas de conversation au téléphone en le tenant ou avec les mains libres [Horrey et Wickens, 2006]. Plusieurs études ont confirmé des perturbations causées par l'utilisation du téléphone au volant même avec les mains libres [Matthews *et al.*, 2003] [Patten *et al.*, 2004] [Hancock *et al.*, 2003]. Par exemple, Strayer et Drews ont montré que les conducteurs sont 18% plus lents, ont une distance inter-véhiculaire 12%

plus grande et prennent 17% de temps en plus pour retrouver une vitesse cible après un freinage, comparé à une situation de contrôle [Strayer et Drew, 2004]. Une étude a même comparé deux situations qui peuvent paraître proches : une conversation passé au téléphone avec kit mains-libres ou avec un passager [Drews *et al.*, 2008]. Les résultats montrent que les erreurs de conduite sont plus nombreuses lors de la conversation au téléphone. Drews *et al.* font le parallèle avec le fait que la complexité de la conversation avec le passager baisse lorsque la situation de conduite se complexifie. En effet, les conditions de conduite ont une influence directe sur les deux occupants de la voiture, qui adaptent leur conversation en fonction, ce qui n'est pas le cas au téléphone.

Selon la loi française, l'utilisation d'un téléphone « tenu en main par le conducteur d'un véhicule en circulation » est tout d'abord interdit en 2003, à cause de la distraction motrice engendrée. Cet article R412-6-1 du code de la route est ensuite modifié en 2015 pour englober également les oreillettes et kit mains-libres. Il est alors surprenant de constater que la loi n'interdit pas les dispositifs intégrés aux véhicules, générant tout autant de distractions cognitives.

Nous allons à présent nous intéresser à différentes études comparant ces types de distractions. Les écrans tactiles actuellement implantés dans les véhicules engendrent des distractions motrice et visuelle, alors que notre proposition d'interface gestuelle va proposer des distractions motrice et cognitive. Il est donc intéressant de comparer ces deux détournements de l'attention, notamment pour savoir quelles sont les données de conduite qui y sont sensibles.

3.2.5 Comparaison et conséquences sur la conduite

La comparaison rigoureuse des effets des distractions visuelles et cognitives sur la conduite est difficile à mettre en place, même si elle peut être riche en enseignements. Il est en effet malaisé de déterminer deux distractions ayant rigoureusement la même difficulté, pour observer seulement l'effet attentionnel. Les recherches que nous allons étudier sont donc dépendantes du choix des distractions proposées. D'un point de vue pratique, il peut cependant être choisi de comparer des manipulations usuelles dans un véhicule : l'utilisation du téléphone ou de l'interface des médias par exemple. Ce choix peut être justifié par une application directe : expliquer les causes des accidents de voitures liés à des manipulations secondaires à la conduite. C'est le point de vue adopté par la plupart des études sur lesquelles nous allons à présent nous pencher.

Victor *et al.* se sont par exemple intéressés à l'influence du type de distraction sur le regard des conducteurs [Victor *et al.*, 2005]. Il a été observé qu'une concentration du regard survient dans les deux types de distraction : les conducteurs se focalisent sur l'information essentielle de la distance inter-véhiculaire, possiblement pour éviter les accidents. L'augmentation de la difficulté des tâches visuelles secondaires perturbe également les conducteurs : ils regardent moins le centre de la route, et pendant moins longtemps. Cependant, cette influence de la difficulté de la tâche n'est pas retrouvée sur les mouvements des yeux pour les distractions cognitives, qui semblent moins perturbées de ce point de vue.

Engström *et al.* ont également montré que les deux types de distraction peuvent conduire à des effets différents sur les paramètres de conduite, dans une étude menée en simulateur et sur route [Engström *et al.*, 2005]. La tâche visuelle consiste à déterminer si une cible était présente parmi un ensemble de stimuli visuels, et les sujets doivent compter le nombre d'occurrences d'un son cible dans un flux sonore. Avec la tâche visuelle, l'écart-type de l'écart latéral est significativement plus élevé qu'en situation de contrôle, alors que l'effet inverse est observé pour la tâche cognitive. Cet effet témoigne d'un contrôle visuel de la route intermittent avec la tâche visuelle, corrigé par des

mouvements abrupts du volant. Au contraire, la tâche cognitive requérant de l'attention, il est probable que des mouvements oculaires inutiles à la conduite ont été supprimés, pour se concentrer sur la route. Ce même effet est constaté dans des études de Törnros et Bolling [Törnros et Bolling, 2005] ou de Hurwitz et Wheatley [Hurwitz et Wheatley, 2002]. De plus, les conducteurs ont tendance à réduire leur vitesse avec les distractions visuelles, que ce soit en simulateur ou en conduite réelle. La réduction de la vitesse est interprétée comme un effet compensatoire, pour réduire la sollicitation visuelle de la tâche de conduite et ainsi maintenir des performances à un niveau acceptable. Cet effet n'est pas retrouvé dans le cas des distractions cognitives, montrant une meilleure répartition de l'attention dans ce cas.

Une autre étude précise que dans les deux conditions de distraction, les conducteurs ont un temps-avant-collision – ou *time-to-collision* – plus faible qu'en situation de contrôle, ce qui montre le problème sécuritaire d'une tâche secondaire [Hamish Jamson et Merat, 2005]. Lamble *et al.* se sont également penchés sur ce temps-avant-collision avec une tâche de suivi d'un véhicule en conduite réelle sur autoroute [Lamble *et al.*, 1999]. Trois conditions distrayantes sont proposées : une distraction visuelle et motrice venant de la composition de numéros sur un téléphone ; une distraction cognitive par un calcul mental ; et une condition de contrôle sans distraction. L'analyse des temps-avant-collision indique une diminution significative pour la distraction cognitive par rapport à la condition de contrôle, ce qui n'est pas le cas pour la distraction visuo-motrice – $p = 0,056$. La tâche de calcul mental semble donc mener à une situation plus dangereuse que la composition d'un numéro, ce qui est expliqué par Lamble *et al.* par une grande habitude du maniement des téléphones mobiles. En mettant cette étude en parallèle avec l'étude d'Engström *et al.*, il est aussi possible que les conducteurs aient réduit leur vitesse avec la distraction visuelle, ce qui conduit à un temps-avant-collision élevé. Le temps de réaction au freinage – ou *brake reaction time*, défini comme la différence des temps de début de freinage des véhicules suivi et suiveur, donne également des résultats intéressants. Des augmentations significatives de 0,48 et 0,5 seconde sont constatées entre les conditions de contrôle et de distractions visuo-motrices et cognitives. Les deux distractions provoquent donc un effet comparable sur cet indicateur. Globalement, cette étude met en lumière la demande attentionnelle des tâches secondaires, dégradant les performances de conduite en terme de réaction à un danger. Cependant comme nous l'avons évoqué précédemment, les résultats de ces études dépendent fortement des tâches cognitives et visuelles choisies, mais les distractions cognitives semblent parfois moins perturber les conducteurs. Un test en simulateur peut être l'occasion d'observer des différences, comme nous le verrons plus loin dans ce document.

Les études relatées dans cette section montrent des effets disparates selon les tâches proposées aux sujets. Les manipulations secondaires semblent concentrer le regard du conducteur sur le véhicule situé devant lui, pour anticiper au maximum un freinage d'urgence. Pourtant, le temps de réaction au freinage semble augmenter pour des distractions visuo-motrices ou cognitives indifféremment, suggérant que ces différences proviennent de l'exécution même d'une tâche secondaire. Les distractions visuelles semblent cependant mener les sujets à ralentir leur véhicule, à la différence des tâches cognitives. Ce phénomène semble compenser le manque d'informations visuelles qu'ont les sujets sur la situation de conduite : ralentir leur véhicule permet de réduire le flux d'informations visuelles, et donc la nécessité d'utiliser la vue pour la conduite, ce qui permet in-fine de faire les tâches secondaires convenablement. Les paramètres de conduite comme la vitesse ou l'écart latéral – cités dans de nombreuses études – seront donc des variables intéressantes à enregistrer lors du test de notre interface en situation de conduite.

Conclusion du chapitre

L'utilisation en situation de conduite d'une nouvelle interface pilotée avec des gestes et proposant des retours sonores soulève plusieurs questions : comment les utilisateurs pourront-ils la prendre en main dans les meilleures conditions ? Comment mesurer les efforts de concentration consentis ? Du point de vue de la conduite, quels sont les paramètres impactés par son utilisation ? Comment réagissent les conducteurs à différents types de distractions ?

Pour répondre à ces questions, plusieurs définitions de la théorie de la charge cognitive ont tout d'abord été présentées, notamment concernant les mémoires de travail et long terme. Ces notions sont en effet nécessaires pour comprendre l'automatisation de certaines tâches grâce à des schémas cognitifs, permettant une baisse de la concentration globale du conducteur après plusieurs utilisations d'un système. Il est en effet nécessaire que la charge cognitive des utilisateurs de l'interface gestuelle que nous voulons proposer ne soit pas trop élevée, pour permettre une conduite sécuritaire. Lors du test de l'interface gestuelle, il sera donc impératif de mettre en place des mesures de la charge cognitive et de les comparer à une situation de contrôle, pour évaluer les répercussions de l'utilisation de ce nouveau type d'IHM sur la concentration mentale des conducteurs.

L'impact de tâches distrayantes a ensuite été étudié, en se concentrant tout d'abord sur l'apport des simulateurs de conduite dans les études perceptives. La validité de ces simulateurs est relative, mais permet néanmoins d'observer les effets attentionnels liés à l'utilisation d'une interface conjointement à la conduite sécuritaire d'un véhicule. Les actions perturbatrices pour la tâche de conduite, comme les managements des IHM par exemple, ont ensuite été discriminées selon les modalités mises en jeu. Les interfaces actuelles, par l'utilisation d'écrans tactiles, sollicitent la vue des conducteurs et nécessitent une main pour être utilisées. L'interface gestuelle que nous proposons d'étudier introduit quant à elle des distractions motrice et cognitive. Nous avons donc déterminé certains aspects décrivant les répercussions de ces types de distractions sur la conduite, pour anticiper l'enregistrement des paramètres pertinents lors du test de notre interface. Les vitesses de conduite observées par Engström *et al.* sont par exemple différentes pour les distractions visuelles et cognitives [Engström *et al.*, 2005]. Ce paramètre ainsi que l'écart latéral par exemple peuvent être particulièrement intéressants à scruter.

Deuxième partie

Contribution expérimentale

Construction d'un objet virtuel sonore

Les interfaces tactiles présentes dans les véhicules automobiles peuvent poser des problèmes dans leur utilisation en conduite. Comme nous l'avons vu, les distractions visuelles peuvent amener à une baisse de l'attention globale du conducteur. Des effets compensatoires ont même été observés, notamment au niveau de la vitesse du véhicule : ces situations peuvent être dangereuses pour les occupants du véhicule. Nous avons également vu que les distractions cognitives des conducteurs, comme l'utilisation d'un téléphone avec kit mains-libres, dégradent la qualité de la conduite. Cependant, une meilleure répartition des informations entre les différentes modalités lors des manipulations de l'IHM peut permettre une baisse de la charge cognitive du conducteur, qui pourra donc allouer une plus grande concentration à la tâche de conduite [Penney, 1989] [Mousavi *et al.*, 1995] [Tindall-Ford *et al.*, 1997]. L'association de gestes et de retours sonores pour interagir avec l'interface vise à diversifier les sens utilisés, pour que les yeux des conducteurs puissent se concentrer sur la route.

Nous faisons alors l'hypothèse que l'utilisation d'un objet virtuel (OV) peut avoir un grand nombre d'avantages. Cet objet permet d'une part de donner du sens à des gestes, en mimant l'utilisation d'un objet réel. Cette métaphore vise aussi à faciliter la prise en main de l'interface pour de nouveaux utilisateurs, en mettant en perspective le dictionnaire de gestes autour de ce concept d'objet virtuel. D'autre part, l'objet virtuel est un support adapté pour donner un retour sonore sur les gestes effectués. Les sons de notre environnement sont la conséquence d'une action sur un objet comme nous l'avons vu dans le chapitre 2, notre objet virtuel peut donc permettre de nous placer dans ce paradigme. Enfin, différentes manipulations de l'OV peuvent être liées à plusieurs fonctions d'une interface, ce qui peut permettre d'en faciliter l'accès, notamment en situation de conduite.

Ce chapitre vise donc à poser les bases concernant la construction de l'objet virtuel, et les stratégies de sonification qui l'accompagnent. Nous débuterons par une brève description du système que nous avons mis en place. Par la suite nous nous focaliserons sur les retours sonores et à leur construction, en partant de paramètres acoustiques courants. Une première étape sera de réfléchir aux liens possibles entre les paramètres de l'objet et les paramètres sonores. Faut-il les lier de manière linéaire ou plutôt logarithmique ? Avec quelle polarité ? Après avoir apporté des réponses à ces questions dans une première expérience perceptive, il sera alors question de comparer les stratégies de sonification créées, pour obtenir la stratégie la plus informative possible sur l'OV. Il faudra donc mettre au point une tâche permettant de juger la perception de la dynamique de l'objet virtuel par les sujets. Cette

hiérarchisation des stratégies sonores sera l'objet d'une deuxième expérience, à la suite de laquelle nous retiendrons deux stratégies, que nous conserverons pour la suite de l'étude.

Sommaire

4.1	Système mis en place	81
4.1.1	Choix du capteur et logiciels	81
4.1.2	Construction de l'objet virtuel	82
4.2	Expérience 1 : liens entre paramètres de l'objet virtuel et paramètres sonores	86
4.2.1	Hypothèses initiales et motivations	86
4.2.2	Méthodologie	86
4.2.3	Résultats	89
4.2.4	Discussion	90
4.2.5	Conclusion	92
4.3	Expérience 2 : hiérarchisation des stratégies sonores	92
4.3.1	Hypothèses initiales et motivations	93
4.3.2	Méthodologie	94
4.3.3	Résultats	98
4.3.4	Discussion	101
4.3.5	Conclusion	104

Ce chapitre est basé sur les travaux présentés dans l'article de journal :

BRESSOLETTE, B., DENJEAN, S., ROUSSARIE, V., ARAMAKI, M., YSTAD, S. ET KRONLAND-MARTINET, R. (2018). Harnessing Audio in Auto Control: The Challenge of Sonifying Virtual Objects for Gesture Control of Cars. *IEEE Consumer Electronics Magazine*, 7(2):91–100.

DOI : 10.1109/MCE.2017.2721678

4.1 Système mis en place

Nous allons présenter succinctement la structure du système que nous avons mis en place, ainsi que des hypothèses supplémentaires que nous avons choisies concernant l'objet virtuel.

4.1.1 Choix du capteur et logiciels

Notre système contient trois briques principales : le système de captation de gestes ; un logiciel de 3D pour modéliser l'objet virtuel ; et un logiciel de traitement et de restitution du son. Nous allons maintenant aborder chacune de ces trois briques.

Capteur gestuel

Le capteur de gestes que nous allons utiliser doit répondre à la fois aux exigences d'une étude fondamentale, mais également aux contraintes industrielles. Les systèmes nécessitant le port d'un gant ou d'un bracelet n'ont donc pas été retenus pour notre étude, puisqu'il est exclu que le conducteur soit obligé de s'équiper à chaque utilisation dans son véhicule.

Parmi les capteurs permettant une reconstruction de la main, nous avons choisi la Leapmotion développée par la société du même nom¹. Ce capteur d'un coût relativement faible est fourni avec une suite logicielle complète – ou SDK pour *Software Development Kit*. Il est ainsi possible d'avoir accès à un grand nombre de paramètres de la main, ce qui est un avantage considérable par rapport aux autres capteurs qui nécessitent des algorithmes de traitement d'images et de reconnaissance de gestes. Dans le cadre de cette thèse, les paramètres gestuels sont une donnée d'entrée. De plus, ce capteur permet une captation relativement fiable, même si certaines limites ont été démontrées [Guna *et al.*, 2014] [Adhikarla *et al.*, 2015].

Logiciel de 3D

Un logiciel de traitement 3D est nécessaire d'une part pour reconstruire la main de l'utilisateur, mais également pour ajouter notre objet virtuel. Nous avons choisi le logiciel Unity mis au point par la société *Unity Technologies*², qui est très utilisé dans le domaine du jeu vidéo. Les sociétés *Leap Motion* et *Unity Technologies* ont d'ailleurs travaillé conjointement pour permettre une intégration facile de la Leapmotion, notamment pour la réalité virtuelle et augmentée, ce qui a facilité la mise en place de notre système.

Unity permet également de mettre au point une interface visuelle, ce qui sera utile dans nos prochaines expériences.

Logiciel de traitement et de restitution sonore

Le logiciel Max développé par la société *Cycling'74*³ a été choisi pour le traitement du son, mais également comme centre de gestion de l'ensemble du système. Ce logiciel permet de générer ou de lire des contenus sonores en temps réel, et de nombreux outils que nous utiliserons sont accessibles avec Max, comme l'outil de spatialisation du son développé par l'IRCAM⁴ qui utilise un algorithme de *panning* d'intensité VBAP [Carpentier, 2015] [Pulkki, 1997].

1. <https://www.leapmotion.com/>

2. <https://unity3d.com/fr/>

3. <https://cycling74.com/>

4. <http://forumnet.ircam.fr/fr/produit/spat/>

La connexion de ces briques est représentée sur la figure 4.1. Unity et Max dialoguent par l'intermédiaire du protocole OSC – pour *Open Sound Control*, pour transmettre les informations concernant l'objet virtuel ou encore l'interface.

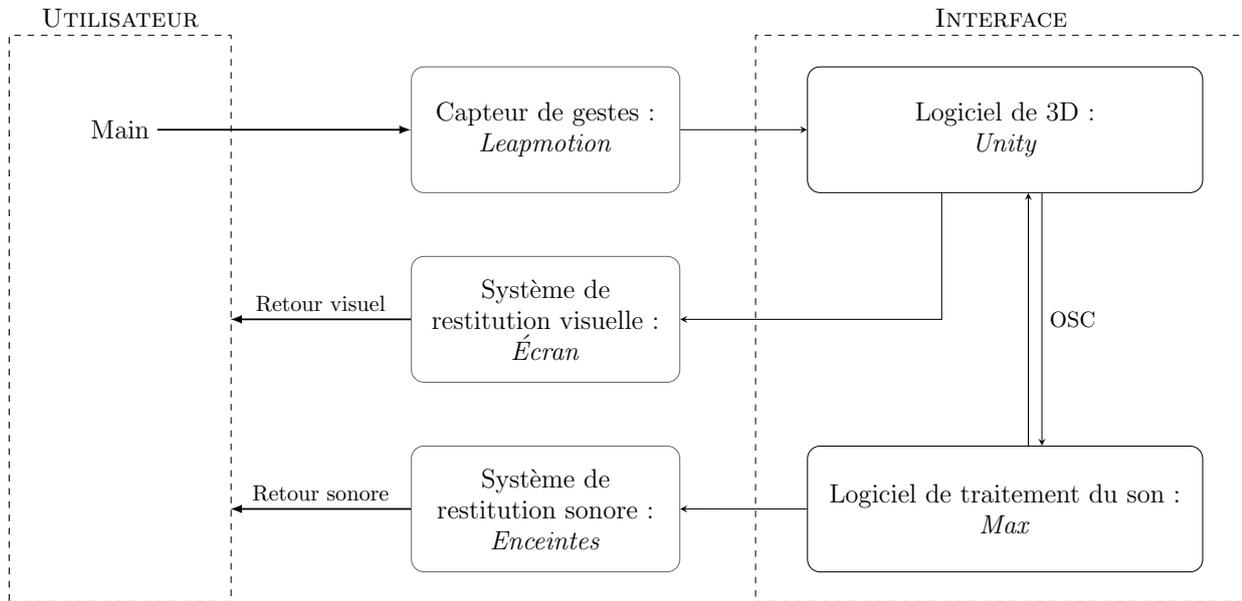


FIGURE 4.1 – Schéma des connexions des différentes briques du système mis en place.

4.1.2 Construction de l'objet virtuel

Objectifs

Nous avons formulé l'hypothèse d'utiliser un objet virtuel tout au long de cette thèse pour plusieurs raisons. Tout d'abord, il peut servir de support aux gestes effectués. L'utilisation d'une métaphore de la manipulation d'un objet réel permet de donner du sens aux gestes effectués, puisqu'ils sont tous liés à l'objet virtuel. Il sera alors possible de le pousser, de le déplacer, voire de l'écraser comme un véritable objet. Cet aspect sera plus précisément décrit au chapitre 6. De plus l'utilisation d'un OV conduit naturellement au paradigme action – objet que nous avons vu au chapitre précédent : une interaction avec un objet crée un son. Ce principe peut donc permettre d'ajouter un son naturellement à l'objet virtuel via une stratégie de sonification adaptée, qui véhicule du sens. Enfin, l'objet virtuel peut conduire à revoir la construction du menu d'une interface : le choix de la fonction avec laquelle l'utilisateur veut interagir peut se faire directement avec le choix de la manipulation. Pour reprendre l'exemple précédent, le placement de l'objet peut être lié à une première fonction, et l'action de l'écraser à une seconde.

L'objet virtuel est donc central dans le système mis en place : il peut permettre de redéfinir un menu, et est le lien entre les gestes de l'utilisateur et les retours sonores qui lui sont restitués.

Construction

Avant de commencer nos expériences, nous avons dû faire plusieurs choix concernant l'objet virtuel. Il a tout d'abord été nécessaire d'en décider la géométrie. La notion d'affordance introduite par Gibson peut être intéressante pour réfléchir à cette question [Gibson, 1979] : « les affordances de l'environnement concernent ce qui est offert aux êtres vivants, ce qui est à leur disposition, ce qui leur est fourni ». Par exemple, si une surface est relativement plate, étendue et rigide, alors cette surface *afforde* l'action de supporter un objet. En revenant à la géométrie de l'objet virtuel, nous avons choisi de le représenter par une sphère, qui afforde l'action de prise en main, ainsi que l'action de rouler – ce qui n'est pas le cas pour une pyramide ou un cube par exemple. Cette géométrie est intéressante puisqu'elle ne semble pas limiter les interactions avec l'objet virtuel qu'il est possible d'imaginer.

De plus, nous avons ajouté une force de rappel grâce à un ressort virtuel, pour diriger l'objet virtuel naturellement vers la main. La force est de la forme $\vec{f} = -k \cdot \vec{x}^\alpha$, où k représente la raideur du ressort, α est une valeur arbitrairement fixée et x est la distance entre la main et l'objet. Ce lien a semblé utile en se projetant dans une utilisation du système en véhicule : lorsque l'utilisateur va vouloir interagir avec l'objet virtuel, il faut qu'il soit facilement accessible. Le ressort permet de guider l'objet vers la main dès que celle-ci est détectée par la Leapmotion. Enfin, nous avons restreint les mouvements de l'objet sur un axe gauche-droite, pour lui éviter de faire des mouvements incontrôlables, et d'en faciliter la maîtrise notamment si l'utilisateur ne peut pas le voir.

D'un point de vue physique, les mouvements de l'objet virtuel selon l'axe \vec{e}_x gauche-droite peuvent être modélisés par l'équation suivante :

$$\ddot{x} + \omega_0^2 \cdot x^\alpha = 0$$

avec $\omega_0 = \sqrt{k/m}$. Cette équation montre que la dynamique de l'objet virtuel est influencée par sa masse m et par la raideur k du ressort. Ce paramètre du ressort est donc un paramètre à transmettre au travers des stratégies de sonification.

Stratégies de sonification

La stratégie de sonification que nous voulons mettre en place doit véhiculer le maximum d'informations sur la dynamique de l'objet virtuel. En effet, l'objectif est de pouvoir utiliser l'objet virtuel en situation de conduite sans le voir, en étant guidé uniquement par le son. Le manque de retours visuels et haptiques de l'OV par rapport à un objet physique doit donc être contrebalancé par une richesse d'informations transmises par les stratégies sonores.

De plus, notre hypothèse concernant l'utilisation d'un objet virtuel tend à opter pour une sonification naturelle, qui pourrait permettre une prise en main du système facilitée, les sons étant facilement reconnus par l'utilisateur.

Pour mettre au point les stratégies de sonification, nous avons déjà dû déterminer quels paramètres de l'objet virtuel pouvaient être pertinents à transmettre aux utilisateurs. Notre choix s'est porté premièrement sur la distance entre l'objet et la main, qui permet de repérer l'objet virtuel dans l'espace. La vitesse de l'objet virtuel est également un paramètre important d'un point de vue perceptif, comme nous l'avons vu dans le chapitre précédent [Rath et Rocchesso, 2005] [Thoret, 2014] [Danna *et al.*, 2013a], Houben ayant même montré qu'il est possible de reconnaître la vitesse de billes roulantes par le son qu'elles produisent [Houben, 2002].

Les paramètres sonores qui transmettent ces informations doivent également être choisis. Nous partons de stratégies bas-niveau dans le même esprit que Parseihian [Parseihian *et al.*, 2016], et nous avons choisi d'utiliser un bruit rose comme base des stratégies sonores, pour être le plus neutre possible en terme d'évocations. Ce choix permet de s'intéresser dans ce chapitre uniquement au lien entre les paramètres de l'OV et les paramètres sonores d'un point de vue général. Nous verrons dans la suite de ce document qu'il peut être intéressant d'utiliser une texture sonore comme base des stratégies de sonification, dans un but esthétique par exemple. L'application au contexte industriel du Groupe PSA sera abordée dans le chapitre 5 de ce document, où une section est dédiée à la mise au point d'une texture sonore, avant de discuter notamment de la question des évocations sonores. Nous avons donc considéré deux catégories de paramètres sonores : les paramètres temporels et les paramètres spectraux. Concernant les paramètres temporels, nous avons choisi la modulation d'amplitude – notée MA – qui peut être facilement appliquée à un bruit rose continu. Ce choix a été motivé par la volonté de donner une information rythmique, dans le même esprit que les radars de recul par exemple, qui lient la distance à l'obstacle au temps entre deux modulations successives. Du côté des paramètres spectraux, notre choix s'est porté sur la hauteur et la brillance. La brillance semble être un paramètre proche du barycentre évoqué par Houben, et il peut être intéressant de lier ce paramètre à la vitesse notamment [Beauchamp, 1982] [Houben, 2002]. La hauteur est quant à elle un paramètre sonore fondamental, auquel l'oreille est particulièrement sensible [Moore, 1973]. Dans le cas de la hauteur, nous avons ajouté une compression exponentielle puisque nous supposons une perception logarithmique de la perception de la hauteur, comme dans l'étude de Parseihian *et al.* [Parseihian *et al.*, 2016] et plus généralement dans le fonctionnement de l'oreille.

L'association de ces paramètres de l'objet virtuel et du son permet de mettre au point six stratégies sonores, détaillées dans le tableau 4.1. Chaque ligne représente une stratégie de sonification, liant le paramètre de l'objet listé en tête de colonne au paramètre sonore indiqué sur chaque ligne. Par exemple, la stratégie 1 correspond au lien distance – hauteur, alors que la stratégie 6 désigne le lien vitesse – brillance.

Stratégie sonore	Distance	Vitesse
1	Hauteur	–
2	MA	–
3	Brillance	–
4	–	Hauteur
5	–	MA
6	–	Brillance

TABLEAU 4.1 – Les six stratégies de sonification mises au point.

Nous allons maintenant nous intéresser rapidement à la construction des sons qui seront utilisés pour la modulation d'amplitude, la hauteur ou la brillance. Les stratégies sonores sont bâties à partir de bruit rose, qui possède un spectre riche tout en restant neutre.

Modulation d'amplitude

Le signal modulé est créé à partir de la multiplication du signal de bruit rose par un signal de basse fréquence. Dans notre cas, nous avons choisi de générer un signal avec la partie positive d'un signal triangulaire, pour obtenir un effet plus net par rapport à un signal sinusoïdal. Nous avons choisi une modulation de 80% du signal – 0% représentant aucune modulation et 100% une modulation totale – pour transmettre une information clairement perceptible. Le paramètre sonore qui va véhiculer de l'information est alors la fréquence de modulation, qu'il est possible de faire varier entre 1 et 10 Hz.

Hauteur

Le bruit rose est ici filtré par un filtre passe-bande mis au point suivant la méthode développée par Mathews et Smith avec un facteur de décroissance de 25 [Mathews et Smith, 2003]. La fréquence centrale de ce filtre varie entre f_{min} et f_{max} . Nous avons ajouté une compression exponentielle avant le filtrage, puisque nous supposons une perception de la hauteur logarithmique, comme l'ont également fait Parseihian *et al.* [Parseihian *et al.*, 2016]. Nous obtenons donc la fonction suivante :

$$f(x) = f_{min} \cdot 2^{x \cdot n_{oct}}$$

où n_{oct} représente le nombre d'octaves couverts par cette stratégie sonore, x une variable à sonifier comprise entre 0 et 1 et $f(x)$ la fréquence centrale du filtre. Pour éviter que les variations de hauteur soient trop importantes et perturbent les utilisateurs, nous avons donc choisi de limiter les variations à 1,58 octave, ce qui correspond à 19 demi-tons. Nous avons choisi $f_{min} = 300 \text{ Hz}$, ce qui donne une valeur maximale de $f_{max} = 900 \text{ Hz}$.

Brillance

Pour ce paramètre sonore, le bruit rose est tout d'abord filtré par un filtre passe-haut avec une fréquence de coupure $f_{c1} = 150 \text{ Hz}$ et un facteur de qualité $Q = 0,79$ avant d'utiliser un filtre passe-bas avec une fréquence de coupure f_{c2} et un facteur de qualité $Q = 0,79$. Le paramètre sonore que nous faisons varier est la fréquence f_{c2} , qui peut varier entre 300 et 900 Hz, en gardant les mêmes limites que le paramètre de hauteur. Il est intéressant de noter le lien qu'il existe entre le paramètre sonore de brillance et le centre de gravité spectral : lorsque la fréquence de coupure f_{c2} augmente, des fréquences aiguës du bruit rose ne sont alors plus coupées, et le centre de gravité augmente.

Les paramètres de l'objet virtuel sont donc liés aux paramètres sonores que nous venons de décrire (i.e. 3 paramètres sonores et 2 paramètres à sonifier) pour former les six stratégies décrites dans le tableau 4.1.

La première question qui se pose concerne la manière de lier les paramètres de l'objet virtuel et du son, pour obtenir les stratégies sonores illustrant le plus naturellement possible les déplacements de l'OV. La manière dont le son va évoluer avec l'objet virtuel va en effet fortement influencer sa perception, et il est important que les modifications temporelles du son soient reliées à la dynamique de l'objet virtuel. La première expérience de cette thèse a été l'occasion d'apporter des réponses à ces questions.

4.2 Expérience 1 : liens entre paramètres de l'objet virtuel et paramètres sonores

Lors de cette expérience, nous avons cherché à déterminer les liens entre les paramètres de l'objet et les paramètres sonores qui soient les plus naturels possibles. Nous avons proposé un test de manipulation de l'objet virtuel grâce à des gestes de balayage de la main. Ces gestes permettent de pousser l'objet virtuel dans une direction, dont la vitesse va augmenter sous l'effet de son inertie. Le ressort reliant la main à l'objet virtuel va alors se tendre, ce qui va progressivement arrêter l'objet pour ensuite le ramener vers la main de l'utilisateur. Cette manipulation semble entraîner des mouvements de l'objet virtuel suffisants pour le déroulement de cette étude : ce geste sépare l'OV de la main, permettant de se pencher sur la véritable sonification de l'OV ; et il modifie la distance main-OV ainsi que la vitesse de l'OV.

Pour cette première expérience, nous avons choisi de nous limiter à des balayages de la main droite, avec une direction allant de droite à gauche. L'objet part donc vers la gauche du sujet avant de revenir dans sa main droite.

4.2.1 Hypothèses initiales et motivations

Pour déterminer le lien le plus naturel entre les paramètres gestuels et sonores, nous allons ici nous appuyer sur la perception multisensorielle basée sur les modalités visuelles, proprioceptives et sonores. Dans cette expérience, il sera possible pour les sujets de voir les mouvements de l'objet virtuel, et d'entendre les sons produits par les six stratégies de sonification. Les gestes et les stimuli visuels de la main et de l'OV sont fortement liés, puisque ces derniers sont construits à partir de la captation par la Leapmotion. La variable d'étude est donc la fonction liant le paramètre gestuel au paramètre sonore. L'hypothèse ici est que les utilisateurs, par leur forte expérience de manipulation d'objets au quotidien, pourront déterminer si l'association des stimuli gestuels, visuels et auditifs semble naturelle. Nous allons donc faire appel aux habitudes de manipulation d'objets et donc aux schémas cognitifs qui y sont liés, stockés dans la mémoire long terme des sujets, comme le montre la figure 4.2.

4.2.2 Méthodologie

Panel de sujets

Vingt sujets ont participé à ce test, six femmes et quatorze hommes, âgés de 21 à 44 ans – âge moyen : 28,7 ans ; écart-type : 7,0 ans. Tous travaillent au sein du Groupe PSA, et aucun d'eux n'a signalé de problèmes auditifs ou moteurs.

Dispositif expérimental

Le but du dispositif est de donner une impression de réalisme aux sujets, pour qu'ils utilisent au maximum les schémas cognitifs associés aux manipulations d'objets réels. Lors de l'expérience, les sujets sont assis à une table sur laquelle est posé le capteur de geste. Un écran d'ordinateur est placé face à eux, sur lequel il est possible de voir la reconstruction de la main faite à partir des données captées, ainsi que l'objet virtuel. Trois enceintes Fostex PM0.4d sont également placées sur la table, régulièrement espacées sur la gauche du sujet comme le montre la figure 4.3.

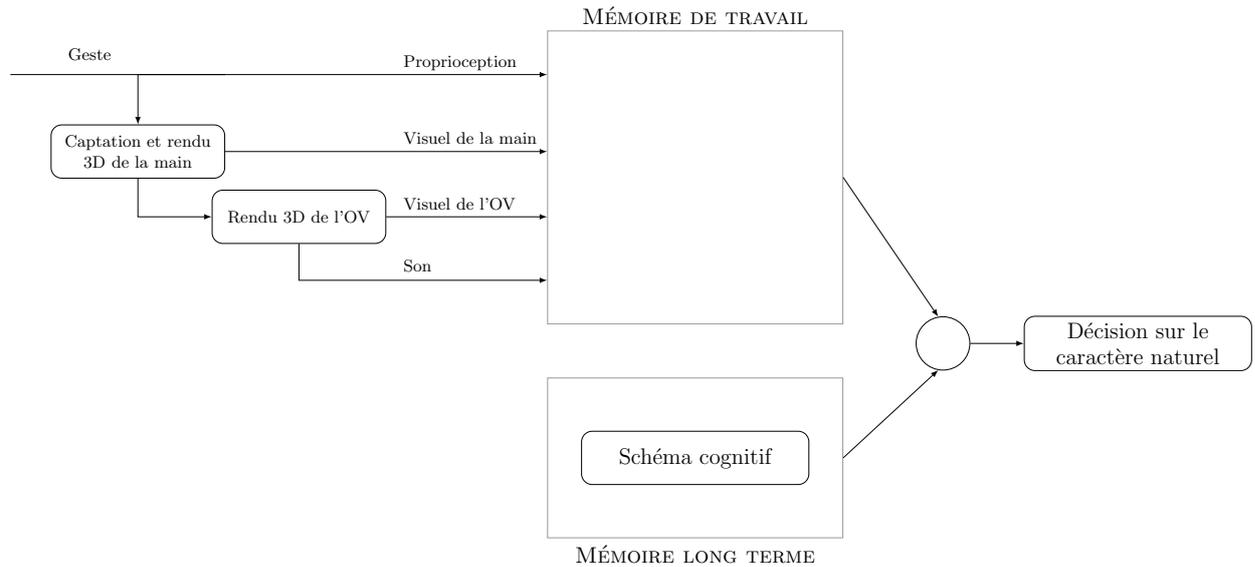


FIGURE 4.2 – Schéma résumant les stimuli proprioceptifs, visuels et auditifs disponibles lors de l'expérience, pouvant être comparés à des schémas cognitifs existants.

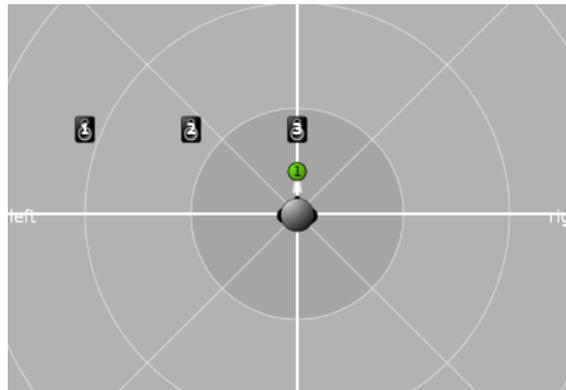
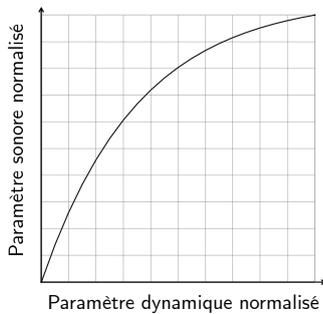


FIGURE 4.3 – Vue du dessus du dispositif donnée par l'interface de l'outil de spatialisation. Il est possible de voir l'auditeur au centre, les positions des 3 enceintes qui sont numérotées, ainsi que l'objet virtuel placé devant l'auditeur.

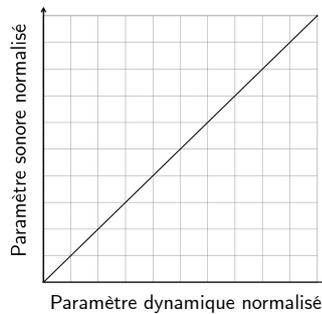
Ce positionnement des enceintes permet un rendu sonore spatialisé, pour augmenter l'impression de réalisme de la scène. Il est alors possible de restituer aux auditeurs les indices interauraux de localisation de source que nous avons abordés en section 2.1.2. Le choix des enceintes par rapport à une restitution par casque est motivé par l'application en véhicule automobile, où le son sera transmis via les enceintes.

Stimuli

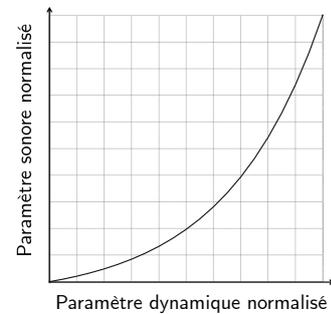
Les 6 stratégies sonores évoquées précédemment ont été utilisées – voir tableau 4.1. Pour chacune de ces stratégies, nous avons proposé différents liens de correspondance – ou *mappings* – entre les paramètres gestuels et sonores. 3 courbes différentes sont proposées – linéaire, exponentielle ou logarithmique, combinées à 2 polarités – croissante ou décroissante, ce qui donne les 6 courbes présentées en figure 4.4. Les sujets ont donc dû évaluer 6 stratégies sonores \times 6 mappings = 36 stimuli sonores différents. Les stimuli visuels correspondent à la reconstruction de la main du sujet, ainsi qu'à l'objet virtuel qui y est rattaché. Par soucis de clarté, et pour ne pas confondre les mappings et les stratégies sonores, nous allons noter log_+ et log_- les mappings logarithmiques croissants et décroissants respectivement, lin_+ et lin_- les mappings linéaires croissants et décroissants, et exp_+ et exp_- les mappings exponentiels croissants et décroissants respectivement.



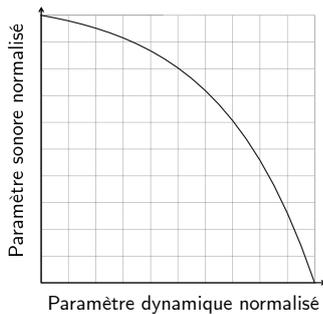
(a) Mapping log_+ : logarithmique croissant



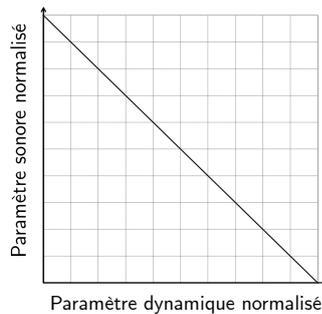
(b) Mapping lin_+ : linéaire croissant



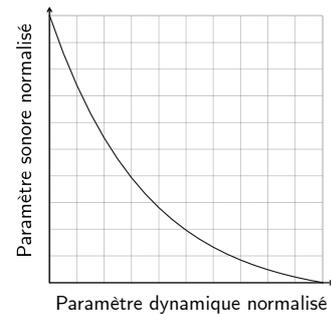
(c) Mapping exp_+ : exponentiel croissant



(d) Mapping log_- : logarithmique décroissant



(e) Mapping lin_- : linéaire décroissant



(f) Mapping exp_- : exponentiel décroissant

FIGURE 4.4 – Les 6 différents mappings sélectionnés, représentés avec des axes normalisés.

Procédure

Pour chaque stratégie sonore, les 6 mappings ont été comparés entre eux grâce à un test de comparaison par paire. Chaque mapping était comparé aux 5 autres, ce qui a conduit à $6 \times (6-1)/2 = 15$ combinaisons différentes, présentées dans un ordre aléatoire. Le sens de présentation des combinaisons est contrebalancé, 10 sujets ayant l'ordre A puis B et les 10 autres l'ordre B puis A. Ces confrontations vont permettre d'obtenir le mapping paraissant le plus naturel pour la stratégie sonore étudiée.

Nous avons appliqué cette méthodologie aux 6 stratégies sonores abordées précédemment, pour déterminer le mapping le plus naturel pour chacun des stratégies. Les 15 comparaisons ont été proposées pour chacune des 6 stratégies, le nombre total de combinaisons testées par chaque sujet est donc de 15×6 stratégies sonores = 90 comparaisons.

Pour un essai donné, les sujets sont invités à faire quelques gestes de balayage d'amplitudes différentes – l'expérimentateur veille à ce que l'ensemble des amplitudes du paramètre dynamique soit bien parcouru pour que toutes les variations du son puissent être perçues. Après avoir observé l'objet virtuel et entendu les sons résultant de cette manipulation, le mapping est changé par l'expérimentateur et le sujet est de nouveau invité à faire des balayages de la main. À l'issue de ces manipulations, il est demandé aux sujets de choisir la manipulation dont un appariement scène visuelle – stimuli sonore est le plus naturel.

Analyse des données

Pour interpréter les résultats, nous avons choisi un test de Bradley–Terry. Ce traitement donne un score normalisé de préférence pour chaque mapping par rapport aux autres.

4.2.3 Résultats

Les scores de Bradley–Terry obtenus à l'issue de l'expérience sont indiqués dans le tableau 4.2. Dans ce tableau, les résultats sont à lire en ligne, chaque mapping ayant été comparé aux autres pour une stratégie particulière. Le test de Bradley–Terry permet de discriminer les scores significativement différents. Nous avons donc indiqué en gras les mappings appartenant au meilleur groupe, et n'étant donc pas significativement différents par rapport à un seuil de significativité de 5%. Le meilleur score est indiqué en italique. Le détail des résultats est disponible en annexe A. Une représentation visuelle en niveaux de gris est également disponible sur la figure A.1, qu'il faut également lire uniquement par ligne.

Les scores obtenus sont assez variables selon la stratégie de sonification. Concernant les stratégies 1, 2 et 3, certaines tendances peuvent être observées, mais aucun des mappings proposés ne semble être nettement plébiscité. Les mappings croissants semblent être privilégiés dans la stratégie 1, même si les différences avec les mappings décroissants ne sont pas significatives. Seul le mapping lin_+ semble se détacher des autres, avec des significativités $p \approx 0,004$ avec le mapping log_- , $p \approx 0,006$ avec le mapping lin_- , et $p \approx 0,004$ avec le mapping exp_- . La stratégie 2 liant la distance à la modulation d'amplitude semble elle avoir un comportement particulier. Il n'y a pas de mapping se détachant nettement, mais au contraire les mappings exp_+ et exp_- sont rejetés – voir figure A.3. Les mappings exponentiels semblent clairement posséder un caractère non naturel.

Stratégie sonore	Mappings					
	log ₊	lin ₊	exp ₊	log ₋	lin ₋	exp ₋
1	0,183	<i>0,259</i>	0,189	0,121	0,126	0,121
2	0,190	0,203	0,106	0,102	0,190	<i>0,210</i>
3	<i>0,244</i>	0,198	0,128	0,132	0,157	0,141
4	<i>0,253</i>	0,228	0,160	0,135	0,155	0,069
5	<i>0,189</i>	0,171	<i>0,189</i>	0,140	0,166	0,145
6	<i>0,488</i>	0,310	0,103	0,040	0,030	0,030

TABLEAU 4.2 – Scores obtenus à l'issue de l'expérience. La meilleure stratégie est indiquée en italique, et le groupe de stratégies contenant la meilleure est indiqué en gras.

Les résultats pour la stratégie 3 semblent montrer une légère préférence des sujets pour le mapping logarithmique ascendant, même si les différences ne sont significatives qu'avec les mappings exp₊, log₋ et exp₋. Les tendances ne sont pas très nettes, comme avec la stratégie sonore 1.

Les résultats avec les stratégies liées à la vitesse de l'objet virtuel donnent en revanche des résultats plus contrastés. La stratégie sonore 4 montre un net rejet du mapping exp₋, et les sujets semblent privilégier les mappings log₊ et lin₊ même si les résultats ne sont pas significatifs.

Pour la stratégie 5 liant la vitesse à la modulation d'amplitude, les résultats sont particuliers puisqu'il n'y a aucune différence significative entre les mappings. Les sujets n'ont pas réussi à départager les mappings sur l'aspect naturel de cette stratégie de sonification.

Enfin, les résultats les plus nets de cette expérience proviennent de la stratégie sonore 6. Les mappings log₊ et lin₊ sont largement plébiscités pour cette stratégie combinant la vitesse à la brillance. Ces deux mappings sont perçus significativement plus naturels par rapport aux autres comme le montre la figure A.7. Il est intéressant de noter que les stratégies décroissantes log₋, lin₋ et exp₋ sont particulièrement rejetées par rapport aux stratégies croissantes. Enfin, le mapping log₊ n'est pas significativement différent du mapping lin₊, avec une significativité $p \approx 0,15$.

4.2.4 Discussion

Les résultats de cette expérience renseignent de différentes manières. La stratégie de sonification liant la vitesse à la modulation d'amplitude ne semble donner aucune préférence pour les six mappings testés. Plusieurs hypothèses peuvent expliquer un tel comportement. Il est par exemple possible que cette stratégie ne soit pas évaluée comme naturelle quels que soient les mappings, les réponses étant alors quasi aléatoires. Il est également possible que le son ne corresponde pas au visuel proposé quelque soit le mapping testé. Enfin, les choix que nous avons fait concernant les paramètres des modulations peuvent avoir menés à des stratégies de sonification non naturelles.

Les résultats pour les stratégies 1, 2, 3, et 4 donnent des résultats mitigés, sans véritable mapping dominant et donc sans mapping jugé particulièrement naturel. Les stratégies 1 et 4 semblent indiquer une préférence des sujets pour la polarité croissante, une augmentation de distance ou de la vitesse est donc plus naturellement interprétée avec une hauteur qui augmente.

En revanche, la stratégie sonore apportant le plus de contrastes dans les réponses des sujets concerne sans conteste la sonification de la vitesse par la brillance. Les mappings \log_+ et lin_+ sont jugés les plus naturels, avec une préférence pour le mapping logarithmique, même si la différence avec le mapping linéaire n'est pas significative – $p \approx 0,15$. Il est cependant clair que le sens critique des sujets s'est tout particulièrement exprimé avec ce type de sonification : le score de Bradley–Terry pour le mapping \log_+ est de 0,488 pour cette stratégie de sonification, alors qu'il est au maximum de 0,259 pour les autres. Ce phénomène peut provenir du caractère naturel du son et de l'habitude des sujets à entendre ce type de son : si un mapping ressenti comme non naturel est utilisé, il est rejeté sans hésitations comme c'est le cas pour les mappings \log_- , lin_- et exp_- et dans une moindre mesure pour le mapping exp_+ . Il peut être intéressant de rapprocher ces résultats des travaux de Houben concernant la perception de vitesse d'une bille [Houben *et al.*, 2004]. Comme nous l'avons vu dans le paragraphe consacré à la brillance de la section 4.1.2, le paramètre de brillance est lié au centre de gravité spectral. Le barycentre d'intensité spécifique calculé par Houben se base sur le centre de gravité spectral, tout en tenant compte de caractéristiques spécifiques de l'oreille – fonctions de transfert du tympan à la cochlée et filtres auditifs de la cochlée. Nous avons donc calculé ce paramètre pour 11 vitesses de l'objet virtuel, allant de 0 à 1 m.s^{-1} par pas de $0,1 \text{ m.s}^{-1}$. Ces résultats sont renseignés sur la figure 4.5, auxquels nous avons ajouté les résultats de Houben [Houben *et al.*, 2004]. Les résultats sont exprimés en ERB pour *Equivalent Rectangular Bandwidth*. Il est possible de relier une grandeur θ exprimée en ERB à une fréquence f en Hertz par la formule $\theta = 21,4 \times \log_{10}(4,37 \times f + 1)$ [Glasberg et Moore, 1990].

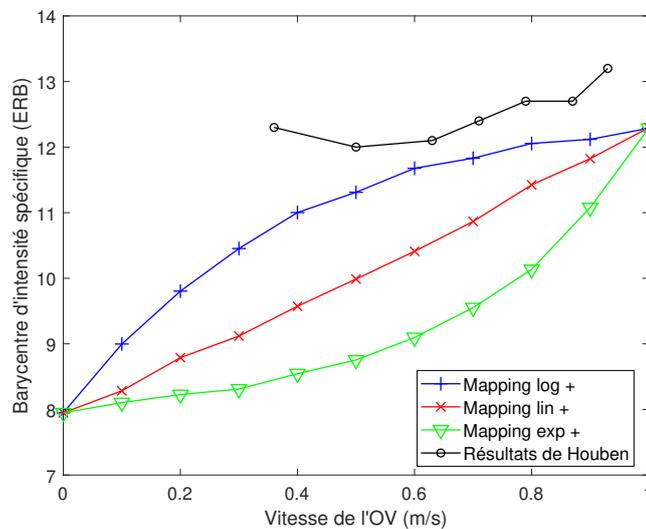


FIGURE 4.5 – Barycentre d'intensité spécifique calculé pour 11 vitesses différentes pour la stratégie sonore 6. Les résultats de Houben sont également renseignés.

La première chose à noter est la similarité entre les courbes des mappings \log_+ , lin_+ et exp_+ sur la figure 4.5 avec les figures 4.4a, 4.4b et 4.4c. Nous voyons donc clairement que le barycentre calculé par Houben est proche de notre paramètre de brillance.

La figure 4.5 ne révèle aucune similarité entre les courbes des mappings avec les résultats de Houben en dessous de $0,6 \text{ m.s}^{-1}$. En revanche, à partir de $0,6 \text{ m.s}^{-1}$ le barycentre de toutes les courbes augmente avec la vitesse. Les courbes des mappings \log_+ et lin_+ sont relativement proches des

résultats de Houben. Pour se rendre compte des évolutions des différentes courbes, nous avons calculé les pentes de chacune des courbes à partir de $0,6 \text{ m.s}^{-1}$, et nous obtenons $(13,2-12,1)/(0,93-0,63) \approx 3,67$ pour les données de Houben et respectivement $\approx 1,53$ et $\approx 4,71$ pour les mappings \log_+ et lin_+ . Ces résultats semblent indiquer que l'évolution d'un véritable son de roulement se situe entre les mappings \log_+ et lin_+ . De plus, il est intéressant de noter la proximité entre les courbes des mappings \log_+ et lin_+ avec les résultats de Houben, ce qui peut expliquer pourquoi ces deux mappings ont été significativement plus appréciés que les autres. Il est donc possible que les sujets ont jugé ces mappings comme naturels à cause de cette évolution similaire au-dessus de $0,6 \text{ m.s}^{-1}$. Leur oreille a donc pu être particulièrement critique, ce qui peut expliquer le rejet catégorique des autres mappings.

4.2.5 Conclusion

Cette expérience avait pour but de lier les paramètres de l'objet virtuel et du son par des mappings naturels, pour créer des stratégies sonores cohérentes avec la représentation visuelle que nous proposons. Nous avons fait appel aux schémas cognitifs de la manipulation de véritables objets comme le rappelle la figure 4.2. Six stratégies sonores ont été construites à partir de paramètres de l'objet virtuel – la distance entre l'OV et la main et la vitesse de l'OV – et de paramètres sonores – la hauteur, la brillance et la modulation d'amplitude. Pour chaque stratégie de sonification, six mappings ont été comparés lors d'un test de comparaison par paires.

Le caractère naturel de la stratégie 5 liant la vitesse de l'objet à une modulation d'amplitude semble être indépendant du mapping proposé. Les stratégies 1, 2, 3 et 4 montrent certaines tendances, mais aucun mapping n'est particulièrement privilégié. En revanche, les sujets ont été particulièrement sensibles à la stratégie associant la vitesse de l'OV à la brillance, qui a conduit aux résultats les plus contrastés. Les mappings logarithmique et linéaire croissants ont été plébiscités, alors que les autres ont été clairement rejetés. En rapprochant nos résultats de ceux de Houben, nous pouvons formuler l'hypothèse que cette stratégie propose un contenu sonore proche de sons de roulement, ce qui explique les écarts importants entre les six mappings proposés. Ayant une bonne connaissance de ces types de sons, les sujets ont alors pu être particulièrement critiques lorsqu'un mapping perçu comme non naturel était présenté.

Avant de continuer l'étude de nos six stratégies sonores, il est nécessaire de choisir un mapping pour chacune d'entre elle, même si certains résultats de cette expérience sont mitigés. Nous avons donc choisi de conserver les mappings menant aux meilleurs résultats, repérés en italique dans le tableau 4.2 : le mapping lin_+ pour la stratégie 1, exp_- pour la stratégie 2, et le mapping \log_+ pour les stratégies 3, 4, 5, et 6.

4.3 Expérience 2 : hiérarchisation des stratégies sonores

L'expérience précédente nous a donc permis de construire six stratégies de sonification en choisissant un mapping pour chaque association de paramètres de l'OV et de paramètres sonores. Dans la présente expérience, il est maintenant question de faire un choix entre ces stratégies sonores, pour retenir celle qui est la plus informative sur la perception de la dynamique de l'objet virtuel. En effet, l'application automobile que nous visons présuppose l'absence de retours visuels, pour que les yeux du conducteur soient focalisés sur la route. La stratégie sonore doit donc pallier l'absence d'informations visuelles sur les mouvements de l'OV par un contenu riche d'informations, pour permettre à l'utilisateur de l'interface de ressentir ses déplacements uniquement par les sons.

Comme nous le préciserons dans un paragraphe suivant, nous avons décidé de créer des combinaisons des six stratégies de sonification que nous avons mises au point jusqu'ici, pour tester le plus grand nombre de possibilités de sonification. Il est en effet possible d'imaginer qu'une combinaison des stratégies de sonification peut être plus informative que les stratégies testées indépendamment.

4.3.1 Hypothèses initiales et motivations

Nous désirons à présent déterminer la stratégie de sonification véhiculant le plus d'informations sur le comportement dynamique de l'objet virtuel. Seule la perception de la dynamique de l'OV par le son nous intéresse, cette expérience sera donc réalisée sans retour visuel de l'objet. L'hypothèse que nous faisons ici est que si les informations dynamiques importantes de l'OV sont données par le son, alors les sujets sont capables d'adapter leurs gestes pour faire les manipulations attendues.

Tâche à réaliser

Le geste de balayage de la main a été conservé pour cette expérience. Il est essentiel de sélectionner une tâche adéquate pour juger de la quantité d'information véhiculée par le son, une tâche de précision pourrait par exemple convenir à notre expérience. Cependant, le but n'est pas d'obtenir un geste parfait grâce au son, mais seulement que les utilisateurs puissent se rendre compte des mouvements de l'objet virtuel.

Nous avons donc choisi de discriminer les sons sur une tâche consistant à envoyer l'OV dans une zone de l'espace parmi trois zones délimitées. Cette tâche, sans nécessiter une trop grande précision, permet d'observer avec quelle facilité les sujets comprennent le comportement de l'OV et du ressort qui le relie à la main de l'utilisateur. Pour accomplir la tâche, les sujets doivent pousser l'objet virtuel avec leur main droite. La distance entre la main et l'OV augmente donc jusqu'à une distance maximale d_{max} à laquelle la vitesse de l'OV est nulle. L'objet revient ensuite dans la main du sujet sous l'action du ressort. Les sujets doivent donc faire les gestes de balayages de la main pour attendre la distance d_{max} correcte. Si d_{max} se trouve entre les limites de la zone demandée, alors le lancer est jugé correct.

Construction des zones

Pour déterminer la position des zones, nous avons choisi de suivre la formulation de Shannon [MacKenzie, 1992] de l'index de difficulté associé à loi de Fitts [Fitts, 1954] pour une cible donnée :

$$Id = \log_2 \left(1 + \frac{D}{W} \right)$$

avec Id l'indice de difficulté de la tâche, D la distance de la cible et W sa largeur. Pour avoir un indice de difficulté constant sur les trois zones, il faut donc :

$$\frac{D_{zone1}}{W_{zone1}} = \frac{D_{zone2}}{W_{zone2}} = \frac{D_{zone3}}{W_{zone3}} \quad (4.1)$$

Nous avons choisi de définir trois zones ne se chevauchant pas, nous avons donc obtenu des zones comme l'illustre la figure 4.6.

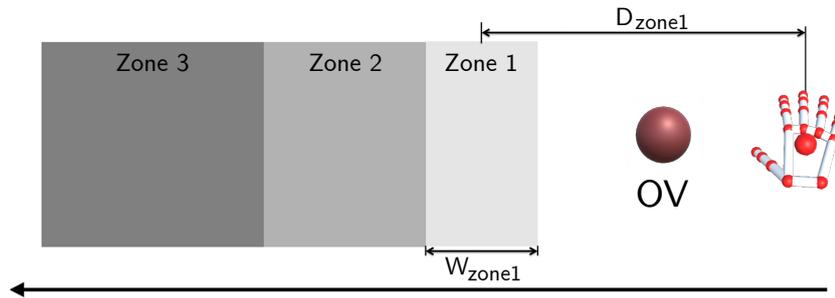


FIGURE 4.6 – Illustration des trois zones à atteindre. La plus petite à droite est la plus proche du sujet, celle de gauche est la plus loin.

L'équation (4.1) implique donc que pour garder une difficulté constante d'après la loi de Fitts, la largeur d'une zone est d'autant plus grande qu'elle est éloignée de la main. La zone de droite est donc fine car elle est proche de l'utilisateur alors que la zone de gauche, plus éloignée, est en conséquence plus large. À ce stade, nous avons choisi de nous appuyer sur la loi de Fitts car la tâche que nous avons choisie s'approche d'une tâche de pointage, même si la précision n'est que le moyen de mesurer la quantité d'information perçue par les sujets dans les stratégies de sonification.

4.3.2 Méthodologie

Panel de sujets

Trente deux sujets droitiers, quatorze femmes et dix-huit hommes, travaillant pour le Groupe PSA se sont portés volontaires pour participer à cette expérience. Ils étaient âgés de 22 à 53 ans – âge moyen : 32,0 ans ; écart-type : 8,1 ans – et n'ont pas déclaré de problèmes auditifs ou moteurs. Certains d'entre eux avaient participé à l'expérience 1.

Dispositif expérimental

Les sujets sont placés dans les mêmes conditions que lors de l'expérience 1 : assis à une table où est posée une Leapmotion. Sur la table sont disposées 3 enceintes Fostex PM0.4d dans la même configuration que la précédente expérience. Face aux sujets est placé un écran où il est possible de voir une interface mise au point pour l'expérience.

Stimuli

Nous pensons qu'il est possible qu'une stratégie contenant une combinaison d'informations sur la distance main-OV et sur la vitesse de l'Ov permette de mieux réaliser la tâche que nous avons mise au point que les deux stratégies prises indépendamment. Pour tenter d'observer ce phénomène, nous avons choisi de créer six autres stratégies sonores à la suite de l'expérience 1, dites *complexes*, en combinant les six stratégies déjà connues, dites *simples*. Une stratégie de contrôle sans son a également été prévue, pour mesurer l'apport du son. L'ensemble des stratégies est donné dans le tableau 4.3, la présentation étant conservée par rapport au tableau 4.1.

Stratégie sonore	Distance	Vitesse
1	Hauteur	–
2	MA	–
3	Brillance	–
4	–	Hauteur
5	–	MA
6	–	Brillance
7	Hauteur	MA
8	Hauteur	Brillance
9	MA	Hauteur
10	MA	Brillance
11	Brillance	Hauteur
12	Brillance	MA
13	–	–

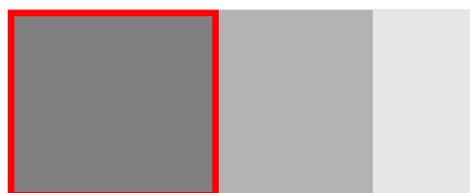
TABLEAU 4.3 – Les treize stratégies de sonification mises au point.

Les sujets ont à leur disposition une interface, observable en figure 4.7, où il leur est possible de voir schématiquement les zones, la zone à atteindre, ainsi qu'une icône bougeant simultanément avec leur main, ce qui peut être important pour confirmer la captation de la main. Cette interface donne certaines informations sur l'accomplissement de la tâche, mais l'objet virtuel n'est pas visible pour l'ensemble de l'expérience.

Phase 1



Zone 3 Zone 2 Zone 1



Pas assez loin !

Tâche à réaliser : Aller en zone 3

FIGURE 4.7 – Interface visuelle mise à la disposition des sujets lors de la phase 1 de l'expérience 2. Les 3 zones sont représentées schématiquement au centre. La zone à viser est entourée et rappelée par le texte en bas de l'interface. Une indication est donnée lorsque le lancé n'est pas réussi, comme il est possible de le voir sur la figure. La main représentée sur la droite suit les mouvements de la main des sujets.

Raideur du ressort

Chaque sujet doit tester l'ensemble des stratégies sonores au cours de cette expérience. Un potentiel biais de l'expérience peut alors provenir de l'apprentissage des trois gestes pour atteindre les trois zones. En effet, les sujets peuvent parvenir à les apprendre pour atteindre les zones demandées en ignorant les sons proposés et en se concentrant uniquement sur la tâche gestuelle. Pour éviter ce cas de figure, nous avons fait varier le paramètre de raideur k entre trois valeurs, représentant un ressort souple, normal et tendu. Lors du changement de stratégie sonore, le paramètre de raideur sera également changé aléatoirement, ce qui obligera les sujets à utiliser les informations sonores pour adapter leurs gestes et ainsi accomplir la tâche correctement. Les sujets n'étaient pas prévenus de ces changements de raideur du ressort.

Chacune des treize stratégies de sonification est testée pour les trois raideurs de ressort, ce qui conduit à $13 \text{ stratégies} \times 3 \text{ raideurs} = 39$ conditions pour chaque sujet.

Procédure

Chacune des 39 conditions est divisée en deux phases ayant lieu l'une après l'autre : une phase avec plusieurs essais successifs dans la même zone – la phase 1 ; puis une phase où les essais sont aléatoirement distribués entre les trois zones – la phase 2. Pour chaque condition, les sujets commencent donc par la phase 1 avant de passer à la phase 2.

Pendant la première phase, les sujets peuvent se familiariser avec la position des zones, ainsi qu'à l'appariement entre les gestes et les sons. Il leur est demandé de faire quinze balayages successifs dans chacune des trois zones. L'ordre de présentation des zones est aléatoire. Pendant cette phase de lancers groupés, les sujets ont un retour visuel sur l'accomplissement de leur tâche : quand la zone cible est manquée, un message « pas assez loin » ou « trop loin » est affiché, pour aider les sujets à trouver le geste correct pour atteindre la zone cible.

La phase 2 est plus proche de notre potentielle application en véhicule et sera donc particulièrement étudiée. Cette phase consiste à exécuter quinze balayages, correspondant à cinq balayages de la main par zone, distribués aléatoirement. L'exécution de la tâche est plus compliquée que lors de la phase 1, puisque le geste demandé varie à chaque essai. De plus, le retour visuel est moins informatif que dans la phase précédente, puisqu'il est seulement indiqué sur l'interface si la zone cible est atteinte ou non. Pour réussir la tâche demandée pendant cette phase de lancers distribués, les sujets doivent tout particulièrement se concentrer sur l'appariement geste – son.

Deux pauses sont proposées aux sujets, après avoir accompli 13 et 26 conditions, pour éviter d'avoir un effet de fatigue trop important.

Analyse des données

Pour cette expérience, les performances sont mesurables par un taux de succès de l'envoi de l'OV dans la zone demandée. Nous avons donc calculé les coefficients de Cochran pour mesurer les effets des stratégies sonores sur les performances pour les phases 1 et 2.

Cependant il peut également être intéressant de prendre en compte la distance à la zone pour un lancer raté. En effet, un lancer juste à côté de la limite de la zone ou très éloigné de la zone est comptabilisé de la même façon par le pourcentage de réussite. Par conséquent, nous avons créé un indice semblable à un taux d'échec, défini par l'équation suivante :

$$I(x) = \begin{cases} 0 & \text{si } x \leq W_{zone}/2 \\ x + 1 - W_{zone}/2 & \text{si } x > W_{zone}/2 \end{cases}$$

où I représente l'indice, x la distance par rapport au centre de la zone à atteindre, et W_{zone} la largeur de la zone.

Ainsi, si le lancer de l'OV est correct, l'indice vaut 0. Pour un mauvais lancer, la distance au centre de la zone est prise en compte comme le montre la figure 4.8, qui donne l'évolution des indices des trois zones selon la position du lancer. Pour comparer les indices de chaque stratégie sonore, nous avons fait une analyse de variance sur l'indice avec les facteurs de la stratégie sonore – 13 – 1 = 12 degrés de libertés ou *ddl*, la raideur du ressort – 2 *ddl*, la zone à atteindre – 2 *ddl*, le numéro du lancé – 14 *ddl*, le facteur sujet – 31 *ddl*, ainsi que leurs interactions simples. Une analyse différente a été effectuée pour chacune des deux phases.

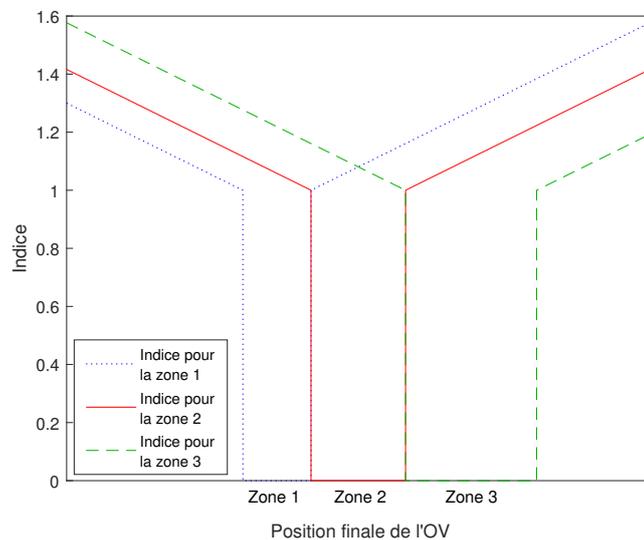


FIGURE 4.8 – Construction de l'indice à partir des positions finales de l'objet virtuel.

4.3.3 Résultats

Performances

Les taux de réussite obtenus lors des deux phases sont donnés figure 4.9. Les taux sont relativement proches, mais les tests de Cochran ont révélé un effet significatif de la stratégie sonore pour la phase 1 – $Q = 21,558$; $N = 4320$; $p = 0,043 < 0,05$ – et pour la phase 2 – $Q = 33,227$; $N = 1440$; $p = 0,001$. Des tests de Mc Nemar ont été effectués en post-hoc avec la correction de Bonferroni [Bland et Altman, 1995], mais ils n'ont pas permis de montrer d'effets significatifs. Les performances sont comprises entre 40 et 50%, ce qui montre la difficulté de la tâche, et elles semblent néanmoins donner un plus grand contraste entre les stratégies sonores dans la phase 2 que dans la phase 1.

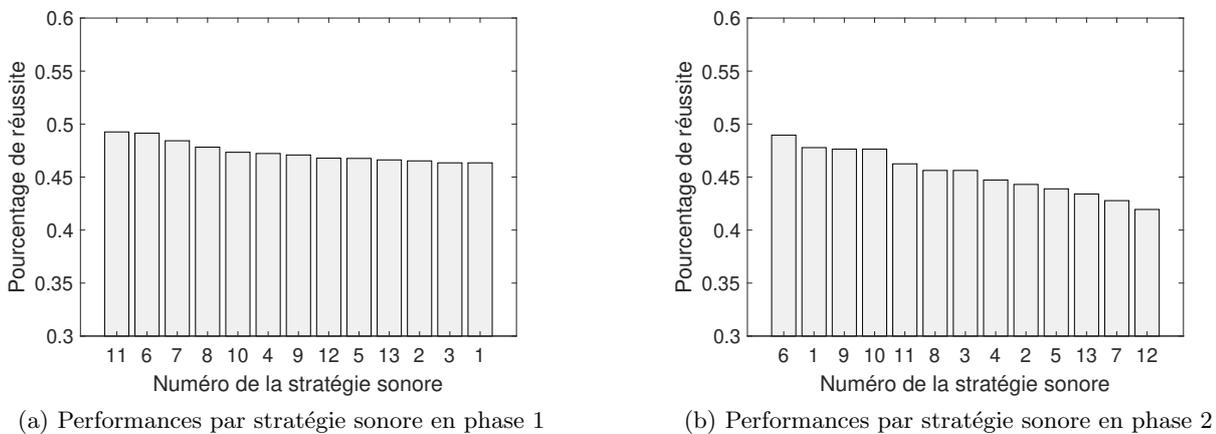


FIGURE 4.9 – Performances par stratégie sonore pour les deux phases de l'expérience 2.

Nous allons maintenant nous intéresser aux indices, qui vont nous renseigner plus précisément sur les balayages effectués par les sujets, en prenant en compte les lancers considérés comme des échecs. Comme expliqué précédemment, l'indice est analogue à un taux d'échec : plus l'indice est proche de zéro et meilleure est la stratégie.

Indices

Les indices pour les deux phases sont donnés par la figure 4.10.

L'ANOVA révèle un effet significatif selon le facteur de la stratégie sonore pour la phase 1 – $F(12,54892) = 1,979$; $p = 0,022 < 0,05$, et pour la phase 2 avec un effet plus prononcé – $F(12,17452) = 2,780$; $p = 0,001$. Les tests post-hoc de Duncan ont cette fois permis de faire ressortir des différences significatives entre les stratégies sonores.

Comme l'application que nous visons sera plus proche de la phase 2 que de la phase 1, nous nous intéresserons particulièrement aux résultats de la phase 2. La phase 1, par les informations qui sont données par l'interface, est moins centrée sur le son et permet principalement aux sujets de saisir la position approximative des zones.

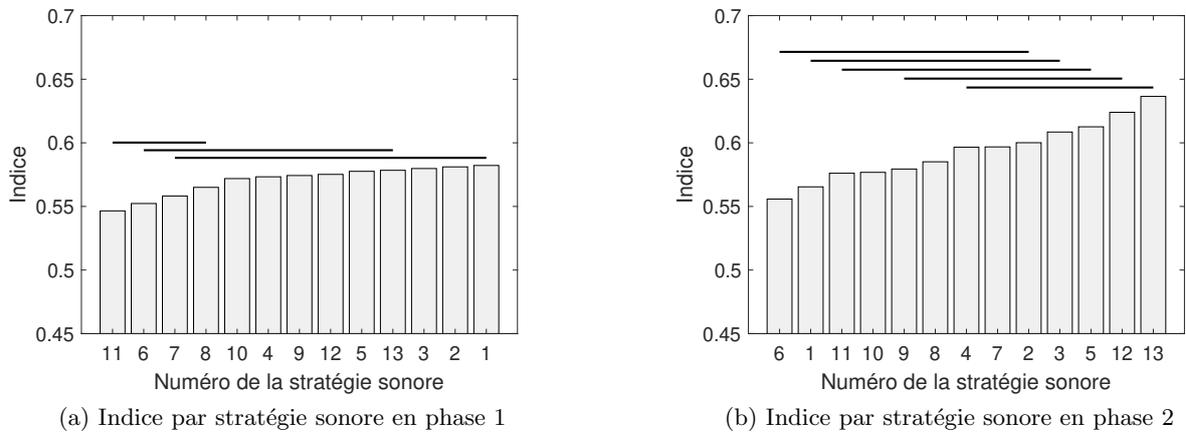


FIGURE 4.10 – Indices par stratégie sonore pour les deux phases de l'expérience 1. Les barres horizontales donnent les groupes de stratégies non significativement différentes ($p < 0,05$).

Pour faciliter la lecture, nous avons fait un rappel des numérotations des stratégies sonores, voir tableau 4.4.

Nous pouvons déjà noter que les stratégies 6, 1, 11, 10, 9 et 8 sont significativement différentes de la stratégie 13 sans son, voir figure 4.10b. Il y a donc bien un apport significatif du son.

En se focalisant dans un premier temps sur les stratégies simples, la stratégie 6 liant la vitesse de l'OV au paramètre de brillance est significativement différente des stratégies 3 et 5. Il est intéressant de noter que ni la sonification du paramètre de distance ou de vitesse ne mène à de meilleurs résultats. Les stratégies 6 et 1 sont les deux meilleures de la phase 2, alors que les stratégies 3 et 5 font parti des plus mauvaises. En ciblant le paramètre sonore de la brillance, nous pouvons voir que l'association n°6 avec le paramètre de vitesse mène à des résultats significativement meilleurs comparé à l'association n°3 avec le paramètre de distance.

En ce qui concerne les sons plus complexes, les stratégies 11, 10, 9 et 8 obtiennent des résultats corrects. Il est intéressant de noter que la stratégie 8 est composée par les stratégies simples 1 et 6, mais n'obtient pourtant pas de meilleurs résultats. Pour les stratégies 9 et 11, il est intéressant de constater que les résultats de ces stratégies sont meilleurs que les stratégies simples qui les composent – 2 et 4 pour la stratégie 9, 3 et 4 pour la stratégie 11 – même si ces différences ne sont pas significatives. Les résultats des stratégies 10 et 8 peuvent provenir de l'association vitesse-brillance de la stratégie 6. Les stratégies 11 et 8 obtiennent de bons résultats malgré le fait qu'elles soient composées de deux paramètres fréquentiels, ce qui aurait pu être déstabilisant et difficile à comprendre pour les sujets. On peut supposer par exemple que les sujets se sont focalisés sur un seul paramètre fréquentiel, et ont ignoré l'autre.

Stratégie sonore	Distance	Vitesse
1	Hauteur	–
2	MA	–
3	Brillance	–
4	–	Hauteur
5	–	MA
6	–	Brillance
7	Hauteur	MA
8	Hauteur	Brillance
9	MA	Hauteur
10	MA	Brillance
11	Brillance	Hauteur
12	Brillance	MA
13	–	–

TABLEAU 4.4 – Rappel des treize stratégies de sonification.

Il est également intéressant de noter que les stratégies 4, 7, 2, 3, 5 et 12 obtiennent des résultats analogues à la situation de contrôle. Il est donc possible que, même si le son a tendance à aider les sujets à accomplir la tâche, certains types de sonification ne donnant pas d'informations suffisamment pertinentes ne soient pas meilleurs qu'une situation sans son. Il est à noter que ces stratégies sonores contiennent des indices de localisation sur la position de l'objet virtuel grâce à la spatialisation, ce qui peut aider les sujets à atteindre la zone demandée. Deux hypothèses peuvent donc être formulées : soit les informations provenant de la spatialisation ne sont pas utiles, soit les stratégies sonores n'aident pas les sujets à accomplir la tâche demandée.

Enfin, chaque association de paramètres semble avoir mené à un résultat particulier. Par exemple, en ce qui concerne les stratégies simples, ni le paramètre de distance – pour les stratégies 1, 2, 3 – ni le paramètre de vitesse – stratégies 4, 5, 6 – n'ont donné des résultats systématiquement supérieurs aux autres. Il en est de même avec les paramètres sonores de hauteur, brillance ou modulation d'amplitude. Les résultats ne dépendent pas seulement des paramètres utilisés, mais bien de l'association des paramètres.

Raideur et zone à atteindre en phase 2

Sur la seconde phase, l'ANOVA montre un effet de la zone à atteindre – $F(2,17452) = 51,343$; $p < 0,001$, mais également de la raideur k du ressort – $F(2,17452) = 6,312$; $p = 0,002$. La raideur 2 du ressort moyen – $k_2 = 1,2 N.m^{-\alpha}$ – donne un indice significativement plus faible que les autres – $p < 0,05$, les raideurs 1 et 3 respectivement des ressorts détendus et tendus – $k_1 = 0,7 N.m^{-\alpha}$ et $k_3 = 1,7 N.m^{-\alpha}$ – ont donc été perturbateurs par rapport à la tâche demandée, d'après le test post-hoc. De plus, la zone 1 la plus proche des sujets conduit significativement à de moins bons résultats par rapport aux deux autres zones, d'après le test de Duncan. L'interaction entre les paramètres de raideur et de zone est également à noter : $F(4,17452) = 10,383$; $p < 0,001$. La figure 4.11 donne le détail de l'interaction entre les deux paramètres.

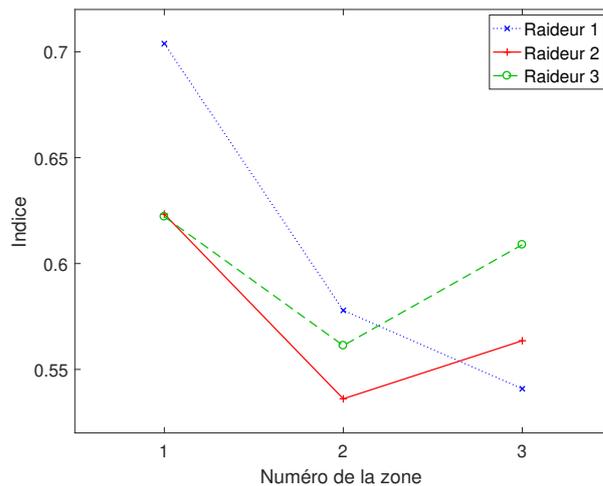


FIGURE 4.11 – Indice détaillé par raideur et zone en phase 2.

Cette figure montre que pour le ressort détendu avec la raideur 1, il est plus difficile d'atteindre la zone la plus proche. En effet pour un geste donné, la force de retour exercée sur l'objet est plus faible pour un paramètre k faible. L'OV pourra parcourir une distance plus élevée avant d'être arrêté, et atteindra une zone plus éloignée. Les lancers en zone 3 donnent de bons résultats pour cette raideur : la position éloignée de cette zone combinée à la faible valeur de raideur la rend relativement facile à atteindre. En ce qui concerne les lancers pour les autres raideurs, le ressort avec la raideur 2 semble conduire à de meilleurs lancers en zones 2 et 3.

Ces résultats semblent montrer que la loi de Fitts n'est pas globalement valable dans notre cas de figure : la zone 1 a été significativement plus difficile à atteindre que les deux autres zones. Nous pouvons cependant noter que la raideur n'intervient pas dans la définition de l'indice de difficulté – $Id = \log_2(1 + D/W)$ – alors qu'elle a une grande incidence sur les résultats des lancers. En effectuant une ANOVA sur chaque raideur individuellement avec les mêmes facteurs que décrits précédemment, un effet de la zone est également constaté : $F_{raideur1}(2,5092) = 51,438$; $p_{raideur1} < 0,001$; $F_{raideur2}(2,5092) = 15,246$; $p_{raideur2} < 0,001$ et $F_{raideur3}(2,5092) = 6,7228$; $p_{raideur3} = 0,001$. Pour la raideur 1, les lancers en zone 3 conduisent à de significativement meilleurs résultats ; pour la raideur 2, les zones 2 et 3 obtiennent les meilleurs résultats mais ne sont pas significativement différents ; et la zone 2 obtient un indice significativement plus bas pour la raideur 3. La loi de Fitts n'est donc pas respectée non plus pour chaque raideur prise individuellement. Cependant, il semble exister un lien entre raideur, distance et largeur de la cible qu'il reste à déterminer. Nous pensons donc que la loi de pointage de Fitts peut être appliquée dans notre cas de figure, en incluant la force de rappel dans le processus de calcul de l'indice de difficulté.

Prise en main lors de la phase 1

Un aspect intéressant concerne la prise en main de l'objet virtuel. Une stratégie sonore a-t-elle permis de se rendre compte plus rapidement de la dynamique de l'objet virtuel dans la phase 1 ? Cette stratégie pourrait être particulièrement intéressante dans la prise en main de l'objet virtuel, lors des premières utilisations du système.

Il est tout d'abord nécessaire d'observer si un effet global d'apprentissage a lieu sur les quinze lancers successifs. Nous avons donc fait une ANOVA en nous intéressant uniquement au numéro de lancers, sans détailler par sujet, raideur, zone ou stratégie sonore dans un premier temps. Les résultats montrent un effet significatif sur les numéros de lancers – $F(14,54892) = 39,504$; $p < 0,001$ pour la distance, $F(14,54892) = 19,293$; $p < 0,001$ pour l'indice. Le test post-hoc de Duncan révèle un effet d'apprentissage entre le premier et le sixième lancer pour la distance, et entre le premier et le quatrième lancer pour l'indice.

4.3.4 Discussion

Tout d'abord, la combinaison d'informations que les stratégies complexes véhiculent n'ont pas eu d'effet particulièrement bénéfique par rapport aux informations des stratégies de sonification simples. Au mieux, elles ne sont pas significativement différentes des meilleures stratégies simples, comme c'est le cas des stratégies 11, 10, 9, 8 et 7 en phase 2. La seule stratégie qui obtient des résultats significativement moins bons que la stratégie 6 est la stratégie 12, liant le paramètre de distance à la brillance et le paramètre de vitesse à la modulation d'amplitude. En particulier, ses résultats sont significativement différents de la stratégie 11, qui diffère seulement sur la sonification de la vitesse de l'OV. Nous pouvons alors penser que la sonification de la vitesse par la modulation d'amplitude entraîne les mauvaises performances de la stratégie 12. Les stratégies 5 et 7 ayant cette

même sonification de la vitesse obtiennent également des résultats mitigés, les résultats de la stratégie 5 étant significativement différents de ceux de la stratégie 6. Ce n'est en revanche pas le cas de la stratégie 7, nous ne pouvons donc pas conclure sur l'aspect non informatif de la sonification de la vitesse de l'OV par la modulation d'amplitude.

Cette discussion sur les différences entre stratégies simples et complexes, à cause de résultats dont les différences ne sont pas souvent significatives, est difficile à mener comme le montre ce début de discussion. De manière générale, il semble que les résultats des stratégies complexes ne sont pas systématiquement meilleurs que les stratégies simples les composant. Il peut donc exister des effets d'interaction entre les sonification de la distance main-OV et de la vitesse de l'OV. Une expérience plus poussée, avec plus de sujets, pourrait permettre de comprendre les stratégies mises en place par les sujets avec ces stratégies complexes : les sujets utilisent-ils les informations portées par les deux sonifications simultanément ? Se focalisent-ils sur une seule sonification, en essayant d'ignorer l'autre, la seconde sonification devenant alors gênante ?

Nous allons à présent nous focaliser sur les stratégies 6 et 1, qui paraissent les plus prometteuses d'après leurs résultats en phase 2 de l'expérience.

Les résultats semblent montrer qu'une stratégie liant la vitesse de l'OV à un paramètre de brillance est intéressante pour transmettre des informations sur la dynamique de l'objet virtuel. Le paramètre de vitesse a été utilisé à plusieurs reprises en sonification, pour illustrer une balle virtuelle [Rath et Schleicher, 2008] ou encore pour recréer un son de frottement [Thoret, 2014]. Les indices obtenus en phase 2 révèlent que ce paramètre ne donne pas systématiquement de bons résultats, ce que révèle la stratégie 5 par exemple, liant les paramètres de vitesse de l'OV et de modulation d'amplitude. La stratégie 6 associant ce même paramètre à la brillance obtient les meilleurs résultats en phase 2, et donne de bons résultats en phase 1. Ces résultats vont dans le sens des travaux d'Houben, qui ont montré un lien entre la perception de la vitesse de balle roulante et un paramètre spectral [Houben *et al.*, 2001]. Plus précisément, le barycentre spectral semble être un paramètre influant de la perception de la taille et de la vitesse de balles roulantes [Houben *et al.*, 2004].

Comme nous l'avons discuté précédemment, le paramètre de brillance que nous utilisons semble lié au barycentre spectral évoqué par Houben *et al.*. Nous avons de plus remarqué lors de l'expérience 1 que les sujets avaient été particulièrement sensibles aux changements de mappings pour la stratégie 6 reliant la vitesse de l'OV au paramètre de la brillance. Si un mapping non naturel est proposé, il est alors fortement rejeté comme le suggère le tableau 4.2. Il semble donc que les sujets ont pu s'appuyer sur des indices sonores connus avec les sons de roulement, pour ressentir la dynamique de l'OV au travers de sa vitesse. Cette supposition est également appuyée par les bonnes performances de la stratégie 6 dans les deux phases de l'expérience 2. Nous pouvons donc considérer la stratégie 6 comme une stratégie de sonification *naturelle*, de par sa proximité avec des sons de roulement naturels.

D'autre part, la sonification de la distance entre la main et l'OV par la hauteur donne de bons résultats en phase 2, mais de mauvais résultats en phase 1. Nous avons également vu en expérience 1 que les sujets n'ont pas réussi à mettre en avant un mapping paraissant particulièrement naturel pour cette stratégie sonore, voir le tableau 4.2. Puisque le critère de jugement des lancers est la distance, les stratégies de sonification de la distance peuvent être avantagées. La hauteur étant un paramètre sonore perçu avec précision [Moore, 1973], la combinaison de la distance à ce paramètre sonore peut créer une stratégie de sonification particulièrement efficace pour l'envoi de l'objet virtuel dans les trois zones. Par opposition à la stratégie 6, cette stratégie peut être qualifiée de *non naturelle* ou d'*abstraite* même si elle semble adaptée à la tâche demandée.

Pour discuter de la vitesse de prise en main des stratégies sonores, nous avons voulu détailler les résultats des lancers en phase 1 selon les stratégies sonores. Les analyses indiquent un effet non significatif de l'interaction (stratégie sonore \times numéro de lancer) – $F(168,54892) = 1,054$; $p = 0,301$ pour la distance, $F(14,54892) = 1,114$; $p = 0,149$ pour l'indice. Malgré ce constat, nous avons voulu représenter les lancers de plusieurs stratégies pour observer des tendances. Nous ne pouvons en revanche tirer aucun résultat de cette analyse, les résultats n'étant pas significatifs. Nous avons représenté les moyennes glissantes de quatre stratégies sonores qui nous semblaient illustratives des tendances en figure 4.12.

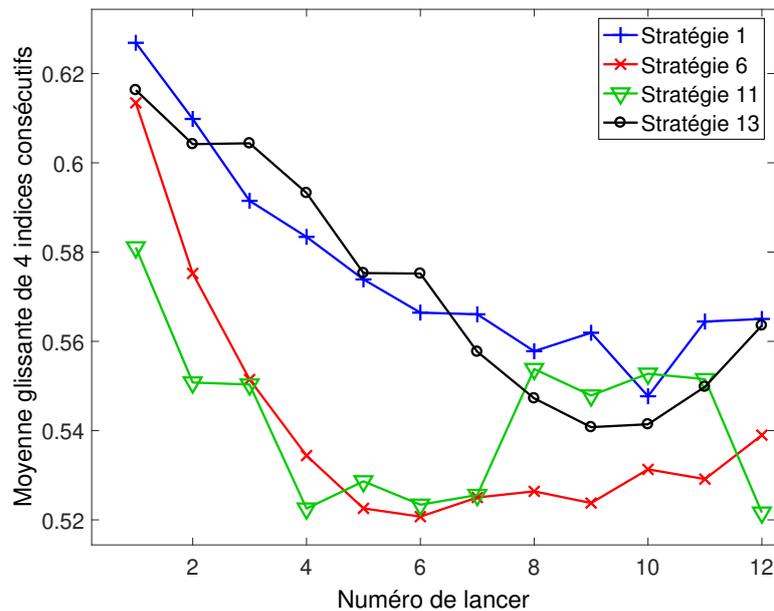


FIGURE 4.12 – Évolution de moyennes glissantes sur quatre indices de quatre stratégies sonores. Les différences entre les courbes ne sont pas significatives.

Les courbes représentées semblent indiquer deux comportements différents : tout d'abord une phase d'amélioration où l'indice diminue, puis une phase de stabilisation avec des oscillations. Malgré des différences non significatives, il semblerait exister deux tendances différentes : les stratégies 1 et 13 ont une phase de prise en main relativement longue, contrairement aux stratégies 6 et 11.

Ces observations peuvent alimenter notre discussion précédente, même s'il est nécessaire de garder à l'esprit que l'interaction (stratégie sonore \times numéro de lancer) n'est pas significative, nous ne pouvons donc pas tirer de conclusions de cette discussion.

La prise en main de la stratégie 6 semble relativement rapide, les sujets ont probablement utilisé des caractéristiques sonores connues grâce à leur connaissance de bruits de roulement pour atteindre rapidement la phase de stabilisation. Ils n'ont potentiellement eu qu'à adapter le schéma cognitif lié à la manipulation d'objets roulants pour accomplir la tâche demandée.

À l'inverse, les sujets ont pu mettre plus de temps à maîtriser la sonification de la distance main-OV par le paramètre de hauteur à cause de son caractère non-naturel. Les sujets n'ont pas développé l'habitude d'utiliser le paramètre de hauteur dans ce contexte, et ont donc dû construire un schéma cognitif pour associer leurs gestes, les sons perçus et les résultats d'envoi de l'OV. Ce temps de constitution de schéma cognitif peut être à l'origine de la progression relativement lente de la stratégie 1 sur la figure 4.12.

Cette expérience peut être rapprochée de celle menée par Rath et Schleicher que nous avons évoquée précédemment, où il est question de guider une balle virtuelle vers une zone précise d'une barre en bois par le son [Rath et Schleicher, 2008]. Les stratégies sonores utilisées sont un son de roulement, et un son qualifié d'abstrait. La première obtient de bons résultats avant et après une phase d'entraînement, alors que la seconde donne tout d'abord de mauvais résultats, puis de bons résultats après l'entraînement. Dans certains cas, le son abstrait obtient de meilleurs résultats que le son de roulement, même si les résultats ne sont pas significatifs. En revenant à notre expérience, la sonification de la vitesse de l'objet par la brillance peut être qualifiée de naturelle, car les sons produits se rapprochent de sons de véritables objets roulants. Au contraire, la stratégie liant la distance main-OV à la hauteur peut être considérée comme abstraite car elle ne peut être rapprochée d'un phénomène physique. Nous pouvons alors supposer que ces deux stratégies sonores peuvent suivre des trajectoires similaires à celles observées dans l'étude de Rath et Schleicher : la stratégie naturelle obtient de bons résultats rapidement, alors que la stratégie abstraite nécessite un temps d'adaptation. La figure 4.12 montre des trajectoires singulières, même si ces différences ne sont pas significatives. La stratégie 6 semble avoir un apprentissage rapide, contrairement à la stratégie 1. Cette dernière stratégie ne semble pas naturelle, mais est adaptée à la tâche demandée, les sujets ayant intégré les informations portées par cette stratégie sonore après la phase 1.

De futurs travaux visant à expliquer plus précisément les liens entre apprentissage, performance et précision pourront se baser sur les travaux d'Éric Boyer, qui a notamment travaillé sur la rétention de tâches visuo-motrices avec retours sonores [Boyer *et al.*, 2017] [Boyer, 2015].

4.3.5 Conclusion

L'expérience que nous avons menée ici vise à hiérarchiser les stratégies sonores dont nous disposons, pour déterminer la stratégie transmettant le plus d'informations sur la dynamique de l'objet virtuel. Nous avons mis au point une tâche basée sur des gestes de balayages de la main, consistant à envoyer l'objet virtuel dans l'une des trois zones dont nous disposons. Nous avons choisi les trois zones avec un indice de difficulté constant, les zones les plus éloignées des sujets sont donc plus larges d'après la loi de Fitts. Nous avons utilisé les six stratégies sonores dites « simples » construites lors de l'expérience 1, auxquelles nous avons ajouté six stratégies dites « complexes » obtenues par combinaison de ces dernières ; ainsi qu'une stratégie de contrôle, sans son. Ces treize stratégies sonores ont été testées pour sonifier l'objet virtuel qu'il n'était pas possible de voir, et transmettre des informations sur sa dynamique. Deux phases ont été proposées : une première phase avec des lancers groupés, et une seconde phase avec des lancers distribués.

Les résultats montrent tout d'abord que plusieurs stratégies obtiennent des résultats significativement meilleurs que la stratégie sans son : le son est donc un vecteur d'information dans ces cas. En particulier, la stratégie liant la vitesse de l'OV au paramètre de brillance a obtenu les meilleurs résultats pour la phase avec lancers distribués, qui est la plus proche de la potentielle future utilisation en véhicule. La sonification de la distance main-OV par le paramètre de hauteur a également

retenu notre attention, par ses résultats singuliers : cette stratégie obtient les moins bons résultats en phase de lancers groupés, mais fait partie des meilleures stratégies sonores pour la phase de lancers distribués.

En rapprochant cette expérience des travaux de Rath et Schleicher, nous pouvons supposer que les stratégies 6 et 1 se rapprochent de sons naturels et abstraits respectivement [Rath et Schleicher, 2008]. Il serait alors intéressant de mesurer les performances avec ces deux stratégies après un long entraînement, pour voir si le phénomène d'inversion que semblent constater Rath et Schleicher se produit ici : la stratégie sonore abstraite peut-elle surpasser la stratégie sonore naturelle ? Nous avons donc choisi de conserver ces deux stratégies sonores, pour les étudier plus spécifiquement. Nous déciderons par la suite de la stratégie sonore à retenir pour l'illustration de l'objet virtuel pour utiliser l'interface gestuelle.

Conclusion du chapitre

La construction de l'objet virtuel a été au centre de nos préoccupations dans ce chapitre. Nous avons tout d'abord posé plusieurs hypothèses sur la construction de cet objet : nous l'avons choisi de forme sphérique, fixé sur un axe gauche-droite, et lié à la main de l'utilisateur par un ressort. Nous avons ensuite choisi deux paramètres de l'objet virtuel qu'il peut être intéressant de transmettre aux utilisateurs, à savoir la vitesse de l'OV et la distance entre la main et l'OV ; et trois paramètres sonores, la hauteur, la modulation d'amplitude et la brillance, qui peuvent être de bons vecteurs d'information. Nous avons choisi de nous focaliser sur un geste de balayage de la main, qui met en valeur des mouvements de l'OV différents de ceux de la main.

Dans une première expérience, nous avons mis au point une tâche audio-visuelle proposant de juger entre deux mappings sonores différents d'une même stratégie sonore. Le but était de déterminer le mapping s'associant de manière la plus naturelle au visuel proposé de l'objet virtuel. Après avoir constaté des résultats plus ou moins prononcés selon les stratégies sonores, nous avons choisi le mapping ayant obtenu les meilleures performances pour chaque stratégie de sonification pour la suite de l'étude.

La question du choix de la stratégie sonore transmettant le plus d'informations sur la dynamique de l'objet virtuel s'est alors posée. Pour permettre des manipulations de l'objet virtuel sans visuel ni retour tactile, il faut en effet une stratégie de sonification riche en informations. Nous avons donc mis au point une tâche permettant de juger le caractère informatif des stratégies sonores. L'analyse d'un indice que nous avons mis au point, mêlant les performances et la distance entre la position d'arrêt de l'OV et le centre de la zone, a permis d'observer plusieurs comportements intéressants. En particulier, la stratégie liant la vitesse de l'objet au paramètre de brillance semble prometteuse, et nous avons vu qu'elle peut se rapprocher d'un son de roulement. Cette proximité pourrait expliquer les résultats de la première expérience, où les sujets ont été particulièrement critiques avec cette stratégie sonore. La stratégie liant la distance main-OV au paramètre de hauteur semble également intéressante, même si les résultats de la première expérience étaient mitigés. Nous avons donc choisi de conserver ces deux stratégies pour la suite de l'étude.

Nous allons maintenant nous pencher sur la question de l'influence des stratégies sonores sur la rapidité de prise en main de l'objet virtuel : est-ce que la stratégie liant le paramètre de vitesse à la brillance permet réellement une prise en main plus rapide que la stratégie 1, comme semble le suggérer la figure 4.12 ? D'autre part, quelle est l'influence de la présence visuelle ou non de l'objet virtuel ? Cette question pourrait se poser lors de la première présentation du système aux sujets. Mais tout d'abord, nous allons nous pencher sur le design des sons. Nous voulons en effet proposer des sons plus plaisant à écouter que des sons basés sur du bruit rose. Le caractère agréable des stratégies sonores est par exemple requis pour que le système d'interface gestuelle soit accepté par les conducteurs. Ce nouveau design des sons posera alors des questions sur le côté évocatoire du son, que nous aborderons également.

Prise en main de l'objet virtuel

Dans le cadre d'une utilisation courante, les objets physiques sont manipulés aisément grâce à la présence de schémas cognitifs spécifiques. En prenant l'exemple d'une balle de tennis, les stimuli perceptifs qui lui sont propres sont connus : sa couleur, sa forme, son caractère rebondissant, le son associé au rebondissement ou encore les signaux tactiles liés à sa saisie. L'accumulation d'expériences passées permet d'anticiper les signaux que les utilisateurs vont percevoir à travers leurs sens, ainsi que leurs associations : une balle de tennis est a priori jaune, sphérique, relativement dure et rebondit avec un son reconnaissable. En revanche, lors de la présentation d'un nouvel objet physique, un temps d'adaptation est nécessaire pour comprendre les signaux perceptifs rattachés au nouvel objet, et les associer les uns aux autres.

L'introduction d'un objet qui a la propriété particulière d'être virtuel peut perturber cette aggrégation de signaux perceptifs. Notre ambition est de compenser cette perte par des informations sonores, et les deux premières expériences perceptives nous ont permis de mettre au point des stratégies de sonification qui semblent bien illustrer les manipulations de l'OV pour la suite de notre étude.

Au cours de ce chapitre, nous allons présenter des travaux initiant l'interface gestuelle que nous ambitionnons de mettre au point. Nous allons tout d'abord déterminer des applications automobiles qui permettront de structurer notre étude. Par la suite, nous partirons des stratégies sonores retenues au précédent chapitre pour effectuer un travail de design sonore, dans le but d'améliorer le caractère plaisant du son tout en conservant les informations transmises. Cette étape s'avère être d'une importance particulière, puisque les futurs utilisateurs de l'interface peuvent tenir compte du caractère agréable du son dans leur acceptation de l'IHM. À la suite de ce travail, nous nous sommes alors interrogés sur les évocations qu'il est possible de créer avec nos stratégies sonores pour des sujets n'ayant pas connaissance de l'existence de l'OV : quels concepts peuvent être perçus et compris par l'intermédiaire des sons ? Des sons ayant un fort pouvoir évocatoire peuvent en effet souligner la métaphore que nous voulons créer, et peuvent accompagner la création ou l'utilisation d'un schéma cognitif relatif à l'objet virtuel. L'expérience 3 permettra de s'interroger sur ces notions, où nos deux stratégies sonores seront jugées par des sujets naïfs.

En prenant en main notre objet virtuel, les utilisateurs de l'interface gestuelle vont devoir adapter des schémas cognitifs existants pour en créer un propre à l'OV. Celui-ci peut se baser sur la manipulation d'une balle réelle en modifiant les attributs relatifs aux perceptions tactiles et sonores par exemple. Une question se pose alors : de quelle manière présenter l'objet virtuel pour que le schéma cognitif soit facilement créé ? Nous tenterons de répondre à cette question dans l'expérience 4, où nous avons supposé qu'une présentation audio-visuelle est la plus pertinente dans un premier temps.

À la suite de ces travaux, nous choisirons l'une des deux stratégies sonores pour poursuivre notre étude. Nous pourrions ainsi nous concentrer sur les phénomènes attentionnels qui peuvent exister lors de la manipulation d'une interface conjointement à une tâche de conduite.

Sommaire

5.1	Premières briques de l'interface gestuelle	109
5.1.1	Choix de cas d'usages	109
5.1.2	Design des sons	111
5.2	Expérience 3 : évocations sonores	117
5.2.1	Évolution de l'objet virtuel	117
5.2.2	Hypothèses initiales et motivations	117
5.2.3	Méthodologie	117
5.2.4	Résultats	120
5.2.5	Discussion	126
5.2.6	Conclusion	127
5.3	Expérience 4 : prise en main des stratégies sonores	129
5.3.1	Hypothèses initiales et motivations	130
5.3.2	Méthodologie	130
5.3.3	Résultats	132
5.3.4	Discussion	143
5.3.5	Conclusion	148

5.1 Premières briques de l'interface gestuelle

Après une première étape de définition de l'objet virtuel, où nous avons posé différentes hypothèses de construction, nous nous sommes intéressés au son permettant de le représenter. Il est en effet nécessaire d'utiliser un contenu sonore illustrant au mieux le comportement dynamique de l'objet virtuel, pour qu'il soit possible de le manipuler sans informations visuelles. Nous avons ainsi pu isoler deux stratégies sonores qui sont particulièrement informatives pour la suite de l'étude : une stratégie liant la vitesse de l'objet virtuel au paramètre de brillance, et une stratégie liant la distance main-OV à un paramètre de hauteur du son.

Nous allons à présent commencer à construire notre interface gestuelle. Tout au long de ce travail de thèse, notre pensée s'est structurée autour de situations d'utilisations potentielles de l'interface gestuelle que nous avons imaginées, que nous allons désigner par le terme « cas d'usages ». Ils ont permis de construire progressivement l'IHM, en se posant les questions d'intégration des fonctions sélectionnées dans un contexte de conduite. Nous allons présenter les deux cas d'usages que nous avons choisis dans un premier temps : le premier est relativement courant, et le second plus prospectif pour mettre en application nos réflexions. Il sera cependant possible d'imaginer d'autres exemples d'applications.

Nous avons ensuite voulu travailler sur le contenu des stratégies sonores, pour éviter l'expérience désagréable de l'écoute d'un bruit rose qui peut contribuer à un effet de fatigue voire de rejet des utilisateurs de l'interface. Nous avons donc proposé un design des stratégies sonores, duquel doivent résulter des sons à la fois informatifs et plus agréables à écouter qu'un bruit rose.

5.1.1 Choix de cas d'usages

Nous avons voulu déterminer des cas d'usages, illustrant une utilisation possible de notre système dans un contexte automobile. Dans la mesure du possible, nous souhaitons nous focaliser sur des fonctions dont les interactions peuvent être améliorées. De cette manière, notre dispositif pourra éventuellement servir de preuve de concept pour le Groupe PSA, s'il est envisagé de poursuivre le développement de cette technologie. De plus, le choix de fonctions couramment manipulées par les conducteurs peut leur permettre de rapidement se rendre compte du potentiel gain d'ergonomie introduit par l'objet virtuel.

Fonctions fréquemment utilisées en conduite

Une étude interne réalisée par le Groupe PSA nous a orienté sur le choix d'une interface de ventilation. Les fonctions liées à cette IHM sont utilisées à de nombreuses reprises, ce qui peut souligner l'utilité de notre système. De plus, il existe plusieurs types de fonctions liées aux interfaces de ventilation : réglage de la température individualisé ou non, utilisation du dégivrage, orientation du flux d'air, ou encore réglage de l'intensité de ventilation. Ce nombre de fonctions permet de proposer un nouveau type de menu comme nous l'avons précédemment évoqué, où un geste de manipulation de l'objet virtuel peut être lié à une fonction précise. Un retour sonore permet ensuite d'informer l'utilisateur sur le geste effectué ainsi que sur la fonction manipulée. Dans ce premier cas, nous sommes dans la configuration d'une *tâche orientée geste*, où le son doit permettre d'accomplir le bon geste [Boyer *et al.*, 2013].

Comme nous l'avons évoqué dans le chapitre 1, les interfaces relativement anciennes donnent des retours d'informations visuels et tactiles par l'intermédiaire de boutons poussoirs ou rotatifs. Les véhicules récents munis d'un écran central multifonction renseignent l'utilisateur, lors de l'utilisation de fonctions liées à la ventilation par exemple, grâce à la modalité visuelle comme nous l'avons vu sur la Peugeot 308 de seconde génération – voir figure 1.2. Ces retours d'informations sont parfois associés à des sons, comme des *earcons* par exemple.

En mettant de côté les données transmises par l'interface, l'utilisateur peut avoir accès à un certain nombre d'informations grâce à ses différents capteurs sensoriels lors de la manipulation de la ventilation. L'intensité du flux d'air peut par exemple être estimée grâce à la pression exercée par le jet d'air sur la peau de l'utilisateur, ainsi que par l'intensité du son de la ventilation. Le positionnement du jet d'air dans le véhicule – sur les pieds, sur le torse, ou sur la tête – est également détectable par la sensation tactile de l'air expulsé sur les parties du corps concernées de l'utilisateur, pour peu qu'elles ne soient pas recouvertes. La température en revanche n'est pas souvent directement accessible, le système de ventilation nécessitant un temps d'adaptation pour modifier la température. Un retour d'information supplémentaire peut alors être pertinent, que nous allons donner par l'intermédiaire du son.

Nous avons également choisi un second cas d'application, qui concerne une interface de gestion de la musique. Nous pouvons ainsi nous baser sur le contenu sonore écouté par les utilisateurs pour donner un retour informatif direct aux utilisateurs. Cette application semble particulièrement illustrative des potentialités offertes par la sonification, car le son n'est plus seulement un intermédiaire pour donner de l'information, ce sont les modifications de ses propriétés qui sont recherchées. Dans ce cas de figure, nous sommes plus particulièrement dans une *tâche orientée son* que nous avons abordé en section 2.4.2, où le but des manipulations est d'obtenir un son avec les caractéristiques voulues [Boyer *et al.*, 2013]. L'utilisateur peut par exemple vouloir modifier l'intensité sonore du morceau écouté : nous pouvons alors proposer un geste pour manipuler ce paramètre, et le retour d'information est directement transmis par la musique. De plus, certaines manipulations comme le positionnement de la source sonore dans l'habitacle semblent être particulièrement adaptées à l'utilisation de gestes, et sont parfois difficilement accessibles dans les interfaces actuelles.

Cependant, nous gardons à l'esprit que l'utilisation de ces interfaces doit être simple et directe, pour que les conducteurs puissent focaliser leur attention sur la tâche de conduite. Nous avons donc décidé de proposer ces deux interfaces basées sur un même principe d'interaction. De plus amples précisions seront données au chapitre 6, où le détail de la construction des interfaces sera abordé. Nous avons donc cerné deux applications où l'utilisation conjointe de gestes et de sons semble pertinente, l'une *orientée geste*, l'autre *orientée son*. Dans le but d'implanter l'interface gestuelle dans de futurs véhicules, il est nécessaire que les retours d'informations sonores soient à la fois utiles pour manipuler l'objet virtuel, mais aussi agréables. Il est donc important de s'intéresser également à l'esthétique des stratégies sonores que nous allons proposer. Nous allons maintenant aborder cette question de design des sons, avant de proposer des nouvelles expériences perceptives.

À partir de maintenant, nous structurons nos travaux plus particulièrement autour du premier concept d'interface de ventilation, cas plus général que l'interface de lecture de musique. Les enseignements que nous tirerons de nos travaux pourront alors être appliqués à d'autres types d'interfaces – interface de GPS ou de téléphonie par exemple.

5.1.2 Design des sons

Dans le chapitre précédent, nous avons choisi de travailler les stratégies de sonification d'un point de vue général, avec un bruit rose. Nous avons ainsi pu étudier spécifiquement l'apport d'informations des stratégies sonores, et nous nous sommes cantonnés à cet aspect dans un premier temps.

Avec la progression vers une application industrielle automobile, nous avons maintenant choisi de définir une texture sonore spécifique. Elle pourra par exemple permettre aux utilisateurs d'avoir une expérience d'écoute plaisante, tout en véhiculant une image de marque pour le Groupe PSA. Il faut donc mener un travail sur l'esthétique des sons sur lesquels sont bâties les stratégies de sonification, tout en conservant leur apport d'informations.

Deux stratégies de sonification prometteuses ont été cernées, et nous avons proposé des explications quant à leur efficacité. Un travail de design des sons est donc possible en ayant à l'esprit les résultats des premières expériences. Par exemple, la stratégie de sonification de la vitesse de l'OV par le paramètre de brillance semble proche d'un son de roulement, nous allons prendre en compte ce constat pour proposer un son à la fois informatif et esthétique.

Objectifs

Il est nécessaire de revoir le design des sons que nous allons proposer pour illustrer l'objet virtuel pour plusieurs raisons. L'objectif est de proposer des stratégies sonores contenant les informations que nous avons étudié précédemment en premier lieu, tout en ayant un caractère agréable à écouter pour les utilisateurs. Comme notre système concerne des fonctions manipulées fréquemment, l'habillage sonore proposé doit être assez satisfaisant pour ne pas dissuader les conducteurs d'utiliser le système. Nous pouvons en effet supposer que l'interface gestuelle peut être rejetée simplement sur le caractère désagréable du son.

De plus, cette étape de design sonore peut permettre de proposer des sons qui ne seront pas masqués par les différents bruits présents dans l'habitacle. Par exemple, le bruit moteur provient principalement des explosions successives dans les différents cylindres du moteur, il est donc harmonique suivant la fréquence des explosions. La fréquence fondamentale de ce bruit se situe généralement entre 30 et 200 Hz, suivant les accélérations et les rapports de la boîte de vitesse. L'environnement sonore de l'habitacle est également composé du bruit de roulement, dont le contact entre le pneu et la chaussée en est à l'origine, qui est un bruit à large bande situé dans les basses fréquences, en-dessous de 500 Hz. Le bruit aérodynamique, dû aux écoulements d'air sur la carrosserie du véhicule, est le troisième type de bruit pouvant apparaître pendant la conduite. Il est également large bande, et ses composantes principales sont situées en-dessous de 1000 Hz. Si notre système est implanté dans un véhicule, il sera important de tenir compte de ces différentes sources de bruit pour proposer des sons facilement perceptibles, qui ne seront pas masqué par ces trois types de bruit.

Les sons peuvent enfin donner une image de marque à notre interface gestuelle, ce qui peut être important pour se démarquer de la concurrence de la part des acteurs industriels.

Stratégie sonore liant la vitesse de l'OV au paramètre de brillance

Comme nous l'avons vu précédemment, cette stratégie sonore semble se rapprocher d'un son de roulement. Nous avons donc voulu comparer les sons utilisés lors de l'expérience 2 avec ceux produits par le synthétiseur de sons naturels que nous avons présenté en section 2.3.4. Cet outil peut en effet nous permettre de générer facilement des sons de roulement, en contrôlant précisément certains paramètres comme la vitesse de roulement. Nous avons donc choisi un son contenant uniquement les invariants transformationnels du roulement, que nous qualifierons de son de roulement « pur », par opposition à des sons mêlant son de roulement et texture sonore.

Pour une vitesse nulle, le son généré par le synthétiseur a une intensité nulle, conformément à la réalité d'un véritable son de roulement. Nous allons cependant modifier ce comportement, en profitant de la flexibilité offerte par notre système numérique. Dans la future utilisation de l'IHM, les utilisateurs ne pourront pas voir l'objet virtuel, mais seulement l'entendre. Pour qu'il soit possible de percevoir l'objet virtuel même lorsqu'il est immobile, nous avons choisi de recalibrer les données d'entrée du synthétiseur, en faisant correspondre un faible roulement à une vitesse nulle. De cette manière, l'intensité sonore du son de roulement n'est pas nulle même si l'objet virtuel est immobile. Nous avons conscience que ce choix peut aller à l'encontre de notre section 2.3.2 plaçant l'événement à l'origine des sons, un son de roulement ne pouvant pas être produit sans déplacement. Nous pouvons alors imaginer que l'objet virtuel est animé d'un mouvement interne, à l'image du moteur d'un véhicule automobile par exemple.

Le paramètre sonore de brillance représente la distribution du contenu spectral d'un son, et provient d'une analogie au paramètre de brillance visuelle. Une manière de quantifier la brillance est de calculer le centre de gravité spectral, puisqu'une forte corrélation existe entre ces deux paramètres [Beauchamp, 1982]. Nous avons donc choisi de nous baser sur ce centre de gravité pour évaluer la brillance des sons dont nous disposons. Le calcul a été effectué pour le son utilisé lors de l'expérience 2, ainsi que pour le son de roulement pur généré par le synthétiseur. La figure 5.1 présente les résultats pour un lancé simulé – voir figure B.1 pour les données d'entrées de notre système. Le but est ici d'évaluer si ce paramètre spectral évolue de manière analogue dans les deux cas.

Dans un premier temps, nous pouvons observer des évolutions similaires entre les deux sons étudiés, même si l'évolution du centre de gravité est beaucoup plus nette avec le son de roulement pur : les différences sont de l'ordre de 1900 Hz entre le minimum et le maximum pour le roulement pur, et de 300 Hz pour la stratégie sonore de l'expérience 2. Nous pouvons alors formuler l'hypothèse que la stratégie sonore utilisée lors des précédentes expériences transmet des informations proches de celles véhiculées par un son de roulement, sans proposer des variations spectrales suffisantes. Nous pouvons observer les spectrogrammes des deux sons sur les figures B.2 et B.3 pour une étude plus détaillée. Le spectrogramme présenté en figure B.3 montre un contenu spectral beaucoup plus riche du son de roulement comparé au son utilisé dans l'expérience 2, notamment au-delà de 400 Hz. La stratégie liant la vitesse de l'OV au paramètre de brillance semble donc proche de celle utilisant le synthétiseur de roulement pur, sans toutefois être aussi présente dans les hautes fréquences. Nous supposons ici que l'utilisation du synthétiseur de roulement permet de transmettre convenablement les informations contenues dans la stratégie sonore alliant vitesse de l'OV à la brillance utilisée précédemment.

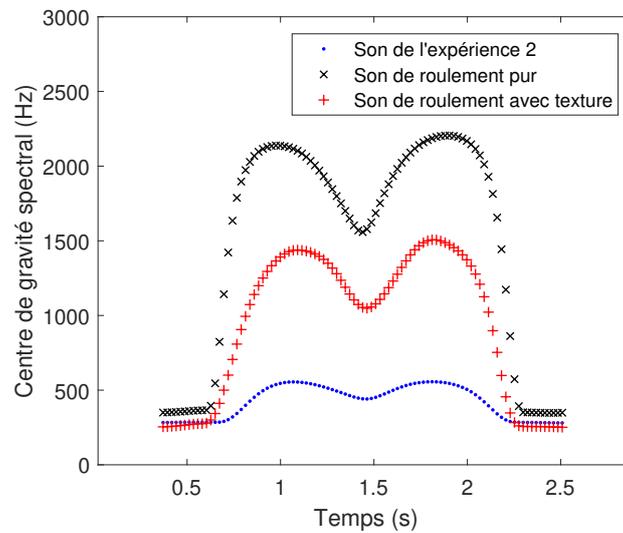


FIGURE 5.1 – Comparaison du centre de gravité spectral du son de la stratégie 6 de l'expérience 2 et de sons générés par le synthétiseur de sons naturels. Chaque point correspond à une fenêtrage du son par une fenêtre de Hanning d'une longueur de $2^{15} = 32768$ échantillons, avec un décalage de $2^{10} = 1024$ échantillons entre chaque point.

Nous pouvons noter que le son de roulement, grâce à un spectre plus riche en hautes fréquences, peut se discerner plus facilement que la stratégie sonore utilisée dans les expériences précédentes.

D'autre part, il est également nécessaire de rendre ce son agréable à écouter par les utilisateurs, comme nous l'avons précisé dans la section précédente. Une version du synthétiseur est pourvu d'un module permettant de remplacer les invariants structuraux provenant d'un matériau prédéfini d'objet résonnant par une texture sonore choisie – voir figure 5.2. Le son résultant est généré en effectuant une convolution du son de roulement par la réponse impulsionnelle de la texture – ce qui reprend le modèle source-filtre évoqué en section 2.3.4. Il est par exemple possible d'obtenir le son d'une balle roulant sur du bois en utilisant la réponse impulsionnelle de ce matériau. Dans le cas présent, nous n'avons pas choisi la réponse impulsionnelle d'un objet réel, qui ne laisse que peu de place au design sonore. Nous avons au contraire profité de la liberté offerte par les outils numériques dans la création de l'objet virtuel pour lui donner une identité sonore propre, qui soit reconnaissable et associée à l'OV. Cette signature sonore peut également intéresser les acteurs industriels pour véhiculer une identité. Nous avons donc choisi arbitrairement une texture sonore semblant agréable. Le spectrogramme de la texture sonore choisie est visible sur la

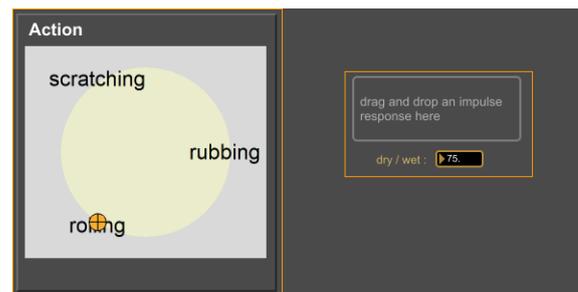


FIGURE 5.2 – Module du synthétiseur de sons permettant de convoluer une action à une texture sonore.

figure B.4. Le centre de gravité spectral du son obtenu, est représenté en par les $+$ en rouge sur la figure 5.1, et le spectrogramme est donné en figure B.5. Le spectrogramme présenté en figure B.5 révèle une évolution analogue au spectrogramme de la figure B.3 du son de roulement pur tout en étant moins marqué. La texture sonore est visible notamment entre les instants $t = 1s$ et $t = 2s$ par différents pics fréquentiels caractéristiques, visibles en figure B.4. L'évolution du centre de gravité spectral semble intermédiaire entre le son utilisé pour les expériences précédentes et le son de roulement pur, avec des différences de l'ordre de 1200 Hz entre le minimum et le maximum. Cette différence avec le son de roulement pur peut être expliquée par des composantes spectrales de la texture sonore moins prononcées entre 3 et 5 kHz comme on peut le voir sur la figure B.4, qui se retrouvent sur la figure B.5. Il est à noter que les composantes spectrales au-delà de 6 kHz semblent filtrées par le synthétiseur. Le son de roulement de l'OV paraît être un bon compromis entre la transmission d'informations permettant de ressentir les mouvements de l'objet virtuel, tout en ayant un son plus agréable qu'un son de roulement seul.

Nous avons donc vu que le son de la stratégie sonore liant la vitesse de l'OV à un paramètre de brillance se rapproche du son de roulement généré par le synthétiseur de sons naturels dont nous disposons. Les spectrogrammes ainsi que l'évolution des centres de gravités spectraux montrent que le son de roulement pur induit une évolution de la brillance plus marquée qu'avec la stratégie sonore que nous avons utilisée jusqu'à présent. Nous faisons l'hypothèse que le son de roulement permet de transmettre les informations dynamiques du comportement de l'OV d'une manière semblable à la stratégie sonore employée jusqu'ici.

De plus, le son de roulement peut permettre de souligner la métaphore de manipulation d'un objet que nous voulons mettre en place. Le son de roulement combiné à une texture sonore semble être un bon compromis entre l'efficacité de transport d'information comme pourrait l'être un son de roulement, et le caractère agréable de la texture sonore choisie.

Après avoir travaillé le son de la stratégie liant la vitesse de l'OV au paramètre sonore de brillance, nous allons maintenant discuter du son liant la distance main-OV à la hauteur. Nous allons donc devoir déterminer une méthode pour combiner les informations transmises par cette stratégie sonore à une texture sonore à identifier.

Stratégie sonore liant la distance main-OV au paramètre de hauteur

Pour cette stratégie sonore, nous désirons également travailler sur le design sonore pour proposer un son plus agréable que le filtrage d'un bruit rose. Le paramètre sonore manipulé est le paramètre de hauteur, que nous avons jusqu'à présent modifié par le changement de la fréquence centrale d'un filtre passe-bande. Nous allons également utiliser une texture sonore, dont il nous faudra modifier la hauteur pour transmettre l'information de distance. Pour modifier la hauteur globale de la texture, nous avons utilisé un algorithme du logiciel Max permettant un décalage fréquentiel – sans pour autant modifier la durée du son comme pourrait le faire la modification de la vitesse de lecture.

Pour choisir la texture sonore, il est important de tenir compte à la fois de l'aspect agréable de la texture de base, mais également du fait que le décalage fréquentiel de la texture va grandement déformer le son. Par exemple, un pic fréquentiel localisé à 100 Hz va se retrouver à 200 Hz pour un facteur de transposition de 2, soit un changement de 100 Hz. Pour un pic à 1000 Hz, la même transposition le déplace quant à lui à 2000 Hz, soit un changement de 1000 Hz. Une texture sonore riche est alors grandement dénaturée par le décalage fréquentiel, et peut perdre son côté agréable. Nous nous sommes donc focalisés sur un son avec peu de composantes spectrales dans les hautes

fréquences, comme le montre le spectrogramme présenté en figure B.6. Cette figure montre que les principales composantes de la texture sont en-dessous de 600 Hz. Il est également possible de voir de petites modulations d'amplitudes à certaines fréquences, qui ont été introduites pour éviter d'avoir un son monotone.

Nous avons ensuite dû définir une plage de transposition de la texture sonore choisie. Dans le même esprit que pour l'expérience précédente, nous avons décidé arbitrairement de faire varier le son de 1,5 octave. Pour comparer le son obtenu au son de la stratégie de l'expérience 2, nous avons également choisi de calculer le centre de gravité spectral. Dans le cas du bruit rose filtré, le centre de gravité représente bien la fréquence centrale du filtre. Dans le cas de la texture sonore, le centre de gravité spectral permet de représenter les variations moyennes de hauteur du son. La figure 5.3 donne l'évolution du centre de gravité spectral dans les deux cas. Les spectrogrammes sont également accessibles en figures B.7 et B.8, pour une représentation détaillée des sons.

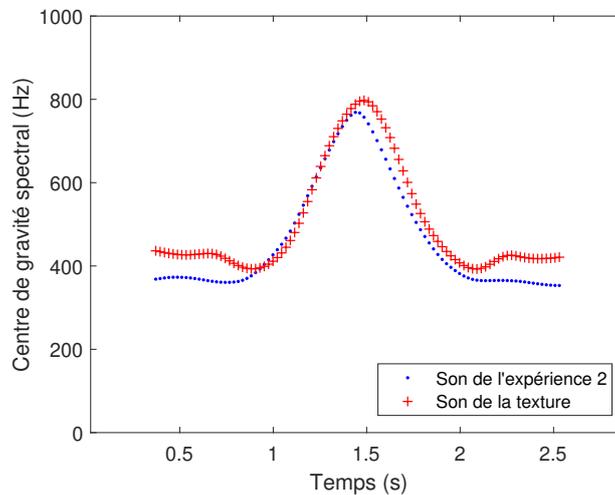


FIGURE 5.3 – Comparaison du centre de gravité spectral du son de la stratégie 1 de l'expérience 2 et de la texture sonore avec décalage fréquentiel. Chaque point correspond à un fenêtrage du son par une fenêtre de Hanning d'une longueur de $2^{15} = 32768$ échantillons, avec un décalage de $2^{10} = 1024$ échantillons entre chaque point.

L'évolution des centres de gravité spectraux est similaire pour les deux sons, avec un faible décalage fréquentiel. Ils se situent aux alentours de 400 Hz lorsque la distance est nulle, et ils atteignent 900 Hz au maximum de distance.

La figure B.7 montre le spectrogramme de la stratégie sonore utilisée en expérience 2. Nous pouvons très nettement distinguer la fréquence centrale du filtre qui se déplace entre les instants $t = 1s$ et $t = 2s$. Dans ce cas, l'évolution de la fréquence centrale est identique à celle du centre de gravité spectral de la figure 5.3. Le spectrogramme de la figure B.8 révèle l'évolution fréquentielle de la texture sonore lors du décalage fréquentiel. Avant l'instant $t = 1s$ et après $t = 2s$, il est possible d'observer la texture sonore non transposée, comme sur la figure B.7. En revanche pour $1 < t < 2$, nous pouvons constater la transposition, avec des décalages des pics fréquentiels plus ou moins importants selon la fréquence du pic original. Nous pouvons constater que les spectrogrammes des figures B.7 et B.8 évoluent de manière similaire.

Les centres de gravité spectraux des deux sons étant similaires, nous avons jugé que la modulation

de la hauteur de la texture sonore permet de transmettre des informations comparables à celles que véhicule la stratégie sonore basée sur le bruit rose que nous avons utilisé dans les expériences précédentes. Nous conservons donc cette nouvelle stratégie pour la suite de notre étude.

Nous avons donc redéfini les stratégies sonores que nous utiliserons dans la suite de l'étude. Dans un souci pratique, nous allons dorénavant désigner respectivement par *sonification de la distance* la stratégie de sonification de la distance main-OV par le paramètre de hauteur ; et par *sonification de la vitesse* la stratégie liant vitesse de l'OV au paramètre de brillance. Nous avons conservé les informations qui ont été utilisées dans les précédentes expériences, tout en redéfinissant les sons utilisés. Deux textures sonores ont été mises au point pour augmenter le caractère agréable des stratégies sonores, dans le but que l'écoute des utilisateurs soit la meilleure possible. Nous pensons en effet qu'un son désagréable peut tout d'abord perturber les conducteurs dans leur utilisation de l'interface gestuelle. Nous pouvons donc constater l'importance que revêt la tâche du design sonore dans la création de nouveaux systèmes. Les acteurs industriels utilisent également le design sonore pour créer une identité de marque à leurs produits. Les sons que nous avons proposé peuvent être remodelés selon les besoins des acteurs industriels, mais il est essentiel que les questions de transmission des informations sonores soient posées, pour que l'interface puisse être utilisée convenablement.

Pour aller plus loin, il peut être intéressant de se pencher sur certains travaux de Philip Tagg à propos de la sémantique sonore, qui concernent certains critères signifiants de la musique – appelés *museme* [Tagg, 2005].

Pour évaluer la production résultant d'un design sonore, il peut être intéressant de se pencher sur les outils développés au sein du projet européen CLOSED¹ [Susini *et al.*, 2006] – pour *Closing the Loop of Sound Evaluation and Design*. Ce projet visait à établir des procédés de mesures liés aux dimensions esthétiques et émotionnelles d'une production sonore provenant de designers sonores, pour les aider dans leurs travaux de production. Les différents livrables de ce projet peuvent guider le design du son lié à l'objet virtuel.

Nous allons à présent nous intéresser au pouvoir évocatoire des sons que nous avons mis au point. Il est en effet possible de véhiculer différentes notions de mouvement grâce au contenu des sons, comme nous l'avons vu dans le chapitre 2 au travers de différentes études [Shove et Repp, 1995] [Eitan et Rothschild, 2011] [Merer, 2011]. Dans l'industrie automobile, une question importante est soulevé par la sonification extérieure de véhicules électriques [Petiot *et al.*, 2013], et des évocations qui peuvent y être associées [Bosc *et al.*, 2014] : le son diffusé doit-il évoquer un danger, l'arrivée d'un objet lourd et imposant ?

Dans notre cas, il peut être intéressant de s'interroger sur les évocations qui peuvent découler du maniement de l'objet virtuel. Quelle situation vient alors à l'esprit des utilisateurs de l'OV instinctivement ? La réaction des stratégies de sonification à leurs gestes peut-elle être mise en parallèle d'un contexte connu ? Il est alors intéressant de déterminer si les stratégies sonores peuvent conduire des sujets naïfs, ne connaissant pas l'existence de l'OV, à l'imaginer intuitivement lors de manipulations en aveugle. Ce constat pourrait permettre de montrer que le son illustre bien l'objet virtuel, et permet aux sujets de saisir la métaphore que nous entendons créer. De plus, ce son évocatoire peut souligner cette métaphore lors des premières utilisations de l'objet virtuel en conduite, ce qui peut se montrer bénéfique en terme de prise en main.

1. <http://closed.ircam.fr/>

5.2 Expérience 3 : évocations sonores

5.2.1 Évolution de l'objet virtuel

Dans les expériences précédentes, les balayages de la main n'étaient possibles que dans une seule direction, de la droite vers la gauche. Pour progresser vers une plus grande liberté d'interaction avec l'OV, nous allons faire évoluer cette limite.

Pour la suite de nos études, nous autorisons les balayages effectués avec la paume de la main, mais également ceux effectués avec le dos de la main. Un algorithme permet de faire passer automatiquement l'objet virtuel d'un côté ou de l'autre de la main lorsque la vitesse latérale de la main atteint un seuil que nous avons déterminé. Ainsi, lorsqu'un balayage est entamé à droite ou à gauche, l'objet virtuel se place automatiquement devant la main pour être poussé dans la direction voulue. Cette évolution permet de proposer des balayages dans deux directions, tout en conservant un unique objet virtuel.

5.2.2 Hypothèses initiales et motivations

Dans cette expérience, nous allons nous intéresser aux évocations qui peuvent être créées par des manipulations gestuelles associées aux sons que nous venons de redéfinir. Chez des sujets naïfs, nous pensons que la manipulation à l'aveugle de l'objet virtuel peut intuitivement conduire à deviner la présence de l'objet virtuel et des concepts qui y sont associés – axe gauche - droite ou présence d'un ressort par exemple. Il nous semble donc pertinent de faire manipuler l'objet virtuel à des sujets sans donner d'éléments précis sur la nature de l'expérience.

Nous avons également proposé une sonification directe de la main des sujets. Ce type de sonification a été écarté lors de la mise en place des stratégies de sonification, car il nous semblait moins pertinent d'un point de vue métaphorique que la sonification de l'OV. Nous sommes cependant intéressés par les évocations qui peuvent être créées, et ainsi observer si ce type de sonification peut faire sens pour les sujets. Nous faisons l'hypothèse qu'il leur sera difficile de rapprocher ce type de sonification d'une situation réelle, alors que la sonification de l'OV peut être plus proche d'un contexte connu. Nous pourrions ainsi constater grâce aux commentaires si l'utilisation d'un objet virtuel est acceptée, ou si un autre lien plus direct entre les gestes et les sons a un pouvoir évocatoire important.

5.2.3 Méthodologie

Panel de sujets

Vingt-huit sujets droitiers ont participé à cette expérience, quatre femmes et vingt-quatre hommes. Tous travaillent pour le Groupe PSA, ils sont âgés entre 19 et 55 ans – âge moyen : 36,6 ans ; écart-type : 10,6 ans – et n'ont pas déclaré de problèmes moteurs ou auditifs. Aucun d'entre eux n'avait participé aux précédentes expériences.

Dispositif expérimental

Lors de cette expérience, les sujets sont assis à une table sur laquelle est placé le capteur de gestes Leapmotion. Deux enceintes Fostex PM0.4d sont placées de part et d'autre d'un écran éteint placé en face des sujets, pour que les informations sonores données par le spatialisateur soient clairement perceptibles.

Stimuli

Dans cette expérience, les sujets ont été divisés en deux groupes distincts : quatorze sujets manipulent uniquement la sonification de la distance, et les quatorze autres sujets utilisent la sonification de la vitesse. À la différence de l'expérience précédente, les sujets ne manipulent pas les deux stratégies sonores l'une après l'autre, car il est essentiel que les sujets soient complètement naïfs lors de la première confrontation avec les stratégies sonores. Aucun stimulus visuel n'est proposé, pour évaluer uniquement les évocations provoquées par les retours sonores à la suite de leurs gestes. Comme évoqué précédemment, deux *types* de sonification ont été testés par l'ensemble des sujets : la sonification directe de la main et la sonification de l'OV, qu'il convient de distinguer des deux stratégies sonores.

Les stimuli sonores relatifs à la sonification de l'OV sont ceux qui ont été décrits en section 5.1.2. En revanche, nous avons dû adapter les liens entre les gestes et les sons pour une sonification directe de la main. Pour la sonification de la distance, nous avons choisi de lier la distance entre la position horizontale de la main et une référence au paramètre sonore de hauteur. La position horizontale de référence se trouve au niveau du capteur de geste. De cette manière, le paramètre de hauteur est minimal lorsque la main est au centre, et maximal lorsque la distance au centre dépasse un seuil que nous avons fixé arbitrairement. En ce qui concerne la sonification de la vitesse, nous avons choisi d'utiliser le paramètre de vitesse horizontal de la main. Ces deux ajustements nous ont semblé les applications les plus directes des deux stratégies de sonification de l'OV à une sonification directe de la main.

Procédure

Tout au long de cette expérience, il est demandé aux sujets de faire le maximum de commentaires sur leurs impressions et leurs compréhensions du système, pour estimer les concepts saisis. Chaque sujet manipule les deux types de sonification – sonification de la main ou de l'OV – dans un ordre contrebalancé.

Pour chacune de ces stratégies, il est tout d'abord demandé aux volontaires d'explorer l'espace avec leur main droite librement, et un retour sonore spatialisé leur est proposé. L'expérimentateur oriente ensuite les sujets vers des gestes de balayage de la main vers la gauche ou vers la droite, pour se concentrer sur le geste que nous étudions. À la suite de cette séquence, une première série de questions est posée aux sujets :

« Quel est votre ressenti par rapport à la relation entre les gestes de votre main et les sons que vous percevez ? »

Les réponses sont alors recueillies par l'expérimentateur, qui enchaîne ensuite par une deuxième question :

« À quoi cette association entre vos gestes et les sons vous fait-elle penser ? »

Ces questions ont pour but d'aider les sujets à exprimer leurs ressentis, en se focalisant tout d'abord sur les concepts de sonification : quels mouvements influencent les sons, quels sont les paramètres sonores modifiés, comment les gestes et les paramètres sonores sont-ils liés ? La deuxième question s'intéresse à l'évocation de contextes connus, en demandant aux sujets de faire un parallèle entre cette expérience de sonification des gestes en temps réel avec des situations de la vie courante auxquels ils ont déjà été confrontés. L'expérimentateur note l'ensemble des ressentis et commentaires des sujets de l'expérience, ainsi que les réponses aux questions.

Après avoir testé les deux types de sonification, il est demandé aux sujets d'évaluer le caractère agréable du contenu sonore de la stratégie de sonification grâce à une échelle de Likert en 4 points labellisés : *très désagréable, un peu désagréable, un peu agréable* ou *très agréable*.

Analyse des données

Les impressions et réponses notées par l'expérimentateur peuvent être analysées avec différentes techniques. Appelées « analyses de contenu », elles servent à faire des déductions valides et fiables à partir d'un texte grâce à son contexte [Krippendorff, 2012]. Ces méthodes d'analyse sont notamment utilisées dans le domaine des sciences humaines et sociales, et visent à décrire le contenu qui est communiqué, ses causes ou ses conséquences. Il est possible d'effectuer des analyses quantitatives, en décomptant par exemple la fréquence d'utilisation d'expressions ou de mots particuliers. Nous avons cependant décidé de faire une analyse qualitative, pour essayer de comprendre les concepts perçus par les sujets de l'expérience, et les évocations provoquées par l'écoute des stratégies de sonification. Nous allons donc analyser le discours des sujets de l'expérience sans recourir à des grandeurs mesurées.

L'analyse de contenu qualitative ne s'appuie pas sur des données chiffrées mais sur une interprétation du discours des sujets qui doit être la plus impartiale possible. Il s'agit donc de déceler les concepts décrits parfois de manière explicite ou *manifeste*, mais aussi d'interpréter des idées exprimées de manière implicite ou *latente* [Marks et Yardley, 2004] [Joffe, 2012]. Pour exprimer leurs idées et sentiments, parfois difficiles à formuler, les sujets peuvent user d'un vocabulaire symbolique qui doit alors être décrypté par l'expérimentateur, sans être mal interprété.

L'analyse thématique nous a semblé pertinente par rapport aux objectifs de l'expérience que nous avons formulé, à savoir déceler les concepts évoqués par nos stratégies de sonification. De nombreux ouvrages ont permis de connaître les spécificités de cette analyse [Braun et Clarke, 2006] [Miles et Huberman, 1994] [Marks et Yardley, 2004] [Boyatzis, 1998] [Guest *et al.*, 2011].

Le questionnaire qui a clôturé l'expérience visant à évaluer le caractère agréable des sons a quant à lui été analysé grâce à un test exact de Fisher.

5.2.4 Résultats

Analyse thématique

L'analyse thématique a pour objectif de comprendre et d'analyser des données textuelles en les structurant autour de thèmes, pour identifier ceux « qui sont importants dans la description du phénomène à l'étude » [Daly *et al.*, 1997]. Selon Boyatzis, ces thèmes sont définis au minimum par une description et une organisation des observations, et peuvent parfois aller jusqu'à l'interprétation de certains aspects du phénomène observé [Boyatzis, 1998]. Ils sont articulés autour d'un code, qui va permettre de les relier par des liens logiques : « l'analyse thématique est un processus d'encodage de l'information qualitative » [Boyatzis, 1998]. Ce code peut être composé de parties déductives – découlant de l'étude théorique – ou inductives – provenant des données de l'expérience. Nous avons donc mis au point un code, présenté figure 5.4. Cette mise en forme va nous permettre de structurer nos réflexions.

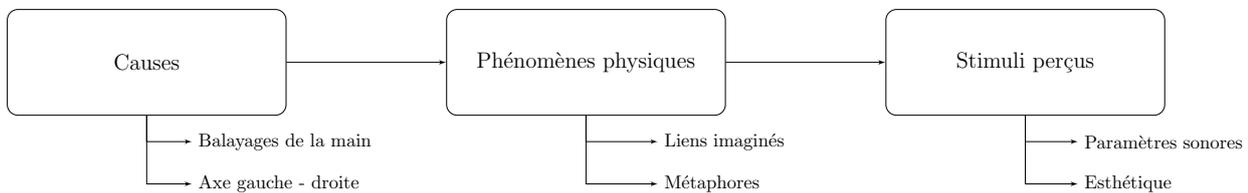


FIGURE 5.4 – Code de l'analyse thématique, bâti autour des phénomènes physiques illustrés par les stratégies de sonification, les gestes qui en sont les causes, et leurs conséquences sur les stimuli sonores.

Nous allons donc développer notre analyse thématique en nous focalisant sur chaque thème, et nous illustrerons nos réflexions en citant directement les sujets de l'expérience. Pour chaque thème, les distinctions entre les stratégies et les types de sonification sont faites. Dans notre présentation des résultats, nous allons présenter à la fois les réponses aux deux questions que nous avons posé, ainsi que les commentaires exprimés spontanément par les sujets.

Causes

Ce thème regroupe les éléments cités par les sujets se référant aux causes des fluctuations des stratégies sonores de cette expérience : nous allons particulièrement nous focaliser sur les gestes et sur les paramètres gestuels influençant les sons.

Globalement, il a été facilement compris que les variations des sons proviennent de mouvements latéraux de la main, pour les deux types de sonification et les deux stratégies sonores, même avant l'orientation de l'expérimentateur vers des balayages gauche - droite. Cependant, il semble que les sujets ont eu différentes interprétations sur d'autres aspects liés aux gestes.

SONIFICATION DE LA DISTANCE DE LA MAIN

Cette stratégie de sonification de la main a permis aux sujets de bien ressentir les mouvements latéraux, grâce à des changements sonores liés « à la distance au centre » (sujet 08, noté S08) : « plus on éloigne la main du centre, plus le son est aigu » (S03). Les sons semblent « proches des gestes » (S12) ce qui montre que « la position de la main est bien prise en compte » (S13). Plusieurs sujets ont apprécié la réactivité des sons par rapport aux gestes, ainsi que son « information précise » (S12).

Cependant, un sujet a trouvé que l'information transportée est « brouillonne » (S10) et que « le geste n'est pas bien reconnu ». Le sujet 01 a également été étonné par cette sonification, « liée à la position alors qu'il s'attendait qu'il soit lié à la vitesse » de la main.

SONIFICATION DE LA DISTANCE DE L'OV

La sonification de l'OV a divisé les sujets en deux catégories. Certains d'entre eux ont trouvé qu'il existe un « manque de synchronisation » entre gestes et sons (S11), « on a la sensation de perte de contrôle » (S02), cette sonification est « complexe, on a l'impression qu'il y a un bug » (S08). L'intermédiaire de l'objet virtuel semble être une source de perturbation.

D'autre part, plusieurs volontaires ont apprécié l'OV : « ça réagit tout de suite » (S03), « on comprend bien ce qu'on est en train de faire », « on sent que le geste est pris en compte (S13) ». La notion de distance par rapport à la main n'a pas été mentionnée, au contraire plusieurs sujets ont parlé de vitesse ou d'accélération : « il y a une notion d'accélération quand on fait le balayage de la main qui est appréciable » (S09), « la sonification est précise, en rapport avec l'accélération du geste » (S10), « le déplacement du geste est pris en compte plutôt que la position » (S13), « il se passe quelque chose quand on met de la vitesse, ou quand on accélère » (S01), « plus la main va vite plus le son accélère, le son est lié à la vitesse » (S14). Nous pensons qu'à travers ces témoignages, les sujets ont voulu exprimer les différences de sons constatés pour des gestes de même amplitude mais exécutés dans un temps plus ou moins long. Un sujet a également noté une particularité de cette sonification : « lorsqu'il y a deux gestes qui se suivent, on n'en entend qu'un » (S03).

SONIFICATION DE LA VITESSE DE LA MAIN

Dans cette configuration, nous avons choisi de sonifier la vitesse de la main des sujets grâce au synthétiseur de sons de roulement, ce qui a été globalement bien perçu par les sujets. « Le son varie avec la vitesse » (S18), « on ressent que la vitesse est prise en compte » (S15), « la vitesse de la main est indiquée de manière intuitive » (S27) sont des exemples des témoignages recueillis pendant cette phase de l'expérience. L'aspect réactif de cette sonification a également été apprécié, qui est « fidèle et qui suit vraiment le mouvement » (S23).

Par ailleurs, plusieurs sujets ont été perturbés : « on a pas l'impression d'arriver à la fin de quelque chose » (S16), « il n'y a pas d'information de confirmation de détection du balayage de la main » (S18), « on ne sait pas si le geste est bien été pris en compte » (S15). Ces témoignages peuvent exprimer l'étonnement des sujets vis-à-vis de cette stratégie, qui illustre leurs mouvements sans pour autant apporter d'information supplémentaire.

SONIFICATION DE LA VITESSE DE L'OV

Les volontaires de l'expérience ne semblent pas s'être focalisés sur le paramètre de la main qui est sonifié, même si le sujet 19 a mentionné que la « vitesse semble faire varier le son » (S19). Ils se sont plutôt concentrés sur la sensation « d'incohérence » (S16) ou de « confusion » (S15) qu'ils ont pu ressentir. Le sujet 16 explique plus précisément : « il y a une impression d'incohérence, mais vu que c'est répétable on comprend que ce n'est pas une erreur ». Même si la corrélation avec les gestes semble difficile à comprendre, il existe la « sensation qu'il y a une notion de fin » (S15). Il y a une impression de « temps de retard » qui peut être « embêtant pour les gestes rapides » (S26), mais « on sent qu'il y a un mouvement induit, ça donne une impression de mouvement » (S21). Le sujet 21 analyse plus précisément que « pour une vitesse lente le son suit la main, mais le décalage arrive pour des vitesses plus importantes ». Même si les mécanismes ne sont pas toujours compris – voir la section *phénomènes physiques*, les sujets semblent comprendre que le son véhicule une finalité du geste, et certains l'accueillent même positivement.

Phénomènes physiques

Nous allons maintenant nous intéresser au cœur de notre système : les liens entre les gestes et les stratégies sonores. Un intérêt particulier sera porté sur les concepts évoqués par les stratégies de sonification. Dans le cas de la sonification de l'objet virtuel, quels indices concernant ce phénomène physique sont perçus par les sujets ? D'autre part, quels sont les mécanismes pouvant être rapprochés de la stratégie de la sonification de la main ?

SONIFICATION DE LA DISTANCE DE LA MAIN

Pour certains sujets, cette sonification donne un « lien direct entre les gestes et les sons » (S04), « le son suit le geste » (S08). Les sujets semblent ici apprécier la relation facilement établie entre les gestes et les conséquences qu'ils ont sur le son. Cependant, cette analyse n'est pas partagée par tous, le sujet 05 trouvant par exemple que « le son n'est pas bien coordonné aux gestes ». Par la suite, peu de sujets ont réussi à rapprocher cette sonification de phénomènes physiques connus. Les sujets 03 et 05 ont pensé au Thérémine, instrument de musique électronique que nous avons cité plus haut. Cette référence semble pertinente, puisque les gestes influencent directement la fréquence des sons produits par l'instrument. En revanche, le jeu du Thérémine ne renvoie pas directement à une situation de la vie courante, mais à un instrument de musique électronique. Le sujet 10 a quant à lui cité plusieurs images intéressantes : « je pense à des balles d'un jeu de raquette qui sifflent quand elles se rapprochent de moi ». Cette image se réfère probablement au son produit par des balles qui fendraient l'air en passant à proximité du sujet. Cette image paraît juste, mais elle ne semble cependant pas tenir compte du geste, en s'appuyant uniquement sur le contenu sonore perçu. Ce sujet 10 évoque enfin un concept d'apesanteur et de champ magnétique, provenant probablement du point central où la fréquence est la plus basse, il indique que ce point particulier attire les gestes. Cette attraction est également évoquée par un autre sujet : « on se sent attiré vers les graves et le centre » (S14).

SONIFICATION DE LA DISTANCE DE L'OV

De même que dans la catégorie des explications des causes, la sonification de l'OV a été accueillie de deux manières différentes par les sujets. Plusieurs sujets n'ont pas réussi à comprendre le concept de balle virtuelle. L'un d'eux a eu « l'impression que le mouvement est un peu décorrélié du son » (S02), qu'il y a un « décalage temporel entre mouvement et son, une perte de contrôle, j'aimerais que le son soit synchronisé au mouvement » (S11). Ce temps de décalage provoque « parfois l'impression de ne pas être bien compris, même si je perçois le défilement de quelque chose » (S08). Ces témoignages montrent que les sujets ont perçu que leurs gestes ne sont pas directement reliés aux sons, même si le lien entre les deux n'a pas été déterminé. Le sujet 08 semble même comprendre une notion de défilement, même si elle reste assez floue.

Pour d'autres sujets cette sonification est plutôt appréciée, comme l'affirme le sujet 05 : « les gestes et les sons se coordonnent bien, si on prend le temps le son suit bien les gestes » (S05). Le sujet 01 évoque lui un son « synchronisé ». Certains concepts semblent bien saisis par des sujets, où « la direction du mouvement est bien comprise » (S11) et la « situation comprise les yeux fermés » (S09). Concernant l'évocation de situation connue, les sujets semblent ici créatifs dans leurs propositions : « une vague qui passe de droite à gauche » (S03), un « lancer de pages » (S02), une « roue qu'on lance » (S08), ou encore une « barrière magnétique, qui donne une sensation d'emprisonnement » (S10). Cette dernière description provient probablement du ressort, qui donne la sensation d'une évolution limitée du son, qui revient inexorablement à une origine. Les deux premières évocations semblent elles se focaliser respectivement sur les sons et sur les gestes, sans pour autant arriver à relier les deux. Le décalage entre les gestes et les sons est interprété « comme si quelqu'un d'extérieur répondait au geste » (S14).

SONIFICATION DE LA VITESSE DE LA MAIN

La sonification de la vitesse de la main n'a pas provoqué une grande quantité d'illustrations, elle ne donne « pas le sentiment de pousser quelque chose », mais elle donne « un sentiment de détection » (S17). L'image du Thérémine a été citée de nouveau – après le cas de la sonification de la distance de la main, même si des sujets différents ont testé les deux types de sonifications. Pour le sujet 25, le « son accroche un peu à la main », et cette impression donne lieu à une image du sujet 22, qui a « l'impression d'être dans l'eau ». Le fait de bouger la main plus ou moins rapidement à la surface ou dans l'eau provoque en effet des sons caractéristiques, et cette image semble intéressante à mettre en parallèle de notre stratégie de sonification. Par ailleurs, un grand nombre de sujets n'ont pas réussi à rapprocher la sonification d'une réalité physique.

SONIFICATION DE LA VITESSE DE L'OV

Même si certains volontaires n'ont pas réussi à illustrer leurs impressions concernant un phénomène physique lié à cette stratégie sonore, nous pouvons observer que certains sujets ont eu une grande compréhension des mécanismes à l'œuvre. Le sujet 25 a par exemple noté une « inertie du son », le sujet 17 est même allé plus loin : « on a le sentiment de pouvoir bouger l'air, de pousser de l'air, on a l'impression d'agir sur quelque chose de matériel sans regarder la main ». De la même manière, le sujet 22 a le sentiment d'un « mouvement de l'air localisé, on voit le déplacement de la perturbation de l'air », ce sujet détaille qu'il a une impression « d'amplification du bruit de l'air qui bouge, le son correspond au brassage d'air qu'on semble faire avec le geste, on voit le déplacement de la perturbation de l'air ».

Certains sujets ont réussi à imaginer des situations où ils pourraient retrouver les mêmes sensations : « il y a un effet de boomerang, qui accélère et revient vers la main » (S20), « il y a présence d'un écho comme une vague ou un yoyo » (S22), « il y a un effet de ressort ou d'élastique, comme si on tendait quelque chose qui revient » (S23). Plusieurs sujets ont même réussi à comprendre le concept de notre objet virtuel, en évoquant « une bille qui roule sur une table et qui revient » (S22), « une bille qui se déplace et qui revient au même endroit » (S23), un « objet roulant qui revient en arrière, avec un élastique et une inertie » (S18) ou encore « un bol dans lequel roule une bille, qui revient à son point de départ » (S22). Cette image du bol vise probablement à évoquer la force de rappel du ressort, et cette personne a dû penser instinctivement à la gravité. Certains sujets de cette expérience, naïfs et donc ne connaissant pas l'existence de notre objet virtuel, ont réussi à l'imaginer uniquement grâce aux retours sonores que nous avons proposés.

Stimuli perçus

Dans ce paragraphe, nous allons nous focaliser spécifiquement sur les informations concernant les stimuli sonores, en s'intéressant à la fois aux paramètres sonores des stratégies de sonification, mais également à l'aspect esthétique, que nous avons voulu travailler dans la section 5.1.2.

SONIFICATION DE LA DISTANCE DE LA MAIN

Les sujets ont tout particulièrement pointé que le son émis même lorsque la main n'est pas au-dessus du capteur est « perturbant » (S11), n'est pas « intuitif » (S01), « il est dommage qu'il reste un bourdonnement quand aucun geste n'est exécuté » (S10). Le sujet 10 évoque également un « brouhaha de son » lors de l'exécution de balayages, qui peut provenir des différentes modifications des fréquences fondamentales du son, voir le spectrogramme en figure B.8. La même impression paraît être ressentie par un autre sujet : « l'association fonctionne sur un geste assez lent, mais si on augmente la vitesse le son devient plus confus » (S06). Globalement, les sujets ont bien repéré que « plus on éloigne la main du centre, plus le son est aigu » (S03).

SONIFICATION DE LA DISTANCE DE L'OV

De nombreux volontaires ont noté l'aspect contre-intuitif de la présence d'un son alors qu'aucun geste n'est exécuté : « je suis surpris qu'il y ait un bruit constant » (S12), « le son reste présent alors qu'il n'y a pas de mouvement » (S11). Comme nous l'avons évoqué dans le paragraphe concernant les causes, l'influence de l'accélération sur les stratégies de sonification a été notée, avec une « hauteur du son qui dépend de la rapidité du geste » (S06). Le sujet 05 témoigne également de ce ressenti : « lorsque je fais un geste rapide, le son devient plus intense » (S05).

SONIFICATION DE LA VITESSE DE LA MAIN

Plusieurs sujets ont noté que « si on arrête de bouger la main, le son est joué continuellement » (S16), ce qui est « gênant » (S17), mais ce son constant peut également « montrer l'activité du capteur » (S22). Le sujet 18 a même trouvé que le son pouvait être « relaxant quand il est fixe » (S18). Le fait qu'un son soit joué constamment semble donc faire réagir les sujets, avec des opinions diverses. Une notion de hauteur a également été évoquée, avec un son « plus aigu quand on se déplace » (S23), le sujet 28 formulant que la « vitesse est reliée à une hauteur du son ».

SONIFICATION DE LA VITESSE DE L'OV

La stratégie de sonification a de nouveau suscité des réactions par rapport à son caractère continu. Le sujet 17 a par exemple trouvé que cette caractéristique donnait une « impression brouillonne ». Du côté des paramètres sonores, plusieurs sujets ont noté les variations des paramètres, même si les explications se sont parfois focalisées sur les phénomènes physiques. Il a été noté que « l'ampleur du mouvement est traduite par la fréquence », une « ouverture du son, mais pas très nette » (S15), et que « l'amplitude de la modulation sonore change avec l'amplitude du geste » (S28). Le sujet 19 a donné plusieurs explications précises, notant que la « vitesse de déplacement semble définir le volume et la fréquence du son », alors que « sans mouvement, le son est plat » (S19). Il est intéressant de noter que le sujet 21, qui n'a pas saisi le phénomène physique lié à la présence de l'objet virtuel, a trouvé que le son est « métallique ». Cet invariant transformationnel relatif à la nature de l'objet sur lequel roule l'OV a également été cité par le sujet 28, qui a proposé l'explication d'une bille roulante.

Analyse du questionnaire

Un test de Fisher a permis de comparer les réponses des deux stratégies sonores, et les résultats confirment un rejet significatif de la stratégie de sonification de la distance par le paramètre de hauteur $p < 0,01$, comparé à l'autre stratégie de sonification.

Le résultat des réponses à la question concernant le caractère agréable du son est présenté en figure 5.5. La stratégie sonore modulant le paramètre de hauteur du son semble particulièrement rejetée, puisque les sujets ont concentré leurs réponses sur les deux réponses plutôt négatives de l'échelle de Likert : *un peu désagréable* ou *très désagréable*. Les réponses sont en revanche plus mesurées pour la seconde stratégie de sonification, avec une répartition égale entre les réponses *très agréables* et *très désagréables*, et entre *un peu agréable* et *un peu désagréable*.

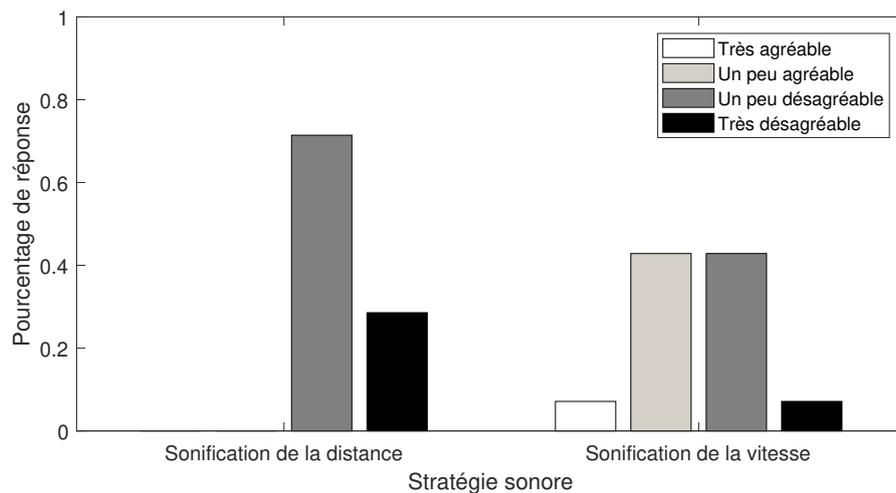


FIGURE 5.5 – Résultats du questionnaire concernant le caractère agréable des stratégies sonores.

5.2.5 Discussion

Nous avons décidé de focaliser notre analyse thématique sur trois parties distinctes de notre concept d'objet virtuel : les gestes causant des changements sonores, les phénomènes physiques sous-jacents, et les conséquences des gestes sur le retour sonore. La partie la plus intéressante du point de vue des évocations sonores concerne la compréhension du lien entre les gestes et les sons, même si les deux autres parties du code vont également nous aider à comprendre le cheminement intellectuel des sujets.

Concernant les stratégies sonifiant directement la main des sujets, les gestes causant les variations des sons ont été facilement déterminés, avec une réactivité souvent louée. La sonification de la main a cependant semblé frustrer certaines personnes, qui auraient aimé que le son valide leurs gestes et montre qu'ils sont bien pris en compte. Une simple icône sonore serait-elle suffisante, ou est-il attendu une réponse du système plus complète? Nous pouvons penser qu'ici se confrontent deux visions : faut-il privilégier la simplicité de compréhension du système, ou véhiculer des informations sur les conséquences des gestes? La réactivité appréciée plaide pour la première solution, alors que le manque de validation appuie la seconde.

Du côté des sonifications de l'OV, une division a également eu lieu illustrant la question que nous venons de soulever : certains n'ont pas compris le temps de retard qui pouvait exister entre leurs gestes et les sons, alors que d'autres l'ont apprécié car il semblait donner une confirmation du geste. Nous pouvons alors nous interroger si les deux discriminations des sujets que nous venons d'évoquer ne sont pas liées : les sujets appréciant les liens directs entre les gestes et les stratégies sonores sont-ils également ceux qui ne comprennent pas le temps de retard lié à l'OV? Au contraire, la frustration de non-confirmation des gestes de la sonification de la main est-elle comblée par l'utilisation de l'OV? Les commentaires des sujets ne nous permettent pas de confirmer ces hypothèses, mais il serait intéressant d'étudier plus précisément les liens existant entre ces notions.

Les idées évoquées concernant les phénomènes physiques des stratégies de sonification ont clairement montré que les sujets ont été plus inspirés pour les stratégies faisant intervenir l'objet virtuel. Plusieurs concepts imagés ont été développés avec la sonification de la distance de l'OV : la propagation d'une vague, ou le lancer de pages qui peut faire penser au lancer du carrousel d'une interface. Le « concept d'emprisonnement » cité par un sujet semble concerner le ressort qui retient l'OV, et de fait retient également la hauteur du son.

La sonification de la vitesse de l'OV a donné les descriptions les plus complètes du concept de l'OV, avec la description d'une bille roulante : ces sujets ont donc saisi que le son ne reflète pas directement leurs gestes, mais leurs conséquences. Les concepts de « bouger de l'air », de boomerang, ou de bille montrent que les sujets ont dissocié les sons qu'ils entendent et les gestes qu'ils font, et ont repéré un intermédiaire. L'évocation de l'inertie de l'OV par un sujet montre également que la physique de l'OV est ressentie. Enfin, les concepts liés au ressort sont également saisis par certains sujets, parlant de boomerang, élastique, bol ou ressort. L'ensemble de ces termes renvoie au concept de force freinant l'OV pour retourner à son état initial. Les stratégies de sonification de la main ont cependant rappelé à certains sujets des situations physiques, comme un concept d'apesanteur ou de champ magnétique : cette image peut être exploitée pour guider la main des utilisateurs par exemple, Parseihian ayant utilisé ce type de stratégie sonore pour localiser une cible [Parseihian *et al.*, 2016]. Enfin, le concept de son accrochant à la main ou l'impression d'être dans l'eau décrit par deux sujets semble assez juste par rapport à la sonification proposée.

Pour discuter des stimuli sonores, nous pouvons à la fois nous appuyer sur les commentaires des sujets présentés dans l'analyse thématique, mais aussi des réponses au questionnaire sur le caractère agréable du son. La principale caractéristique relevée concerne le caractère continu des stratégies de sonification. Les sujets n'ont pas semblé comprendre cet aspect du son, qui peut paraître désagréable lorsque l'OV n'est pas utilisé. Dans un contexte d'une écoute quotidienne, des événements sont à l'origine de la production des sons suite à une interaction d'objets entre eux. Le choix que nous avons fait peut donc sembler contre-intuitif. Ce choix peut donc être discutable, mais il permet de donner un retour d'information sur l'objet virtuel, qui ne pourra pas être vu par les sujets dans une éventuelle phase de conduite. Dans cette expérience, les sujets ne sont pas dans cette condition particulière, et ne connaissant pas l'existence de l'OV, ils n'ont pas pu comprendre l'apport de ce son. Nous pensons cependant que ces informations pourront s'avérer utiles dans la suite de nos expériences, notamment quand les sujets auront connaissance de l'OV et devront l'utiliser en aveugle. Les résultats visibles en figure 5.5 montrent que les sons créés à partir du synthétiseur de roulement ont été significativement plus appréciés que l'autre stratégie sonore, un sujet l'ayant même qualifié de relaxant. D'autre part, nous constatons qu'il est difficile de proposer un son agréable dont la hauteur est modifiée de manière importante, même si nous avons attaché une importance particulière à ce caractère agréable. Certains sujets, en évoquant la présence d'une bille, se sont focalisés sur les invariants transformationnels du roulement. Le synthétiseur transmet donc bien l'interaction de roulement, qui aurait pu être masquée par la texture sonore. Un sujet a de plus évoqué un invariant structurel en parlant de son métallique, et nous avons vu dans le chapitre 2 que la perception d'un matériau est notamment liée à la décroissance temporelle du son ainsi qu'à sa hauteur [Roussarie, 1999] [Klatzky *et al.*, 2000] [Avanzini et Rocchesso, 2001] [Giordano et McAdams, 2006]. Cette évocation provient donc probablement du choix de la texture sonore, mais aussi du mapping choisi en expérience 1 qui influence la décroissance du son : sur le spectrogramme de la figure B.5, on retrouve la forme de la décroissance de la figure 4.4a.

De manière générale, nous constatons que les sujets ayant saisi les phénomènes physiques ont choisi un vocabulaire se concentrant sur les événements à l'origine des sons, alors qu'une description plus proche des paramètres dynamiques ou acoustiques est adoptée dans le cas contraire. Ces choix de vocabulaire font échos aux études menées sur les écoutes musicales et quotidiennes de Vanderveer et de Lemaitre *et al.* [VanDerveer, 1979] [Lemaitre *et al.*, 2010].

5.2.6 Conclusion

Les évocations provoquées par les sons nous ont particulièrement intéressés dans cette expérience, où nous avons proposé aux sujets plusieurs stratégies de sonification. Il a été également possible d'expérimenter une sonification directe de la main ainsi qu'une sonification de l'objet virtuel. Nous avons comme but d'observer les évocations provoquées par chacune des situations, et nous sommes particulièrement attentifs aux concepts saisis par les sujets.

De manière générale, la sonification directe de la main a été comprise par les volontaires de l'expérience, avec des liens facilement perçus entre les gestes et les modifications des sons. Cependant, peu de situations évoquant des liens similaires entre les gestes et les sons ont été proposées, mis à part le Thérémine pour certains sujets, ou la métaphore de l'eau.

Les évocations sont beaucoup plus riches et nombreuses pour les stratégies de sonification de l'OV, et particulièrement avec la stratégie illustrant la vitesse de l'OV. Pour cette sonification, les présences de l'OV et du ressort permettant un retour à sa position initiale ont été formulées par certains sujets. En revanche, d'autres personnes n'ont pas apprécié les sonifications de l'OV car elles introduisent un temps de retard entre les gestes et les sons, qui paraît gênant lorsque la métaphore de l'OV n'est pas saisie.

Enfin, la texture sonore liée au synthétiseur de roulement a été significativement plus appréciée d'un point de vue esthétique que la seconde texture. Cette dernière a été particulièrement rejetée, aucun sujet n'ayant choisi les réponses *un peu agréable* ou *très agréable*.

L'ensemble de ces observations nous informe d'un pouvoir évocatoire plus fort de la stratégie sonore utilisant le synthétiseur de roulement, conjugué à un son plus agréable. Ces évocations peuvent être utiles lors des premières manipulations de l'OV en aveugle. En effet, l'absence de retours d'informations visuelles peut s'avérer déstabilisant, et des stimuli sonores soulignant la métaphore de l'objet virtuel peuvent être un atout pour la constitution de schémas cognitifs. Par la suite ces évocations peuvent aider les sujets lors de l'utilisation de l'objet virtuel, notamment lors des premières manipulations de l'interface gestuelle, en soulignant certaines caractéristiques de l'OV lorsqu'il n'est plus visible. Une représentation mentale de l'OV peut ainsi être plus facilement créée, permettant de meilleures manipulations.

À la suite de cette expérience, nous pouvons nous interroger sur la méthodologie à suivre pour présenter l'objet virtuel à des sujets novices. En particulier, comment amorcer la création d'un schéma cognitif relatif aux manipulations de l'OV lors des premières manipulations qui optimise sa future utilisation en aveugle ? Nous allons également nous intéresser aux deux stratégies sonores que nous avons choisies à la suite de l'expérience 2, pour n'en garder qu'une seule dans la suite de nos expériences.

5.3 Expérience 4 : prise en main des stratégies sonores

L'expérience 2 présentée précédemment se base sur une tâche de manipulation de l'objet virtuel, et nous a permis de classifier les stratégies sonores. À la suite de cette expérience, nous avons choisi d'en conserver deux qui nous semblent particulièrement informatives : une stratégie liant la distance main-OV à un paramètre sonore de hauteur, et la seconde mettant en relation la vitesse de l'OV au paramètre de brillance.

Après nous être limité à deux stratégies sonores, nous avons pu progresser vers une implantation en véhicule en travaillant le contenu sonore des stratégies, comme nous l'avons présenté dans la section 5.1.2. Cette étape nous permet, tout en conservant les informations transmises par les sons, de proposer des stimuli sonores plus agréables d'un point de vue esthétique que ceux basés sur du bruit rose. Il est en effet essentiel que les sons soient acceptés par les utilisateurs pour une utilisation du système à long terme.

L'expérience 3 nous a permis de constater les interprétations liées aux deux stratégies sonores que nous avons conservé. En particulier, la stratégie liant la vitesse de l'OV au synthétiseur de roulement a évoqué à plusieurs sujets ne connaissant pas notre système la présence d'un objet roulant, pouvant être manipulé par des gestes de la main. Ces évocations peuvent permettre de faciliter la prise en main de l'OV et de l'interface gestuelle, en rappelant la métaphore que nous avons choisi.

Nous allons effectuer un nouveau pas vers la mise au point de l'interface gestuelle, en reliant cette fois les manipulations de l'objet virtuel à des fonctions utilisées en véhicule automobile. Nous avons détaillé en section 5.1.1 les cas d'usages que nous avons ciblés pour développer notre IHM, et nous avons ici choisi d'appliquer les managements de l'OV au contrôle d'une interface de ventilation. En particulier, il nous a semblé pertinent de lier l'envoi de l'objet virtuel à des manipulations d'un axe de température de ventilation. En effet les gestes s'effectuent selon une unique dimension, l'axe gauche-droite, et la température est souvent représentée sur un axe qu'il est possible d'augmenter ou de diminuer. De plus, les zones que nous avons délimitées pour l'expérience 2 peuvent permettre des manipulations de différentes intensités, comme le montre la figure 5.6 : un geste avec une faible amplitude est lié à une modification de température de 1°C, alors qu'un balayage envoyant l'OV dans la zone la plus éloignée change la température de 5°C. Lors de cette expérience, nous avons choisi trois zones de la même manière que lors de l'expérience 2, même si les zones sont légèrement plus larges.



FIGURE 5.6 – Liens entre zone atteinte par l'objet virtuel et changement de température. Une représentation de la main et l'OV sont visibles au centre de la figure, au-dessus de l'affichage de la température. Les balayages de la main vers la gauche diminuent la température, alors que ceux vers la droite l'augmentent.

Nous avons choisi de ne pas tenir compte de la tâche de conduite lors de cette expérience : cette question est au cœur de l'expérience présentée dans le chapitre 6. Nous allons donc nous intéresser principalement à la prise en main de l'objet virtuel. Dans le cas d'une phase de prise en main de l'OV à l'arrêt, pour acquérir les schémas cognitifs relatifs à sa manipulation, les conclusions de la présente expérience pourront être utiles. En revanche, les conclusions que nous allons tirer de cette expérience ne pourront donc pas directement s'appliquer en situation de conduite, où la manipulation de l'interface gestuelle sera secondaire. Néanmoins, des enseignements pourront être tirés pour choisir la stratégie sonore illustrant l'objet virtuel, et ils pourront guider nos choix dans la mise en place de l'interface gestuelle.

5.3.1 Hypothèses initiales et motivations

Nous nous sommes interrogés sur la prise en main de l'IHM : comment proposer l'interface gestuelle lors de la première utilisation du système ? Dans cette première phase, les stratégies sonores que nous avons sélectionnées se différencient-elles l'une de l'autre ? Nous pensons qu'il est nécessaire, dans un premier temps au moins, de proposer le visuel de l'objet virtuel qui malgré des sons plus ou moins évocatoires, permet une bonne représentation de l'objet virtuel. Les utilisateurs peuvent ainsi avoir une image mentale de l'OV, et ils peuvent alors développer un schéma cognitif centré sur l'objet virtuel, en intégrant ses caractéristiques visuelles – forme, dynamique, force du ressort – ainsi que les stratégies sonores. Nous avons vu au chapitre 3 que ces informations multisensorielles sont traitées d'un côté par la boucle phonologique, et de l'autre par l'ardoise visuo-spatiale [Baddeley, 1992]. Plusieurs études ont montré que l'utilisation conjointe des modalités visuelle et auditive conduit à de meilleurs résultats que les situations unimodales [Penney, 1989] [Mousavi *et al.*, 1995] [Tindall-Ford *et al.*, 1997]. Nous pensons que leur association peut être bénéfique pour acquérir efficacement le schéma cognitif concernant l'objet virtuel.

Par la suite, nous imaginons une seconde phase où il n'est plus possible de voir l'objet virtuel, pour simuler une situation de conduite par exemple où nous désirons que les yeux du conducteur soient exclusivement concentrés sur la tâche de conduite. Dans ce deuxième cas, comment évoluent les performances de manipulation de l'objet virtuel ? Les informations sonores, associées au schéma cognitif créé peuvent alors permettre d'avoir une représentation de la dynamique de l'objet virtuel, et facilitent donc sa manipulation.

5.3.2 Méthodologie

Panel de sujets

Trente sujets ont participé à cette expérience, vingt-cinq hommes et cinq femmes, travaillant pour le Groupe PSA. Ils sont âgés entre 19 et 55 ans – âge moyen : 36,1 ans ; écart-type : 10,6 ans – et n'ont pas déclaré de problèmes moteurs ou auditifs. La plupart des sujets a participé à l'expérience 3.

Dispositif expérimental

Le dispositif expérimental est identique à celui de l'expérience précédente, à l'exception de l'écran qui est cette fois allumé. Il permet de montrer l'objet virtuel aux sujets de l'expérience, ainsi que l'interface proposée aux sujets. Des retours d'informations sur les tâches effectuées sont donc disponibles.

Stimuli

Les stimuli sonores sont identiques à ceux de l'expérience précédente, et de la même façon quinze sujets manipulent la sonification de la vitesse de l'OV, et les quinze autres testent la sonification de la distance. Pour chaque manipulation, la modification de température à effectuer est tout d'abord indiquée à la voix. Les sujets accomplissent alors le balayage de la main qu'ils pensent nécessaire pour envoyer l'OV dans la zone demandée. Deux icônes sonores indiquent ensuite si le lancer a été correctement effectué ou non. De cette manière, les sujets peuvent avoir un retour sonore sur leurs performances.

Dans la première partie de l'expérience, les sujets peuvent voir leur main, l'objet virtuel, ainsi que la température. Il est donc possible de voir le comportement de l'objet virtuel, et plus précisément sa dynamique liée au ressort. En revanche, dans la seconde partie de l'expérience, il n'est plus possible de voir la main ni l'objet virtuel. L'axe de température reste en revanche visible.

Procédure

Lors de la première partie de l'expérience, que nous allons qualifier d'audio-visuelle, nous avons permis aux sujets de voir l'objet virtuel, mais également de l'entendre pour observer si l'association multisensorielle du visuel et du son pouvait être bénéfique dans la prise en main de l'objet virtuel. Nous avons choisi une durée relativement longue de 120 essais, pour permettre d'observer l'évolution des performances sur une période plus longue que pour l'expérience 2. De plus, nous supposons que ce temps de manipulation est suffisamment long pour permettre aux sujets de développer le schéma cognitif centré sur la manipulation de l'objet virtuel. Ces 120 lancers sont effectués par 20 groupements de 3 lancers de chaque côté, où les 3 zones sont sollicitées dans un ordre aléatoire.

Une deuxième partie de l'expérience propose la même tâche que lors de la partie précédente, mais cette fois les sujets n'ont pas la possibilité de voir l'objet virtuel ou la reconstruction de leur main. Dans cette deuxième partie auditive seule, 54 lancers sont proposés dans un ordre aléatoire, avec 3 lancers de suite de chaque côté. Pour observer si les sujets ont bien acquis le schéma cognitif lié aux manipulations de l'objet virtuel, nous avons décidé de modifier le ressort par rapport à la première partie de l'expérience, en diminuant la force de rappel. De cette manière, les sujets adapteront leurs gestes s'ils ont bien acquis et interprété les informations sonores liées aux déplacements de l'objet virtuel.

Analyse des données

De la même manière que pour l'expérience 2, nous allons observer les résultats grâce à l'indice que nous avons introduit en section 4.3.2. L'indice est toujours défini par la formule :

$$I(x) = \begin{cases} 0 & \text{si } x \leq W_{zone}/2 \\ x + 1 - W_{zone}/2 & \text{si } x > W_{zone}/2 \end{cases}$$

où I représente l'indice, x la distance par rapport au centre de la zone, et W_{zone} la largeur de la zone. Il tient compte à la fois du succès du lancer, mais aussi de la distance entre le lancer et le centre de la zone en cas d'échec. Les zones ayant légèrement évolué – la limite de la zone la plus éloignée a été choisie à 0,9 au lieu de 0,8 pour essayer de rendre la tâche moins difficile – nous rappelons l'indice associé à chaque zone en figure 5.7.

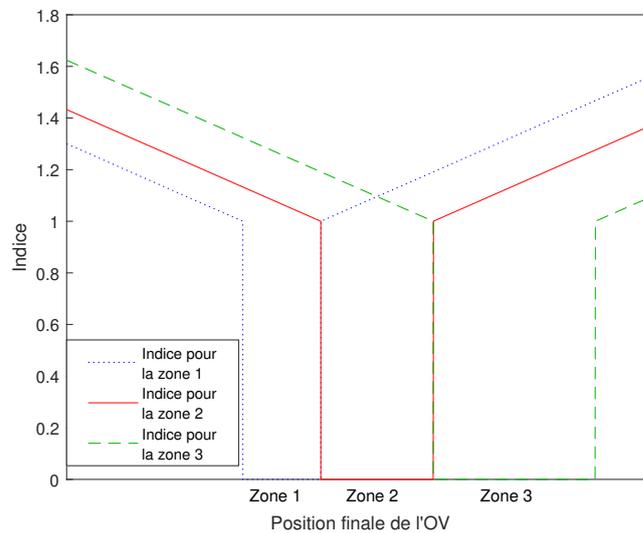


FIGURE 5.7 – Nouvelle construction de l'indice à partir des positions finales de l'objet virtuel.

5.3.3 Résultats

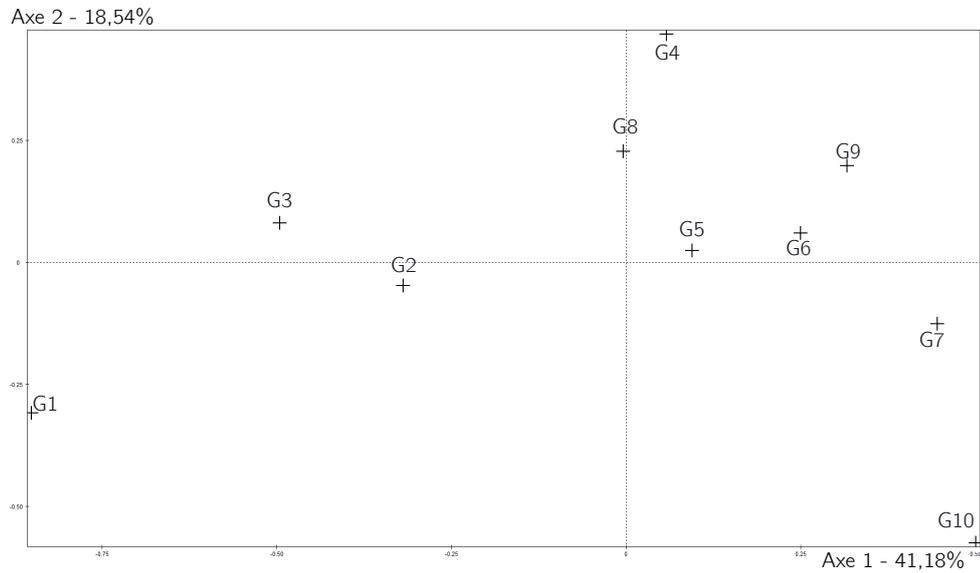
Effet des stratégies sonores dans la première partie de l'expérience

Dans cette section, nous nous intéressons particulièrement à l'évolution des performances dans le temps, en détaillant par stratégie sonore pour observer un éventuel effet des stratégies sonores sur cette évolution.

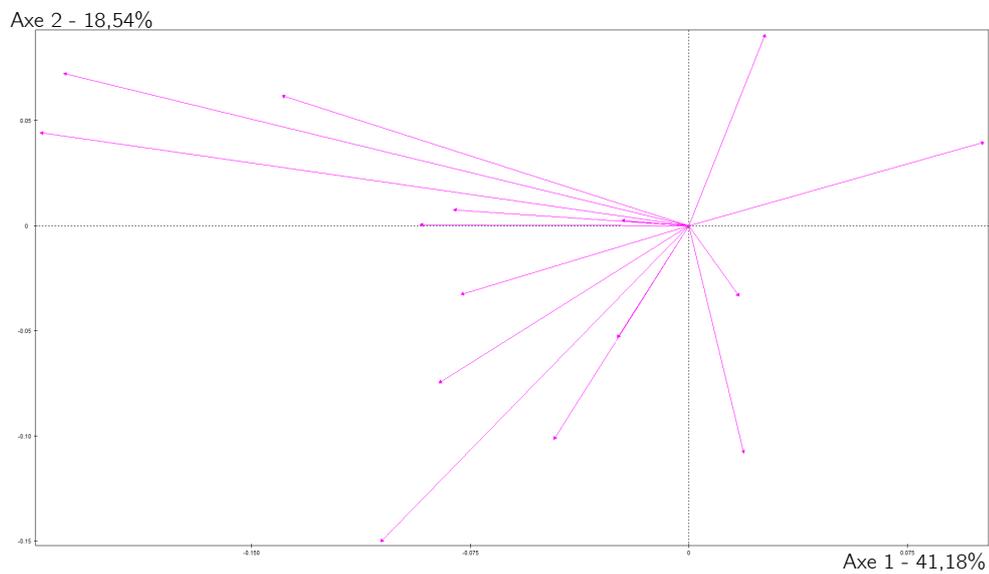
Nous avons choisi de grouper les lancers en 10 groupes de 12 pour faire abstraction des effets des zones, ou de l'effet de la latéralité des balayages de la main, pour se concentrer uniquement sur l'évolution de l'indice. Cette moyenne sur 12 essais permet également d'augmenter la lisibilité des résultats, et regroupe donc 6 lancers de chaque côté, et 2 lancers dans chaque zone. Nous allons tout d'abord nous intéresser au caractère consensuel des comportements des sujets grâce à une analyse en composantes principales – ou ACP.

ACP

Les résultats des ACP que nous avons réalisées sont présentées en figures 5.8 et 5.9, chaque figure représentant une ACP pour une stratégie sonore.

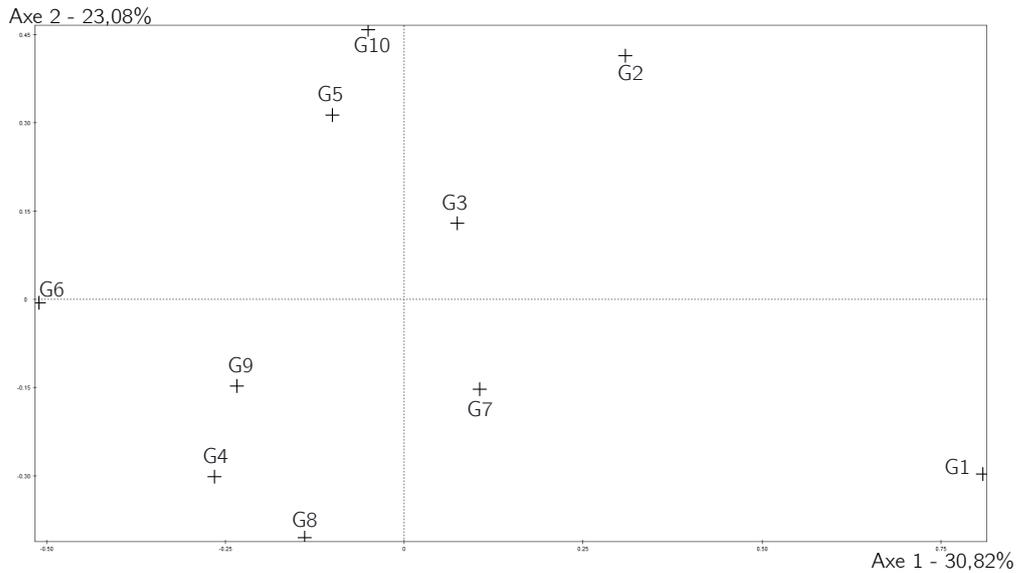


(a) Évolution moyenne de l'indice par groupe de lancers

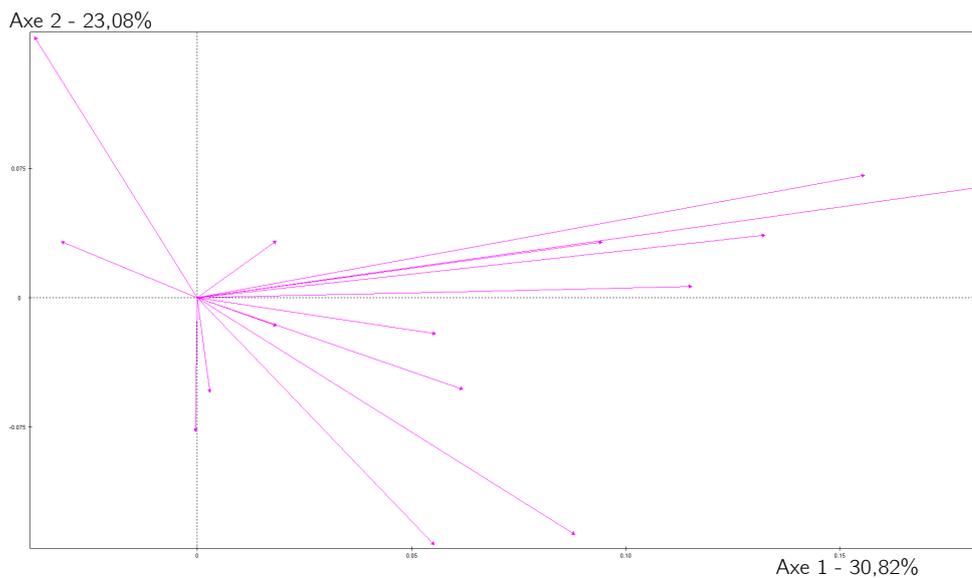


(b) Évolution moyenne de l'indice par sujet

FIGURE 5.8 – ACP réalisée sur l'indice des lancers successifs par rapport aux groupes de lancers, pour la stratégie sonore liant la distance main-OV à la hauteur.



(a) Évolution moyenne de l'indice par groupe de lancers



(b) Évolution moyenne de l'indice par sujet

FIGURE 5.9 – ACP réalisée sur l'indice des lancers successifs par rapport aux groupes de lancers, pour la stratégie sonore liant la vitesse de l'OV au paramètre de roulement du synthétiseur.

Concernant la sonification de la distance, nous pouvons tout d'abord noter les axes principaux expliquant à 41,18% et 18,54% la variabilité des résultats sur les figures 5.8a et 5.8b. Les deux axes principaux de la seconde stratégie sonore représentent seulement 30,82% et 23,08% des variations des résultats, comme nous pouvons l'observer sur les figures 5.9a et 5.9b. Les figures 5.8a et 5.8b permettront donc de décrire la variabilité des résultats avec plus de précision.

Pour lire les figures 5.8 et 5.9, il faut tout d'abord s'intéresser à la direction globale pointée par les flèches des figures 5.8b et 5.9b. Les têtes des flèches représentent les indices les plus élevés. En prenant la figure 5.8 en exemple, il semble que les indices les plus élevés se situent globalement sur la gauche de la figure 5.8b. Il faut alors « superposer » la figure 5.8b à la figure 5.8a : les indices les plus forts se trouvent sur la gauche de la figure 5.8a.

Les figures 5.8a et 5.9a présentent l'évolution de l'indice suivant les 10 groupes de 12 lancers que nous avons créés – notés Gi , $i \in \llbracket 1, \dots, 10 \rrbracket$. Sur les deux figures, il est possible d'observer un regroupement de points – sur la droite de la figure 5.8a et sur la gauche de la figure 5.9a. Cette concentration de points représente un phénomène de stagnation des performances, où les résultats paraissent équivalents à partir du groupe 4 sur la première figure, et du groupe 3 sur la seconde. Il semble donc qu'il y ait une amélioration des performances entre le groupe 1 et les groupes 4 et 3 pour les deux figures, respectivement. Ces résultats sont précisés dans l'ANOVA que nous allons présenter dans le paragraphe suivant.

Les figures 5.8b et 5.9b montrent le caractère consensuel des 15 sujets qui ont participé, chacun étant représenté par une flèche. Sur la première figure, nous pouvons observer qu'un faisceau de flèches se dirige vers la gauche, indiquant un comportement relativement similaire entre les sujets. Seules quatre flèches sont dirigées vers la droite, dont une seule a une composante importante selon l'axe 1. Nous considérons ces sujets comme ayant un comportement anecdotique, et nous effectuerons donc une analyse de variance sur l'ensemble des sujets. En ce qui concerne la figure 5.9b, trois flèches représentant trois sujets sont orientées dans une direction opposée à la majorité des flèches, dirigées vers la droite. De la même manière que précédemment, nous allons procéder à une analyse de variance sur l'ensemble des sujets.

ANOVA

À la suite des analyses en composantes principales, nous avons effectué une analyse de variance sur notre indice avec un facteur sujet imbriqué dans le facteur stratégie sonore – 28 ddl, un facteur groupe de lancer – 9 ddl, un facteur stratégie sonore – 1 ddl et une interaction entre les facteurs groupe de lancer et stratégie sonore – 9 ddl. Cette analyse va nous permettre d'observer plus précisément l'évolution des manipulations de l'objet virtuel par le panel de sujets.

Cette analyse montre un effet significatif du facteur *groupe de lancer* – $F(9,3552) = 5,495$; $p < 0,001$ – comme on peut le voir en figure 5.10. Un test post-hoc de Duncan témoigne d'une progression jusqu'au troisième groupe de lancers. Par la suite, les résultats semblent se stabiliser. Nous pouvons donc affirmer qu'une progression a bien lieu, mais au-delà du 36^e lancer les résultats semblent plafonner et ne s'améliorent plus. En ce qui concerne le facteur des stratégies sonores, aucun effet significatif n'est constaté – $F(1,3552) = 0,235$; $p = 0,628$. Les sujets de l'expérience n'ont donc pas obtenus de meilleurs résultats avec l'une des deux stratégies sonores lors de cette phase de l'expérience.

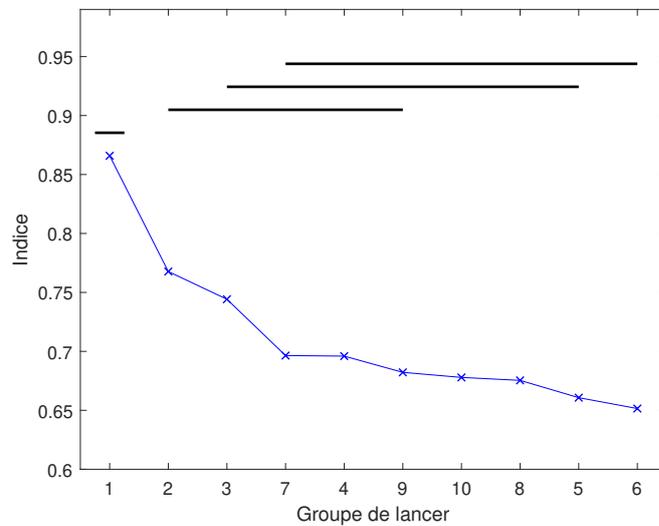


FIGURE 5.10 – Évolution moyenne de l'indice selon les groupes de lancers. Les barres indiquent les groupes où il n'y a pas de différences significatives.

Effet des stratégies sonores dans la seconde partie de l'expérience

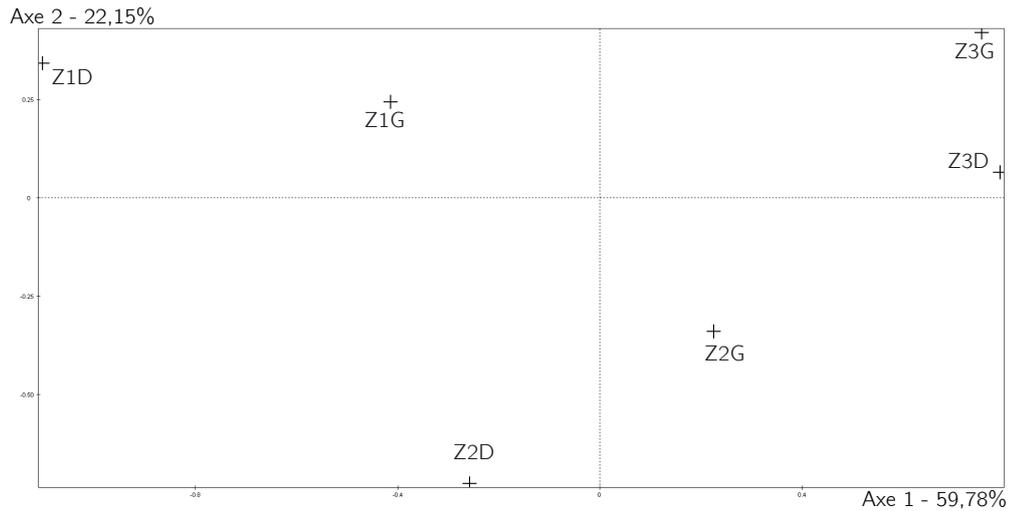
Nous allons à présent nous intéresser à la seconde partie de l'expérience. Le même type de tâche est demandé aux sujets, ils ne peuvent cependant plus voir l'OV : le comportement dynamique de l'objet virtuel est donc seulement perceptible grâce aux stratégies de sonification. Des icônes sonores sont également utilisées comme dans la première partie de l'expérience. Pour observer si le son apporte bien de l'information, nous avons choisi de diminuer légèrement la force de rappel du ressort. Ainsi, il est possible d'observer si les sujets adaptent leurs gestes grâce aux retours sonores : une stratégie sonore transmettant suffisamment d'informations permet de constater la différence de force de rappel, et donc d'adapter les balayages de la main en fonction.

Cette partie de l'expérience peut être mise en parallèle avec l'expérience 2 : il existe en effet un certain nombre de similarités. Cependant, ces deux situations diffèrent par la première phase de l'expérience, où les sujets peuvent voir l'objet virtuel à la différence de la précédente expérience. De plus, cette expérience s'inscrit dans un contexte concret d'interface automobile, où les manipulations de l'OV sont liées à une fonction, à la différence de l'expérience 2 qui s'inscrivait dans une situation moins appliquée et plus abstraite.

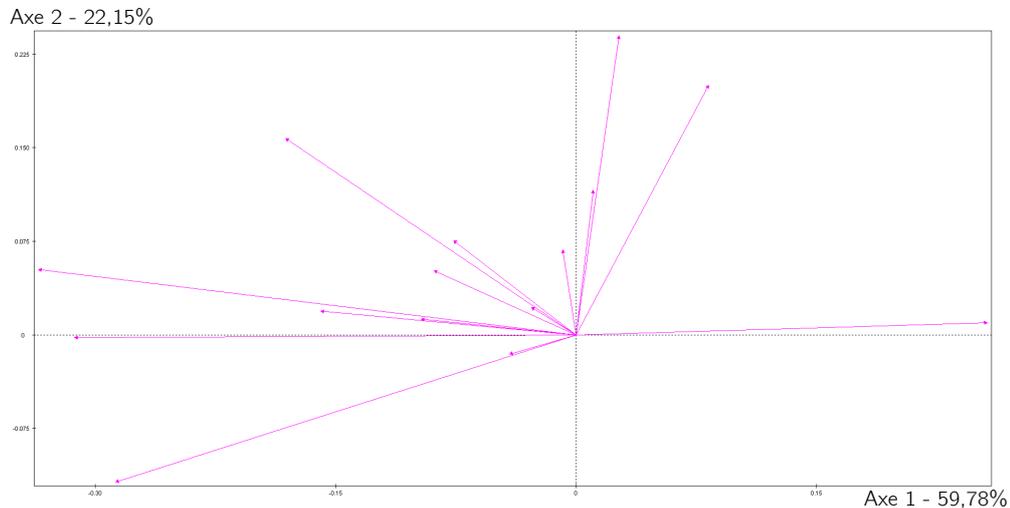
Dans cette section, nous allons particulièrement nous intéresser aux différences qu'il existe entre les stratégies sonores. L'aspect temporel important peu, nous avons choisi un ordre aléatoire des lancers. Ce choix permet de s'intéresser à la latéralité des balayages de la main, ainsi qu'à l'influence des zones dans les résultats.

ACP

Les figures 5.11 et 5.12 montrent les résultats des analyses en composantes principales que nous avons effectuées selon les différentes conditions de l'expérience. Sur les figures 5.11a et 5.12a, nous avons noté les zones à atteindre ZiG ou ZiD , avec $i \in \llbracket 1,2,3 \rrbracket$ respectivement les indices obtenus pour un balayage en zone i à effectuer sur le côté gauche ou droite. Les figures 5.11b et 5.12b montrent quant à elles le comportement des sujets.

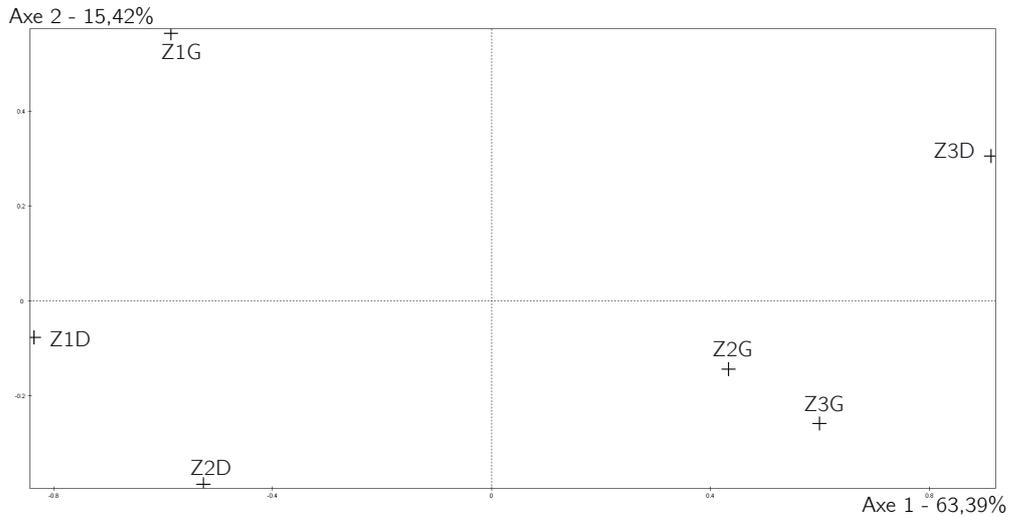


(a) Évolution moyenne de l'indice par condition

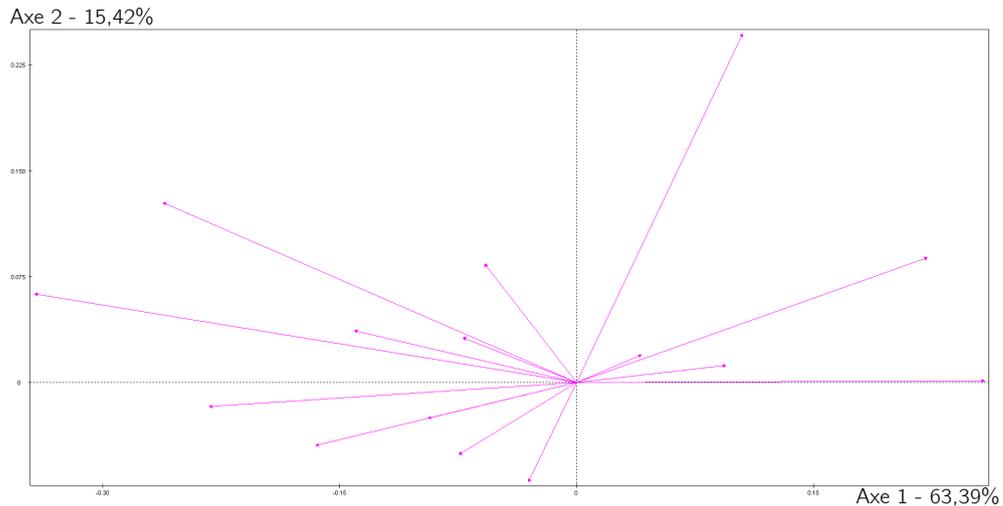


(b) Évolution moyenne de l'indice par sujet

FIGURE 5.11 – ACP réalisée sur l'indice par rapport aux conditions de l'expérience, pour la stratégie sonore liant la distance main-OV à la hauteur.



(a) Évolution moyenne de l'indice par condition



(b) Évolution moyenne de l'indice par sujet

FIGURE 5.12 – ACP réalisée sur l'indice par rapport aux conditions de l'expérience, pour la stratégie sonore liant la vitesse de l'OV au paramètre de roulement du synthétiseur.

Nous pouvons déjà remarquer que les axes 1 des deux figures expliquent environ 60% de la variabilité des résultats. Les axes 2 sont eux plus proches de 20% de variabilité, nous devons donc interpréter en premier lieu la répartition des résultats selon l'axe horizontal, prépondérante, avant d'observer la répartition suivant l'axe vertical.

En regardant la répartition des conditions latérales sur les figures 5.11a et 5.12a, nous pouvons voir que les lancers des zones situées à gauche et à droite Z_iG et Z_iD se situent globalement dans des zones voisines. Les conditions $Z2G$ et $Z2D$ de la figure 5.12a sont les seules où la différence est assez nette. Nous pouvons d'ores et déjà penser que l'aspect latéral des balayages de la main ne donnera pas de différences significatives, ce que nous précisons dans le paragraphe suivant avec l'analyse de variance.

En ce qui concerne les zones, il est possible de voir une répartition particulière : la zone 3 se situe sur la droite des figures 5.11a et 5.12a, la zone 1 plutôt sur la gauche, et la zone 2 entre les deux. Ces différences de positionnement selon les axes 1 semblent indiquer un effet significatif de la zone, que nous verrons plus loin également.

Nous allons à présent observer la répartition des sujets sur les figures 5.11b et 5.12b. Les deux figures montrent un faisceau de flèches orientées vers la gauche. Il existe cependant certaines flèches dirigées vers la droite, 4 sur la figure 5.11b et 5 sur la figure 5.12b, témoignant de comportements singuliers de certains sujets. Il aurait pu être intéressant de faire des analyses de variances séparées, cependant ces analyses ne sont pas pertinentes compte tenu du faible nombre de sujets. Nous avons donc décidé de faire une seule ANOVA avec l'ensemble des sujets de l'étude.

ANOVA

Nous avons procédé à une ANOVA sur l'indice avec un facteur sujet imbriqué dans le facteur stratégie sonore – 28 ddl, le numéro du lancer – 53 ddl, la stratégie sonore – 1 ddl, la zone – 2 ddl, la latéralité – 1 ddl, et les interactions entre la stratégie sonore et la latéralité – 1 ddl, entre la zone et la latéralité – 2 ddl, et entre la zone et la stratégie sonore – 2 ddl. Cette analyse permet de faire ressortir l'influence des stratégies sonores – $F(1,1477) = 13,913$; $p < 0,001$. La stratégie liant la distance main-OV à la hauteur est significativement plus efficace dans cette tâche que la sonification de la vitesse de l'OV, avec des indices moyens d'environ 0,54 et 0,64 respectivement.

L'analyse de variance confirme qu'il n'y a pas d'effet significatif sur le facteur de latéralité des balayages de la main dans les résultats de cette expérience – $F(1,1477) = 0,836$; $p = 0,361$. Les sujets semblent avoir réussi de la même manière les balayages pour baisser et pour augmenter la température. Il existe cependant des variations significatives concernant le facteur *zone* – $F(2,1477) = 19,962$; $p < 0,001$. Un test post-hoc de Duncan a permis de voir que les sujets ont eu tendance à plus facilement envoyer l'objet virtuel dans les zones 2 et 3 qu'en zone 1 – voir figure 5.13. En revanche, il n'y a pas de différences significatives entre les zones 2 et 3. Par ailleurs, les interactions (stratégie sonore \times latéralité), (zone \times latéralité) et (stratégie sonore \times zone) ne montrent pas d'effet significatif.

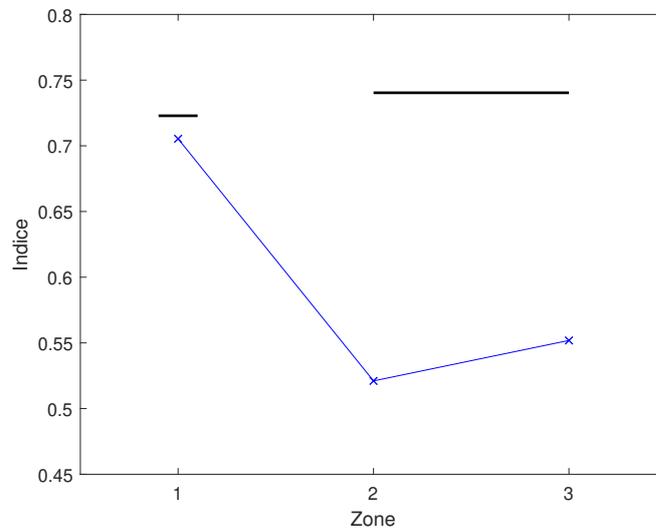


FIGURE 5.13 – Évolution moyenne de l'indice selon les zones pour la seconde partie de l'expérience. Les barres indiquent les groupes parmi lesquels aucune différence significative n'a été constatée.

Influence de la phase de l'expérience

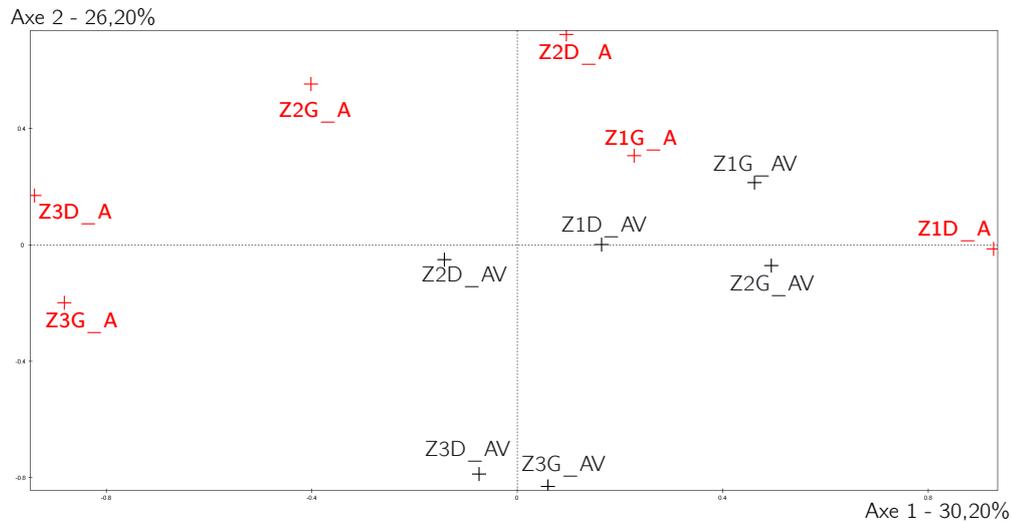
Dans cette section, nous allons nous intéresser aux changements qui ont lieu entre la condition audio-visuelle de la première partie de l'expérience et la seconde partie de l'expérience, où l'objet virtuel est seulement perceptible par le son. Pour rappel, la première partie illustre la prise en main du système lors de la première confrontation de l'utilisateur à l'objet virtuel, alors que la seconde représente la condition d'utilisation du système en conduite, au cours de laquelle nous ne voulons pas solliciter la vue de l'utilisateur.

Pour comparer ces deux situations, nous avons choisi de sélectionner les 54 derniers lancers de la condition audio-visuelle de l'expérience, et de les confronter aux 54 lancers de la condition auditive seule. Comme nous l'avons vu précédemment, il semble que les 54 derniers lancers – qui correspondent aux groupes 10, 9, 8, 7 et à la moitié du groupe 6 – se trouvent dans la phase de stabilisation sur la figure 5.10. Ces lancers peuvent donc représenter le maximum des performances atteintes lors de l'utilisation du système en voyant l'objet virtuel.

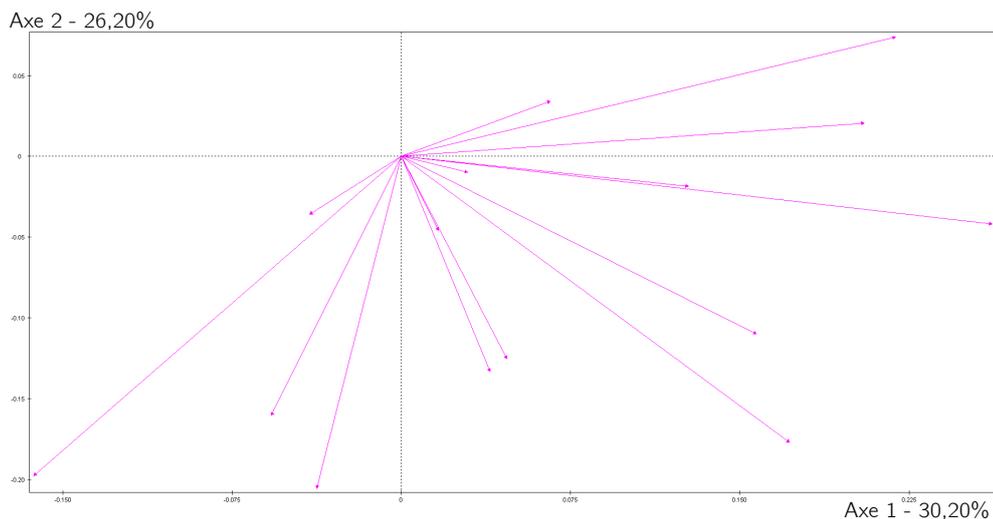
Au cours de cette analyse, nous allons définir un facteur *phase* de l'expérience, où la phase *AV* représente la sollicitation audio-visuelle, alors que la phase *A* correspond à la sollicitation auditive seule couplée à une réduction de la force de rappel du ressort.

ACP

Nous avons procédé à une analyse en composantes principales que nous avons fait sur l'indice par rapport aux différentes conditions de l'expérience, dont les résultats sont visibles en figures 5.14 et 5.15. Sur ces figures, nous avons conservés la notation *ZiG* et *ZiD*, en y ajoutant *AV* ou *A* pour indiquer la phase de l'expérience. Pour plus de lisibilité, nous avons indiqué la phase *A* en rouge et en gras sur les figures 5.14a et 5.15a.



(a) Évolution moyenne de l'indice par condition

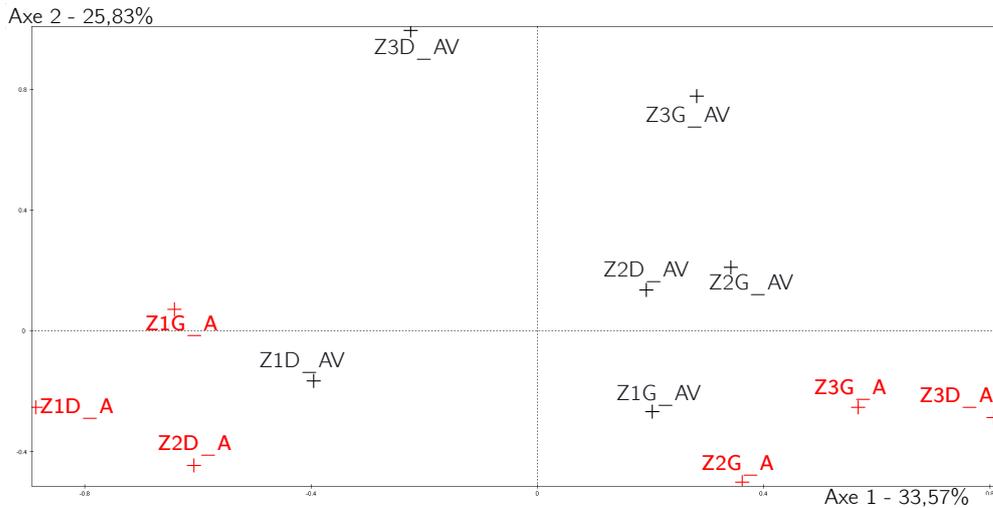


(b) Évolution moyenne de l'indice par sujet

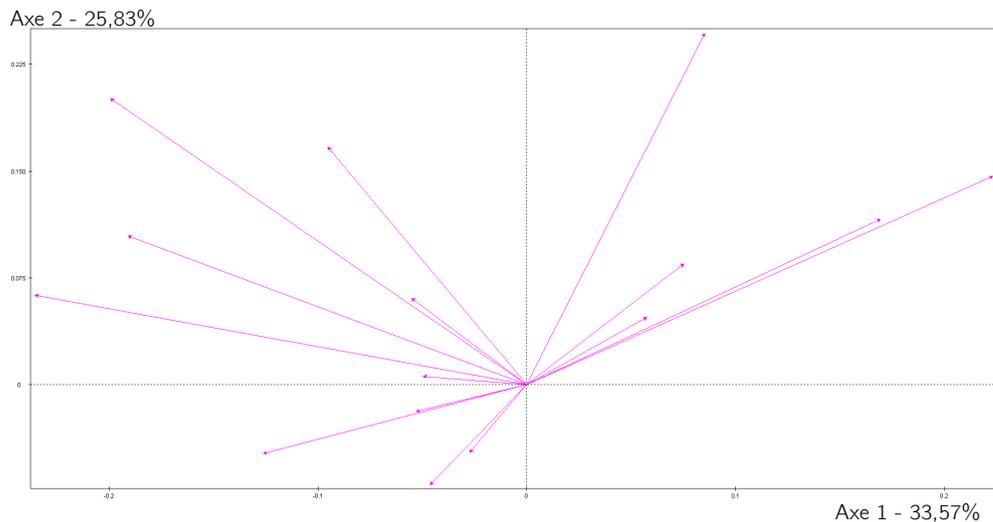
FIGURE 5.14 – ACP réalisée sur l'indice par rapport aux conditions audio-visuelles et auditives, pour la stratégie sonore liant la distance main-OV à la hauteur.

Tout d'abord, nous pouvons observer que les axes 1 représentent une variabilité plus faible des résultats par rapport aux situations précédentes, aux alentours de 30%. Les axes 2 sont eux aux alentours de 25% de variabilité.

En observant les figures 5.14a et 5.15a et en effectuant une comparaison pour une même zone et une même phase, nous pouvons noter que les points pour un lancer de l'OV à gauche ou à droite sont proches : le facteur de latéralité des lancers ne semble pas influencer grandement les résultats. De la même manière que précédemment, les points représentant chaque zone semblent être relativement éloignés les uns des autres en particulier pour la condition auditive, ce qui peut indiquer un effet significatif du facteur *zone*. Enfin, les points représentant les phases AV ou A semblent occuper des positions différentes : les points de la condition AV sont au centre des figures 5.14a et 5.15a, alors



(a) Évolution moyenne de l'indice par condition



(b) Évolution moyenne de l'indice par sujet

FIGURE 5.15 – ACP réalisée sur l'indice par rapport aux conditions audio-visuelles et auditives, pour la stratégie sonore liant la vitesse de l'OV au paramètre de roulement du synthétiseur.

que les points de la phase *A* sont plutôt positionnés sur le bord en haut de la figure 5.14a, et sur le bord en bas de la figure 5.15a. Nous pouvons alors supposer un effet significatif du facteur *phase*. Les figures 5.14b et 5.15b ne semblent pas montrer de caractère consensuel dans le comportement des sujets. Même si les flèches sont plutôt orientées vers le bas et sur la droite de la figure 5.14b, il n'y a pas de faisceau net indiquant des groupes de sujets ayant le même comportement. Dans ce cas, nous choisissons de faire une analyse de variance sur l'ensemble des sujets. Concernant la figure 5.15b, on observe une répartition similaire avec cette fois-ci 5 sujets qui semblent avoir des comportements proches. Cependant, il n'est pas possible d'obtenir des résultats exploitables d'une ANOVA sur 5 sujets, nous avons donc choisi de faire une analyse de variance sur l'ensemble du panel de sujets.

ANOVA

Nous avons donc réalisé une analyse de variance sur l'indice des données des deux phases de l'expérience, avec le facteur sujet imbriqué dans le facteur stratégie sonore – 28 ddl, la stratégie sonore – 1 ddl, la zone – 2 ddl, la latéralité – 1 ddl, la phase – 1 ddl, et les interactions entre ces facteurs. Les résultats indiquent un effet significatif du facteur *phase* – $F(1,3197) = 22,505$; $p < 0,001$. Les résultats sont meilleurs dans la phase *A* que dans la phase *AV*, avec des indices respectifs d'environ 0,59 et 0,68. Un effet significatif des stratégies sonores est observé – $F(1,3197) = 11,570$; $p = 0,001$ – tout comme un effet des zones – $F(2,3197) = 7,051$; $p = 0,001$. Une nouvelle fois, le facteur de latéralité ne donne pas de différences significatives dans les résultats – $F(1,3197) = 0,108$; $p = 0,743$. Aucune des interactions entre facteurs ne donne de résultats significatifs, mis à part pour l'interaction (zone \times phase) – $F(2,3197) = 22,881$; $p < 0,001$ – que nous avons représenté en figure 5.16.

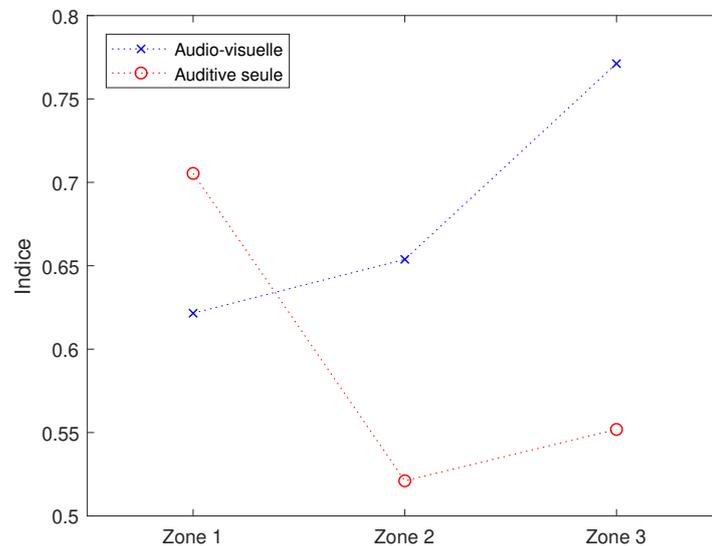


FIGURE 5.16 – Évolution moyenne de l'indice selon les zones et la phase de l'expérience.

5.3.4 Discussion

Effet des stratégies sonores dans la première partie de l'expérience

Dans cette première partie audio-visuelle de l'expérience, les sujets peuvent voir l'objet virtuel mais également l'entendre grâce à nos stratégies de sonification. Des retours d'informations sur la tâche sont également donnés : visuellement avec la modification de la température ; mais aussi par des icônes sonores indiquant l'accomplissement ou non de la tâche demandée.

Le but de cette phase de l'expérience est d'observer si une amélioration des lancers a lieu, et si cette éventuelle amélioration est liée à une stratégie sonore en particulier. Nous avons donc choisi 120 essais, pour observer l'évolution des résultats sur une période relativement longue.

Comme nous l'avons présenté dans la section des résultats, une amélioration des performances a bien lieu. À partir du 4^e groupe de lancers, les résultats se stabilisent : les sujets semblent avoir atteint les performances maximum et ne s'améliorent plus. Nous avons également vu que les résultats ne sont pas significativement meilleurs pour l'une ou l'autre des stratégies de sonification. Nous pouvons alors proposer deux explications. La première supposerait que les sujets ont utilisé les informations sonores des deux stratégies de sonification, qui sont équivalentes. Cette explication est plausible sachant que les stratégies 1 et 6 de l'expérience 2 dont sont issues les stratégies sonores que nous avons testées n'ont pas de différences significatives en phase 2 de la précédente expérience – voir figure 4.10b. Une seconde explication propose que les stratégies sonores n'ont pas été utilisées car les informations provenant de la modalité visuelle ont occulté les informations sonores. En nous référant au chapitre 2, cet exemple peut témoigner une fois de plus de la grande influence qu'ont les informations portées par la modalité visuelle sur celles provenant des autres sens. Plusieurs études vont dans ce sens, avec l'exemple de « l'effet Colavita » que nous avons abordé précédemment [Posner *et al.*, 1976] [Colavita, 1974] [Spence, 2009] [Koppen et Spence, 2007] [Koppen *et al.*, 2009]. L'étude de Pizzighello et Bressan indique pourtant qu'un phénomène inverse existe également, où les informations visuelles sont capturées par l'audition [Pizzighello et Bressan, 2008]. Cette étude impose cependant une tâche auditive aux sujets, qui adoptent alors une stratégie de concentration des ressources attentionnelles sur les stimuli sonores qui leur sont proposés. Notre étude, comme celles relatives à « l'effet Colavita », propose des stimuli audio-visuels sans donner de consigne explicite pour se concentrer sur une modalité.

La découverte de l'objet virtuel et des gestes à exécuter a probablement nécessité une attention importante des sujets, pour acquérir rapidement un grand nombre d'informations. Nous pensons donc que les sujets ont alors utilisé la modalité dominante du système perceptif humain, en ne tenant compte des autres que partiellement, voir pas du tout : nous pouvons alors être dans un cas de « cécité attentionnelle », ou plus précisément de « surdité attentionnelle ».

Effet des stratégies sonores dans la seconde partie de l'expérience

Dans cette seconde partie de l'expérience, les sujets ne peuvent plus voir l'objet virtuel mais ils le perçoivent seulement grâce à nos stratégies de sonification. Les retours visuels et sonores de l'interface – l'affichage de la température et les icônes sonores – sont en revanche toujours disponibles. Cette phase illustre une potentielle utilisation du système en véhicule, où nous ne voulons pas donner de retours visuels de l'objet virtuel, pour que les utilisateurs concentrent leur attention visuelle sur la tâche de conduite. Ils ont théoriquement pu acquérir un schéma cognitif relatif à l'objet virtuel, grâce aux stimuli audio-visuels de la première phase. Plusieurs paramètres nous intéressent tout particulièrement dans cette partie de l'étude : le facteur de stratégie sonore en premier lieu, mais également les facteurs de zone ou de latéralité, pour savoir s'il est plus facile de faire un balayage de la main sur la gauche ou sur la droite.

Pour cette phase de l'expérience, les stratégies sonores sont significativement différentes, avec la stratégie de sonification de la distance qui obtient un indice moyen plus faible que la seconde stratégie sonore. Nous pouvons déjà affirmer dans ce cas que les sujets ont utilisé les informations transmises par nos stratégies de sonification.

Les résultats concernant les stratégies sonores semblent prendre le contre-pied de nos conclusions de l'expérience 2. En effet dans cette précédente expérience, la stratégie 1 dont est issue la sonification de la distance est arrivée en dernière position en phase 1, et en seconde en phase 2 ; alors que la stratégie 6 est arrivée en seconde et première position pour les phases 1 et 2 respectivement. Une première hypothèse qui peut expliquer ces résultats provient du fait que les performances associées aux stratégies de sonification peuvent varier dans le temps, comme nous l'avons illustré au travers de l'étude de Rath et Schleicher [Rath et Schleicher, 2008]. La sonification de la distance a donc pu être apprise dans la première partie de l'expérience, même si les différences avec l'autre stratégie sonore n'étaient pas significatives, pour nettement supplanter l'autre stratégie de sonification dans la deuxième partie. Il faut également rappeler que cette stratégie de sonification est adaptée à la tâche que nous avons proposée aux sujets, puisque les balayages de la main sont jugés sur la distance à laquelle est envoyé l'objet virtuel. Cependant, le changement important de modalité entre les deux phases de l'expérience ne permet pas de se satisfaire de cette analyse.

Une seconde explication, faisant suite à la première de la discussion concernant la première phase, suppose que les différences proviennent de l'intégration multisensorielle qui a pu avoir lieu dans la première partie de l'expérience. C'est d'ailleurs principalement sur les différentes modalités sollicitées que se différencient les expériences 2 et 4 : la première propose une condition sans visuel, au contraire de la seconde. Il est possible que les sujets aient – consciemment ou inconsciemment – associés plus facilement le visuel de l'objet virtuel à la stratégie de sonification faisant varier la hauteur qu'à la sonification de la vitesse, même si ces différences n'ont pas été mesurées. Lorsqu'ils ont été privés d'informations visuelles, les volontaires de l'expérience ont alors pu plus s'appuyer sur la stratégie sonore qu'ils avaient le mieux associé au comportement de l'objet virtuel.

Une troisième hypothèse, qui découle de la seconde de la précédente discussion, propose que les sujets n'ont pas utilisé les informations auditives dans la première partie de l'expérience. Dans cette seconde phase auditive, ils ont alors dû s'appuyer sur la stratégie de sonification la plus adaptée à la tâche que nous avons proposée. La sonification de la distance peut alors être plus efficace que la seconde. Nous pensons que cette hypothèse est la plus probable, et ce raisonnement s'intègre à la réflexion que nous avons développée autour de l'influence des informations visuelles sur celles provenant des sons. Selon cette supposition, les sujets n'ont pas pu créer de schéma cognitif reliant les informations sonores avec la dynamique de l'objet virtuel, ce qui était le but de cette expérience.

Un enseignement intéressant à tirer des résultats est que l'orientation des balayages de la main sur l'axe horizontal ne semble pas influencer les performances de la tâche que nous avons mis au point. Cet effet peut sembler surprenant puisque l'on peut supposer intuitivement que, pour des sujets droitiers, envoyer l'objet virtuel vers la gauche avec la paume de la main est plus facile que l'envoyer sur la droite avec le revers de la main. Ce résultat est intéressant pour la mise au point de l'interface gestuelle que nous allons mettre en place, puisqu'il permet de se dispenser d'une compensation d'un effet latéral.

Enfin les résultats montrent un effet du facteur *zone*, avec un indice significativement plus important pour la zone 1 que pour les zones 2 et 3. La zone 1 peut être trop proche pour pouvoir facilement l'atteindre avec les paramètres du ressort que nous avons choisis, et nous pouvons une nouvelle fois constater que la loi de Fitts ne s'applique pas à notre système.

Influence de la phase de l'expérience

Dans cette partie, nous avons voulu comparer les résultats des deux parties de l'expérience. Pour ce faire, nous avons sélectionné les 54 derniers lancers de la phase audio-visuelle, et nous les avons confrontés aux 54 lancers de la phase où l'objet virtuel est perceptible uniquement grâce aux stratégies de sonification.

Le but est ici de déterminer les causes des évolutions qui ont pu s'opérer entre les deux phases de l'expérience. Pour analyser convenablement les résultats, il est important de prendre en compte le changement de raideur entre les deux phases de l'expérience, que nous avons introduit pour tester si les sujets se repèrent grâce aux stratégies de sonification ; mais aussi le changement des modalités sollicitées.

Le premier résultat surprenant est que la condition audio-visuelle obtient de moins bons résultats que la condition avec son seul. Nous avons intuitivement pensé que l'apport conjoint d'informations visuelles et auditives entraînerait de meilleures manipulations de l'OV. Plusieurs explications peuvent être proposées pour discuter de ce phénomène. Tout d'abord, la figure 5.16 nous indique un fort changement de l'indice moyen pour chaque zone, ce qui semble provenir de la diminution de la force de rappel : pour un même geste, une diminution de la force permet à l'OV de partir plus loin, et donc d'atteindre une zone plus lointaine. La figure témoigne de ce type de phénomène, où les zones 2 et 3 sont plus facilement atteintes en phase auditive seule. À l'inverse la zone 1 est plus difficile à atteindre, le geste à effectuer pour envoyer l'OV dans cette zone peut être d'une amplitude trop faible pour être bien maîtrisé. Cette hypothèse ne tient cependant pas compte des deux stratégies de sonification.

Pour explorer l'influence des stratégies sonores dans les résultats, il peut être pertinent de s'intéresser à l'interaction (stratégie sonore \times phase de l'expérience). Comme nous l'avons indiqué précédemment, il semble qu'il n'y ait pas de différences significatives selon le critère strict d'une significativité en-dessous de 0,05 – $F(1,3197) = 3,505$; $p = 0,061$. Nous allons cependant observer l'évolution des résultats selon cette interaction, puisque la p-value est proche de la limite communément admise. Nous pourrions alors probablement mieux comprendre les comportements des sujets lors du changement de phase. Nous avons donc représenté en figure 5.17 les résultats de l'expérience en détaillant par phase, et par stratégie sonore.

Sur cette figure, on peut tout d'abord constater que les deux stratégies ont des indices proches dans la phase AV, ce qui confirme nos précédentes observations. Nous pensons que les sujets se sont principalement basés sur les informations visuelles dans cette phase de l'expérience, et ne se sont que peu servi des stratégies de sonification. D'une part, nous pouvons observer une diminution globale de l'indice en phase auditive seule. Même si le changement de la force de rappel du ressort semble expliquer ces résultats, nous pouvons également supposer que les informations auditives, ignorées en phase audio-visuelle, peuvent être plus informatives que les informations visuelles de l'OV pour la tâche que nous avons proposé. Une phase proposant seulement des informations visuelles sur l'OV aurait pu permettre de trancher cette question. D'autre part, l'indice baisse nettement plus pour la stratégie de sonification de la distance que pour l'autre stratégie sonore. L'analyse de la phase auditive révèle d'ailleurs des différences significatives entre les deux stratégies de sonification. Il semble donc que le facteur *stratégie sonore* a joué un rôle important dans la diminution de l'indice entre les deux phases. Les informations transmises par la stratégie liant la distance main-OV au paramètre de hauteur se sont révélées être les plus adaptées à la tâche demandée.

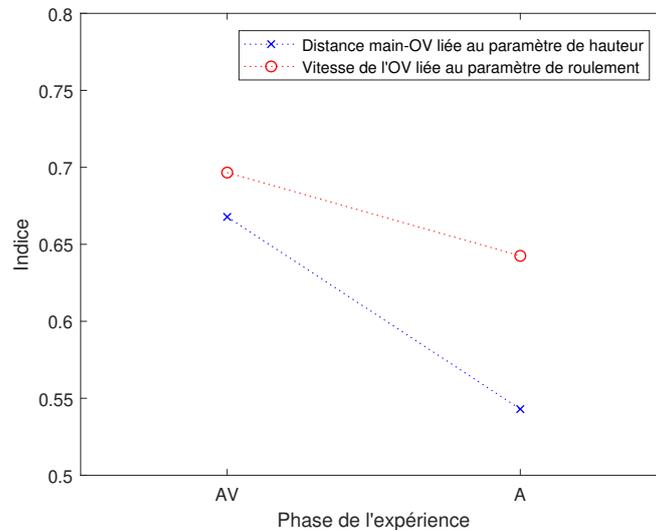


FIGURE 5.17 – Évolution de l'indice selon les stratégies sonores et la phase de l'expérience.

Il est alors possible de résumer les discussions précédentes sous la forme de deux hypothèses. La première suppose que les sujets ont intégré les stimuli audio-visuels concernant l'OV pour créer un schéma cognitif. Les informations audio-visuelles semblent équivalentes puisque mènent à des résultats proches suivant les deux stratégies sonores. Par la suite, en phase *A* le schéma cognitif associé à la stratégie liant la distance main-OV au paramètre sonore de hauteur a permis de mieux accomplir la tâche, ce qui témoigne de l'efficacité de cette stratégie de sonification.

Notre seconde explication, qui nous semble plus convaincante, propose que l'absence de différences significatives en phase *AV* provient de la prédominance de la modalité visuelle sur la modalité auditive : les sujets découvrant un objet virtuel inconnu ont pu concentrer leur attention sur les informations visuelles, phénomène que nous avons déjà observé précédemment [Posner *et al.*, 1976] [Colavita, 1974] [Spence, 2009] [Koppen et Spence, 2007] [Koppen *et al.*, 2009]. Les informations sonores ont alors été probablement ignorées. Dans la seconde phase de l'expérience, les meilleurs résultats sont en partie dus au changement de la force de rappel, ce qui a permis d'atteindre plus facilement les zones 2 et 3, comme le montre la figure 5.16. La stratégie de sonification de la distance s'est par ailleurs révélée plus efficace dans cette phase auditive seule, probablement parce qu'elle renseignait particulièrement sur la tâche à effectuer. Nous pouvons cependant douter de la construction d'un schéma cognitif relatif à l'objet virtuel – but initial de l'expérience – à cause de l'utilisation privilégiée des informations visuelles dans la phase audio-visuelle. Nous devons donc tenir compte de cet enseignement dans la suite de notre étude.

Le véritable comportement des sujets se situe probablement entre ces deux hypothèses. Il est également possible que les informations visuelles et auditives provenant de l'interface aient été utilisées pour moduler les gestes de balayage, notamment pendant la seconde phase de l'expérience où le visuel de l'OV n'est plus disponible. Nous observons cependant des différences significatives entre les deux stratégies sonores dans la phase *A* de l'expérience, ce qui montre bien que ces informations provenant de l'interface de température n'ont pas été les seules à guider les sujets.

5.3.5 Conclusion

Dans cette section, nous nous sommes donc interrogés sur la méthode de présentation de l'objet virtuel à des sujets novices. En particulier, la constitution d'un schéma cognitif relatif aux manipulations de l'objet virtuel nous a particulièrement intéressée. Nous avons donc choisi de donner des informations visuelles et sonores sur l'objet virtuel dans une première partie de l'expérience, pour qu'il soit possible de créer le schéma cognitif le plus riche possible. Nous avons ensuite supprimé les informations visuelles, ne permettant de percevoir l'OV qu'à travers nos deux stratégies de sonification.

Lors de la première phase, une amélioration des résultats a lieu pendant les 36 premiers lancers, les performances stagnant ensuite. Il n'a pas été possible de distinguer les deux stratégies de sonification, qui mènent à des résultats similaires. Nous supposons que les sujets ont concentré leur attention sur les stimuli visuels de l'objet virtuel, et n'ont pas tenu compte des sons représentant la dynamique de l'OV.

Les deux stratégies sonores se distinguent en revanche dans la seconde partie de l'expérience, la sonification de la distance entraînant des résultats significativement meilleurs que la seconde stratégie sonore. Cette stratégie semble plus adaptée à la tâche demandée, ce qui lui a potentiellement permis d'obtenir de meilleurs résultats.

Nous avons ensuite comparé les derniers lancers de la première phase avec les lancers de la seconde phase. Les résultats montrent de meilleures performances dans la seconde phase de l'expérience, où seulement les stimuli auditifs de l'OV sont accessibles. Ces résultats semblent provenir en partie du changement de force de rappel du ressort que nous avons fait entre les deux phases, qui a globalement facilité les lancers. D'autre part, nous pensons que les informations auditives portées par les stratégies de sonification n'ont été utilisées que dans la seconde phase de l'expérience. Ces informations qui semblent utiles, pourraient avoir été ignorées lors de la phase audio-visuelle, et n'ont pas permis de construire un schéma cognitif relatif aux manipulations de l'OV. Elles n'auraient été utilisées que pour la seconde phase, ce qui peut également expliquer l'amélioration des résultats entre les deux phases.

Pour tenter de mieux comprendre le phénomène, il pourrait être intéressant de proposer l'expérience précédente à un groupe de sujets additionnel, qui auraient accès aux mêmes stimuli visuels, mais dont la stratégie de sonification ne serait pas informative. Si les résultats en phase AV sont similaires aux résultats présentés précédemment, alors il est possible que les sujets de notre expérience n'aient pas tenu compte des informations portées par les sons. Il aurait également été possible d'observer si l'amélioration des résultats en seconde phase est due à la modification de la force de rappel, ou à l'utilisation des informations portées par les stratégies de sonification.

Les résultats de cette expérience permettent de tirer plusieurs enseignements. Il semble que le côté droit ou gauche d'un lancer de l'OV n'importe pas, ce qui permet de traiter les deux balayages de manière symétrique. Nous avons également confirmation que certaines zones sont plus difficilement atteintes que d'autres. Enfin, la principale conclusion de cette expérience provient de la « surdit  attentionnelle » que nous pensons avoir observ  : il semble que la cr ation de sch emas cognitifs associant les stimuli visuels et auditifs de l'OV n'aient pas eu lieu. Nous devons tenir compte du mode de pr sentation de l'OV lors de la prochaine exp rience, pour assurer une prise en main optimale de l'interface.

Conclusion du chapitre

Nous avons débuté ce chapitre par des précisions concernant notre interface gestuelle, en choisissant des cas d'usages s'inscrivant dans un contexte automobile. Ces cadres d'application vont nous permettre de structurer le développement de notre IHM, en la construisant autour de cas concrets. Nous avons donc ciblé une interface de climatisation, et une interface de gestion de la musique.

Par la suite une étape de design sonore a été proposée, pour construire des stratégies de sonification agréables à écouter, dans le but qu'elles soient bien acceptées dans une utilisation quotidienne. Ce travail sur les textures sonores ne devait cependant pas empêcher les stratégies sonores de transmettre les informations que nous avons choisies suite à l'expérience 2.

Une troisième expérience perceptive nous a permis d'étudier les évocations sonores provoquées avec différents types de sonification chez des sujets naïfs. La sonification liant la vitesse de l'OV au synthétiseur de roulement semble avoir particulièrement guidé les sujets dans la compréhension des mécanismes en jeu, certaines personnes ayant même décrit intégralement le concept de l'objet virtuel et du ressort.

Enfin, une quatrième expérience s'est focalisée sur l'utilisation de l'objet virtuel appliquée à un réglage de température. Nous avons présenté l'OV avec des stimuli audio-visuels dans un premier temps, pour représenter une phase de familiarisation avec l'OV pouvant se faire à l'arrêt. Par la suite, le visuel de l'OV a été masqué pour simuler une utilisation en conduite. Les résultats semblent montrer qu'un phénomène de « surdité attentionnelle » a eu lieu, les sujets n'ayant pas tenu compte des stratégies de sonification lorsqu'il était possible de voir l'OV. Ils n'ont donc pas pu développer les schémas cognitifs utiles à la manipulation aveugle de l'OV.

Dans la suite de notre étude, nous allons développer une interface gestuelle complète, et l'évaluer en situation réelle de conduite. Pour éviter un trop grand nombre de paramètres dans la prochaine expérience, nous allons choisir une stratégie sonore parmi les deux que nous avons testées au cours des expériences 3 et 4. L'expérience 3 semble plébisciter la sonification de la vitesse, alors que l'expérience 4 favorise la sonification de la distance.

Notre choix s'est porté sur la sonification de la vitesse de l'OV grâce au synthétiseur de roulement, qui correspond le plus à la métaphore que nous voulons mettre en place avec l'OV. Nous pensons que les évocations peuvent permettre une meilleure construction de schémas cognitifs lors des premières manipulations de l'OV. De plus les sujets ont semblé ignorer les stratégies sonores dans la phase audio-visuelle de l'expérience 4 : les résultats de la deuxième phase ne reflètent donc pas les conditions que nous voulions mettre à l'épreuve – la création d'un schéma cognitif audio-visuel puis son utilisation. Enfin, les réponses au questionnaire portant sur le caractère agréable des deux sonifications nous orientent également vers ce choix.

L'application des résultats dans un véhicule automobile va maintenant être au centre de nos préoccupations. Nous allons donc devoir développer l'interface gestuelle complète, avant de l'intégrer à un simulateur de conduite. Enfin, nous testerons différentes hypothèses dans une dernière expérience perceptive.

Utilisation de l'interface gestuelle en situation de conduite

Dans les chapitres précédents, plusieurs expériences ont été menées pour définir la stratégie sonore transmettant les informations liées à la dynamique de l'objet virtuel. À la fin du chapitre précédent, nous avons fait le choix de privilégier la stratégie mêlant un son de roulement à une texture sonore, notamment par la cohérence qu'elle apporte vis-à-vis de la métaphore que nous voulons mettre au point. Cette stratégie de sonification doit permettre aux utilisateurs de se représenter les mouvements de l'objet virtuel en l'absence de stimuli visuels. Nous nous sommes également intéressés aux évocations pouvant être véhiculées par le couplage gestes - sons, qui peuvent être utiles pour les premières utilisations du système. La dernière expérience a permis d'observer l'intégration multisensorielle de stimuli audio-visuels, ce qui nous renseigne sur la méthode de présentation de ces signaux lorsque nous désirerons présenter notre interface gestuelle à des conducteurs.

Dans ce dernier chapitre, nous désirons évaluer l'apport de manipulations gestuelles de l'IHM en situation de conduite. Ce contexte particulier demande une attention visuelle soutenue, et l'exécution de tâches simultanées peut entraîner des perturbations importantes sur la conduite, comme nous l'avons vu dans le chapitre 3. Nous pensons cependant que la répartition des informations selon différentes modalités peut permettre à des distractions cognitives d'entraîner un comportement de conduite plus sécuritaire qu'avec des distractions visuelles.

Nous avons donc mis au point une dernière expérience, dont l'objectif est d'évaluer l'apport potentiel de cette exploitation multimodale pour la réalisation de plusieurs tâches simultanément, en mettant en pratique les connaissances que nous avons accumulées dans les précédentes expériences. Plusieurs interrogations se posent alors à nous : est-il possible de créer une interface manipulable facilement par des gestes en aveugle ? Son utilisation est-elle envisageable en conduite ? La combinaison de gestes et de sons ne provoque-t-elle pas une surcharge cognitive chez les conducteurs, pouvant entraîner des situations dangereuses sur la route ? Pour tenter de répondre à ces questions, il est tout d'abord nécessaire de mettre au point une IHM gestuelle, faisant des liens logiques entre gestes, fonctions automobiles et sons. Les cas d'usages exposés précédemment ont été repris et complétés, pour former des interfaces cohérentes.

L'expérience perceptive dans laquelle notre interface gestuelle est testée se déroule en simulateur de conduite, nous avons donc intégré notre IHM à cet environnement particulier. Nous y avons apporté un certain nombre de modifications, pour que les manipulations du simulateur et de l'interface soient faciles et ne se gênent pas.

Enfin, nous abordons plus précisément les hypothèses de l'étude et le protocole de l'expérience. Pour évaluer la maniabilité de notre interface gestuelle en conduite, nous avons choisi de la confronter à une interface tactile, technologie dominante pour la manipulation d'interfaces sur les véhicules récents. Les résultats qui sont présentés permettent de tirer plusieurs enseignements, et ouvrent un certain nombre de pistes d'améliorations possibles et de perspectives d'études complémentaires.

Sommaire

6.1	Développement d'une interface gestuelle complète	153
6.1.1	Fonctions de ventilation	153
6.1.2	Fonctions de gestion des médias	154
6.1.3	Choix des gestes et manipulations de l'objet virtuel associées	155
6.1.4	Stimuli audio-visuels et construction des interfaces	158
6.2	Intégration du système et simulateur de conduite	162
6.2.1	Simulateur de conduite	162
6.2.2	Verrouillage de l'interface	164
6.3	Expérience 5 : comparaison d'interfaces en situation de conduite	165
6.3.1	Hypothèses initiales et motivations	165
6.3.2	Méthodologie	166
6.3.3	Résultats	176
6.3.4	Discussion	190
6.3.5	Conclusion	195

Les travaux présentés dans ce chapitre font l'objet d'un article en cours de publication, et d'un brevet en cours de dépôt.

6.1 Développement d'une interface gestuelle complète

Dans les travaux que nous avons présentés jusqu'à présent, nous nous sommes focalisés sur la stratégie de sonification la plus adaptée pour transmettre des informations sur la dynamique de l'objet virtuel. Les balayages de la main ont alors paru particulièrement adaptés, et nous ont permis de faire le choix d'une stratégie de sonification pour le reste de nos travaux.

Il est à présent nécessaire de bâtir une interface gestuelle plus complète, permettant de manipuler plusieurs fonctions automobiles à l'image d'une interface tactile. Nous pourrions alors confronter ces deux types d'interfaces, pour juger des avantages et inconvénients de chacune d'elles. La métaphore de manipulations d'un objet virtuel que nous avons mis au point jusqu'ici doit s'accorder avec ces nouvelles interactions, pour que l'OV joue pleinement son rôle de relais entre interactions gestuelles et fonctions de l'IHM. De plus, la réflexion doit également inclure le retour sonore découlant des manipulations, pour que l'ensemble {gestes – manipulations de l'OV – stratégies de sonification – fonctions automobiles} fassent sens [Susini *et al.*, 2012]. Nous pensons en effet que les schémas cognitifs liés à l'objet virtuel et à l'interface seront facilement intégrés si les interactions de l'ensemble des composantes de l'interface sont pensées en amont, avec des relations logiques entre elles.

Nos travaux sont basés sur les principes d'ergonomie des interfaces mis au point par Bastien et Scapin [Bastien et Scapin, 1993]. Nous allons, dans la mesure du possible, respecter ces critères qui concernent par exemple le guidage, la charge de travail des utilisateurs, l'homogénéité ou encore la gestion des erreurs. Nous reviendrons sur ces notions dans la suite du document.

Dans notre démarche de mise au point de l'interface gestuelle, nous nous sommes tout d'abord intéressés aux fonctions que nous voulons modifier, avant de réfléchir aux gestes et manipulations de l'objet virtuel qui y sont associés. Nous allons suivre ce raisonnement dans cette section, en exposant d'abord les fonctions des interfaces que nous avons privilégiées, avant de les relier à l'objet virtuel et à des gestes. Les cas d'usages que nous avons développés en section 5.1.1 ont servi de base de réflexion au développement des interfaces que nous allons présenter.

Il est à noter ici que nous ne visons pas la création d'une interface seulement pilotée par les gestes pour naviguer, valider, revenir en arrière dans les pages de l'interface ou encore manipuler toutes les fonctions. Pour l'instant, notre objectif est de construire une page d'interface simplifiée, où seules certaines fonctions sont accessibles. Si nos essais sont concluants, nous pourrions éventuellement compléter ces interfaces.

6.1.1 Fonctions de ventilation

Dans cette section, nous allons lister les fonctions liées à la ventilation, ainsi que les retours sensoriels qu'il est possible d'avoir lors de leur manipulation autre qu'un retour visuel grâce à l'écran central. En effet, l'ambition de notre interface gestuelle est de donner des retours d'informations par d'autres canaux sensoriels que le canal visuel, qui doit être concentré sur la route.

Nous avons choisi de conserver les modifications de température, qui nous semblent être une fonction pertinente. Elle est fréquemment utilisée en véhicule, et le retour d'information – l'air chaud ou froid arrivant sur le conducteur – n'est pas instantané : un certain temps est nécessaire au système pour arriver à la température cible. Un retour d'information par le biais du son peut alors être intéressant, pour confirmer la prise en compte d'un réglage demandé.

La deuxième fonction que nous avons ciblée est elle aussi fréquemment utilisée conjointement au réglage de la température : l'intensité de ventilation. Cette fonction semble incontournable même dans une interface de ventilation simplifiée. La pression du flux d'air sur la peau des conducteurs peut permettre d'avoir un retour d'information quant au réglage d'intensité, mais ce ressenti tactile est dépendant des habits portés par l'utilisateur, de l'intensité de ventilation et de l'orientation du flux d'air. Un second retour d'information est également présent, avec le bruit émis par la ventilation qui est naturellement modulé par le réglage d'intensité de ventilation. Le retour d'information que nous devons apporter lors des manipulations de cette fonction doit donc être discret, pour ne pas être redondant avec les informations tactiles et sonores.

Enfin, le positionnement de la source de ventilation peut également être une fonction intéressante à piloter grâce à l'interface gestuelle. De la même manière que la fonction précédente, un retour tactile est disponible lorsque cette fonction est modifiée, mais il n'est disponible que si le flux d'air est suffisamment important. L'ajout d'un retour sonore peut alors être une solution pour assurer un retour sensoriel lors de la manipulation de cette fonction.

Nous avons également voulu proposer un deuxième type d'interface, basé sur le second cas d'usage que nous avons évoqué dans le chapitre précédent. Cette interface permet une gestion de la musique, et utilise directement la matière sonore écoutée par les utilisateurs pour donner un retour d'information.

6.1.2 Fonctions de gestion des médias

Cette interface de manipulation de paramètres liés à la musique peut permettre de créer une relation plus directe entre les gestes des utilisateurs et les retours sonores. Avec ce type d'IHM, il n'est pas nécessaire d'ajouter un son pour transmettre de l'information, puisque les modifications des paramètres sonores assurent déjà ce rôle.

Par exemple, les changements d'intensité sonore d'un morceau de musique étant directement perçus, il ne semble pas nécessaire de rajouter un retour d'information par une stratégie de sonification ou des icônes sonores. Cette fonction est par ailleurs incontournable pour une interface automobile de médias : c'est l'une des seules fonctions dont le bouton rotatif est toujours présent, comme dans la 308 de seconde génération présentée en figure 1.2 par exemple. Nous avons donc sélectionné cette fonction pour faire partie de notre interface gestuelle.

Une seconde fonction que nous avons choisie concerne le placement de la source sonore dans l'habitacle. Cette fonction est déjà présente dans les interfaces tactiles – comme celles de la 308 par exemple – mais changer ses paramètres se révèle compliqué : il faut naviguer dans plusieurs couches d'interfaces successives, pour ensuite positionner la source dans le véhicule. Ces manipulations peuvent demander une grande concentration si elles sont accomplies simultanément à une tâche de conduite. Nous pensons qu'une manipulation plus directe et plus naturelle peut être proposée grâce à des gestes. De nouveau, le placement de la source sonore dans l'habitacle étant directement perceptible par les utilisateurs, il ne semble pas indispensable de rajouter des informations sonores.

Enfin, le défilement de titres musicaux est également une fonction fréquemment utilisée lors de l'écoute de musique, et semble une fonction importante à intégrer à notre interface. Les retours d'informations sont une nouvelle fois directement assurés par la modification de la lecture de la musique, par exemple avec l'interruption du morceau lorsque l'utilisateur effectue la commande de lancement du morceau suivant.

Nous avons pour l'instant listé des fonctions fréquemment utilisées sur des interfaces automobiles et donc indispensables à une future interface gestuelle, ou paraissant être particulièrement adaptées pour être manipulées par des gestes. Comme précédemment évoqué, nous allons progresser dans la construction de notre interface par la structuration de ces fonctions autour de gestes que nous allons définir. Il est néanmoins important de garder à l'esprit que les deux cas d'usages que nous avons définis doivent conduire à des interfaces semblables, pour respecter le critère d'homogénéité de Bastien et Scapin [Bastien et Scapin, 1993].

6.1.3 Choix des gestes et manipulations de l'objet virtuel associées

Dans les paragraphes précédents, nous avons choisi trois fonctions pour chacune des deux interfaces. Nous allons les apparier pour les regrouper par famille de fonctions. De cette manière, un seul geste peut être choisi par famille de fonctions et si le choix des regroupements est correct, nous pensons que les sujets pourront créer un schéma cognitif liant geste et famille de fonctions de manière aisée. En passant par exemple de l'interface de ventilation à celle de gestion de la musique, il ne sera pas nécessaire d'apprendre de nouveaux gestes. Nous avons vu au chapitre 3 qu'il est possible de retenir sept éléments simultanément en mémoire de travail [Miller, 1956]. Nous pensons qu'en mettant au point des interfaces similaires, le dictionnaire gestuel à retenir pour les deux interfaces peut être réduit, ce qui peut faciliter l'adaptation des sujets et baisser leur concentration lors des manipulations de l'IHM.

Nous avons donc regroupé les fonctions selon trois familles de fonctions différentes :

- les fonctions d'intensité : la fonction pilotant l'intensité de ventilation ainsi que le volume de la musique, qui est une intensité sonore ;
- les fonctions de placement : le placement du flux d'air de la ventilation, ou le placement de la source de la musique dans l'habitacle s'apparentent toutes les deux au placement dans l'espace d'une source ;
- les fonctions de défilement : les changements de température ou de la piste musicale lue sont en fait des déplacements sur un axe en une dimension, dans un sens ou dans l'autre.

Il est maintenant nécessaire d'associer ces groupes de fonctions à des gestes, mais également à des manipulations de l'objet virtuel pour que la métaphore que nous souhaitons mettre en place soit complète. Trois gestes ont paru intéressants : le balayage de la main que nous avons étudié jusqu'ici, l'ouverture de la main grâce aux cinq doigts, et le déplacement de la main poing fermé. Nous allons maintenant détailler les manipulations de l'OV qui peuvent y être attachées et la famille de fonctions liée à chaque manipulation.

Le balayage de la main est associé aux fonctions de défilement, comme nous l'avons déjà fait pour l'expérience 2. Ce geste permet d'envoyer l'objet virtuel sur la droite ou sur la gauche, avant qu'il ne revienne sous l'effet de la force de rappel du ressort. À la suite des résultats des dernières expériences impliquant les balayages de la main, nous avons choisi de permettre deux réglages différents correspondant à deux zones successives d'envoi de l'objet virtuel. Les performances de manipulations pour trois zones étaient aux alentours de 50% de réussite, ce qui n'est pas suffisant pour une application en véhicule. Pour l'interface de ventilation, les deux zones peuvent correspondre à deux changements différents de température. En ce qui concerne l'interface de gestion de la musique, nous avons choisi d'ajouter la possibilité d'avancer ou de reculer de 10 secondes dans la piste lue. En reprenant les catégories décrites en section 1.1, ce geste peut être classifié comme pantomimique.

Nous avons choisi de relier les fonctions d'intensité au geste d'ouverture de la main, qui peut être relié à un écrasement de l'objet virtuel. Nous pouvons imaginer que les utilisateurs, en fermant leur main, peuvent moduler le rayon de l'OV. Ainsi, une main se refermant écrase l'objet virtuel et diminue son rayon, alors qu'une main s'ouvrant complètement permet à l'objet virtuel de retrouver sa taille d'origine. Il est ensuite possible de figer le rayon de l'objet virtuel entre ces deux extrêmes, ce qui permet de régler finement l'intensité. Ce geste a été imaginé grâce aux manipulations quotidiennes d'objets : il est possible d'écraser des objets avec la main, une balle en mousse par exemple, ce qui en modifie le rayon. Nous avons voulu retranscrire cette affordance qu'offrent certains objets par ce geste pantomimique. Nous tirons en revanche profit de la réalité virtuelle ici pour permettre le retour de l'OV à sa taille d'origine, ainsi qu'une infinité de réglages du rayon de l'objet virtuel grâce à l'ouverture de la main. En imaginant l'objet virtuel comme étant la source de la ventilation ou de la musique, le geste de fermeture de la main évoque également une action d'étouffement de l'OV. C'est pour cette raison que nous avons choisi de lier ces gestes aux fonctions d'intensité. Enfin, nous avons lié les fonctions de placement à un déplacement de l'objet virtuel, en mimant le déplacement d'un véritable objet : si l'on imagine l'objet virtuel comme un petit haut parleur, alors il est possible de le déplacer en le saisissant, en le transportant vers la position souhaitée, et en le relâchant. Il est alors possible d'imaginer l'objet virtuel comme la source de la ventilation ou la source sonore, que l'utilisateur peut placer dans l'habitacle.

Il est intéressant de noter que les gestes pantomimiques se situent à mi-chemin de la classification des gestes selon l'information qu'ils contiennent, que nous avons évoqué en section 1.1. Ils contiennent donc suffisamment d'informations sans pour autant être trop complexes à retenir, comme le sont les gestes symboliques du langage des signes par exemple.

Nous avons résumé les différents gestes et les manipulations de l'objet virtuel qu'il est possible de faire à partir d'un état initial sur la figure 6.1. Sur ce schéma, nous avons utilisé des termes de *geste d'entrée* et *geste de sortie* qui correspondent au geste qu'il faut réaliser pour débiter ou terminer l'interaction avec une fonction. Sans ce principe, l'ouverture de la main serait par exemple constamment reliée à l'intensité, ce qui rendrait l'utilisation de l'interface difficile. Ces gestes permettent donc d'interagir avec l'objet virtuel tout en s'inscrivant dans la métaphore que nous voulons créer. Le fait de tourner la main paume vers le haut fait venir l'objet au creux de la main des utilisateurs, qui peuvent ensuite ouvrir plus ou moins leur main. Lorsque la bonne intensité a été déterminée, l'utilisateur peut simplement faire retomber l'objet de sa main en la retournant, ce qui permet de revenir à l'état initial. Dans chaque bloc de la figure 6.1, le geste à réaliser est précisé, suivi de sa conséquence sur l'objet virtuel. Ces gestes d'entrée et de sortie sont de plus intégrés à la métaphore avec le geste de déplacement : pour déplacer une source, il faut fermer la main pour se saisir de l'OV, se déplacer, pour rouvrir la main pour le relâcher. Il est à noter qu'il n'existe pas ce type de geste pour le balayage de la main, qui consiste à pousser l'objet virtuel sur l'axe horizontal.

Nous pouvons alors nous interroger sur le nombre de gestes qu'il est nécessaire de retenir pour les sujets. Nous pouvons décompter au minimum 10 gestes : 4 gestes de défilement, 3 gestes pour modifier la fonction d'intensité, et 3 gestes de placement. Ce dictionnaire de gestes peut sembler trop important à retenir, la mémoire de travail n'étant capable de manipuler simultanément qu'environ 7 éléments [Miller, 1956].

Nous faisons cependant la supposition que la métaphore de l'objet virtuel permet d'englober plusieurs gestes comme un seul élément : il est possible d'envoyer l'OV dans plusieurs direction, de l'écraser, ou de le déplacer. Il est alors plus juste de compter en terme de manipulations de l'OV, et non de gestes. Dans ce cas, nous sommes en dessous de la limite admise de la mémoire de travail puisque nous proposons 3 manipulations.

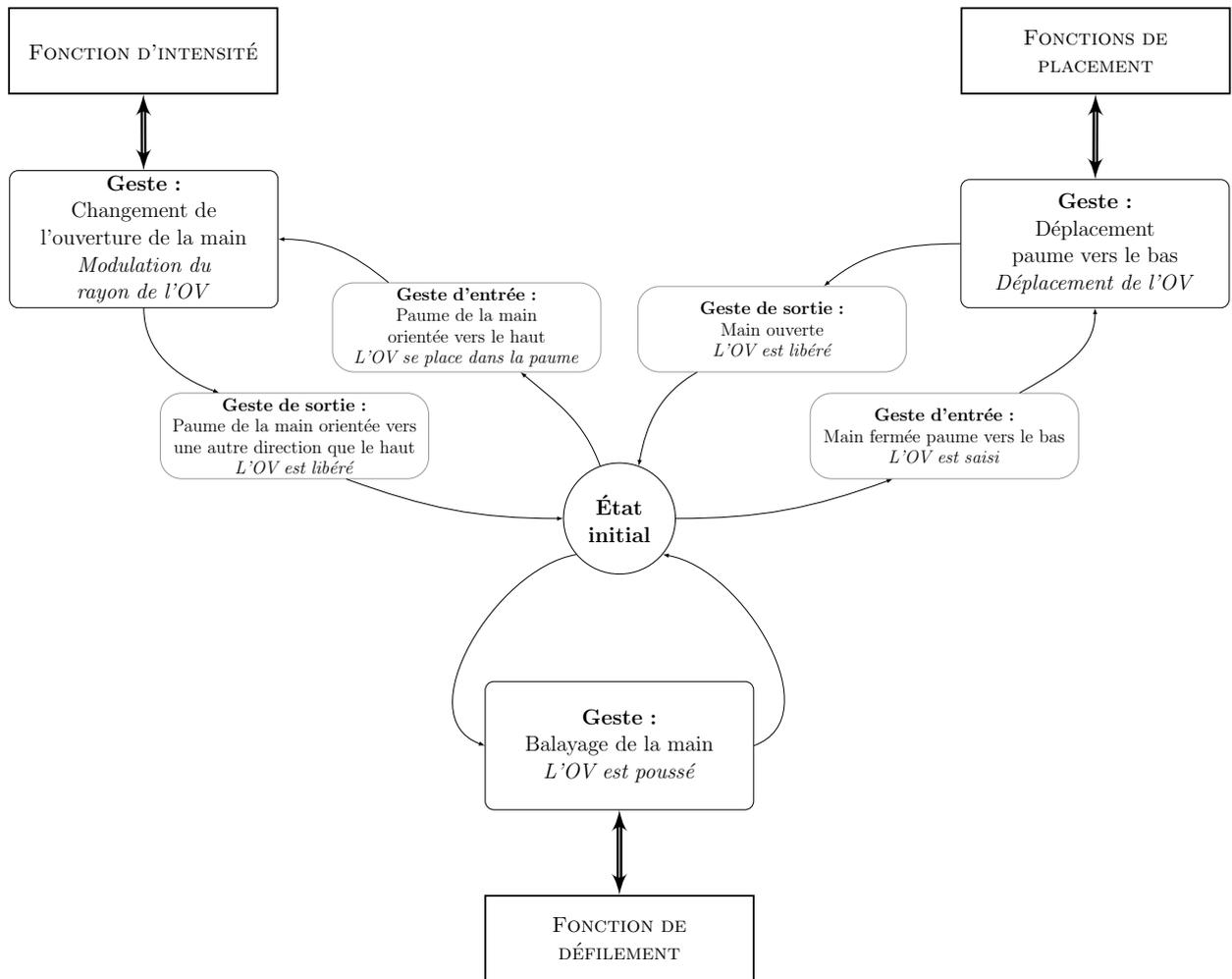


FIGURE 6.1 – Schéma récapitulatif des liens entre l'OV, les gestes et les fonctions. Les descriptions en italique indiquent les actions subies par l'OV dans le cadre de la métaphore que nous avons mise en place. Les familles de fonctions sont liées aux gestes par les doubles flèches.

6.1.4 Stimuli audio-visuels et construction des interfaces

Nous allons à présent présenter succinctement les stimuli visuels et sonores liés aux interfaces de ventilation et de musique. Visuellement, nous avons tenté de conserver la même construction, pour faciliter l'assimilation des fonctions de la même famille.

Interface de ventilation : rendu visuel et fonctionnement

Le rendu visuel de l'interface de ventilation que nous avons mise au point est composé d'une jauge circulaire renseignant sur l'intensité de ventilation, d'un indicateur de la position du flux d'air, et enfin d'un axe de température.

Comme précisé précédemment, l'intensité de ventilation est réglée grâce à l'ouverture de la main, ce paramètre étant obtenu grâce au SDK développé par la société Leapmotion. Lorsque le vecteur normal à la main est colinéaire et opposé à la gravité \vec{g} – avec une certaine tolérance, l'objet virtuel vient alors au centre de la main et son rayon est modulé par ce paramètre d'ouverture de main. Les huit quartiers de la jauge d'intensité s'allument et s'éteignent selon la taille donnée à l'objet virtuel : lorsque la main est ouverte et l'objet est au maximum de sa taille, l'ensemble des quartiers sont allumés. Au contraire, lorsque la main est fermée, la taille de l'objet virtuel est minimale et seulement le premier quartier est allumé.

Nous avons choisi de permettre deux positions différentes pour la source de ventilation. Ce réglage est modifié lorsque l'utilisateur ferme sa main en ayant le vecteur normal de la paume de la main dirigé vers le bas – colinéaire et de même sens que la gravité. À cet instant, la position de la main (x_0, y_0, z_0) est gardée en mémoire et sert de référence. Le réglage est alors mis à zéro. Si l'on note \vec{z} l'axe vertical ascendant et t le temps, les déplacements verticaux de la main fermée $z(t)$ vont alors piloter le réglage du positionnement du flux d'air. Si $z(t) - z_0 < 0$ alors le flux se déplace vers le bas, et se dirige vers les pieds du conducteur. Au contraire, si $z(t) - z_0 > 0$ alors le flux se dirige vers le haut et vers la tête du conducteur. Lorsque l'utilisateur ouvre la main à t_1 , sa position $z(t_1)$ est gardée en mémoire et définit le réglage conservé par l'interface.

Enfin, l'axe de température est initialisé à 20°C, avec la température inscrite au centre. Des signes + et - ainsi que des températures de 19 et 21°C sont inscrits sur les côtés, ce qui permet de suggérer l'orientation horizontale de l'axe des températures. Les balayages de la main arrivant en première zone modifient la température de 2°C, et ceux arrivant dans la seconde zone plus éloignée règlent la température directement au minimum ou au maximum – fixés à 15 et 25°C, respectivement – selon le côté vers lequel l'objet virtuel est envoyé.

Interface de ventilation : retours sonores

SONS CONTINUS

La stratégie sonore de roulement associé à une texture sonore que nous avons utilisée précédemment permet d'illustrer l'ensemble des déplacements de l'OV, et donc de le manipuler en aveugle. Comme nous l'avons précisé au chapitre 5, la stratégie de sonification donne un retour sonore même lorsque l'objet virtuel est immobile, pour que les utilisateurs soient informés qu'ils sont dans la zone d'interaction et que l'OV est présent.

Lors de l'exécution de balayages de la main, la stratégie de sonification donne donc un retour sonore sur les mouvements de l'objet virtuel, ce geste ayant été au cœur des deux chapitres précédents. Nous avons choisi de modifier l'interaction de notre paradigme action - objet lors de la saisie de l'objet virtuel pour le placement de la source de ventilation. Il nous a en effet paru plus cohérent de proposer un son de frottement, correspondant mieux à la réalité physique d'une balle saisie puis déplacée. La hauteur de la stratégie de sonification a également été modulée lors du positionnement vertical de la source, ces dimensions spatiales et sonores paraissant liées [Lemaitre *et al.*, 2017] [Spence, 2011] [Occelli *et al.*, 2009] [Nymoen *et al.*, 2011]. Nous pensons que ce retour d'information peut permettre de guider les sujets sur l'axe haut/bas pouvant être parcouru, reprenant notamment un critère d'ergonomie de Bastien et Scapin [Bastien et Scapin, 1993]. Nous verrons au paragraphe suivant que nous avons fait correspondre ces changements de hauteur avec des icônes sonores.

Enfin, nous n'avons pas jugé nécessaire de rajouter un son continu lors du changement de l'intensité de ventilation, le bruit de ventilation étant naturellement modulé lors des manipulations de son intensité.

ICÔNES SONORES

Nous avons également choisi de recourir à des earcons pour assurer des retours d'informations sur la manipulation des trois fonctions de ventilation. Pour préserver la cohérence de l'interface, nous avons choisi un earcon *principal*, fréquemment utilisé et lu à différentes vitesses pour changer sa hauteur. Nous avons également choisi un *earcon d'entrée* et un *earcon de sortie* pour accompagner les gestes d'entrée et de sortie que nous avons évoqués en figure 6.1.

Lors des manipulations de la température, nous avons diffusé l'earcon principal lu à une vitesse $v = 1$ pour les balayages de faible intensité, augmentant ou diminuant la température de 2°C. Les balayages de grand amplitude, où l'OV est envoyé dans les zones les plus lointaines, sont soulignés par la lecture de l'earcon à différentes vitesses : pour un grand balayage à gauche, l'earcon est lu à une vitesse $v = 0,5$; et à une vitesse $v = 2$ pour un grand balayage à droite. Nous avons choisi de différencier les retours sonores des deux amplitudes de balayages, pour que les sujets puissent connaître le réglage pris en compte par le système en aveugle. Nous avons également choisi de spatialiser les earcons de confirmation à gauche ou à droite.

Cet earcon principal est également joué pour indiquer lorsque la source atteint les limites du réglage de positionnement de la source de ventilation. L'earcon est lu à une vitesse $v = 2$ lorsque la source se trouve sur la tête du conducteur, et à une vitesse de $v = 0,5$ lorsqu'elle atteint ses pieds. La modulation de la hauteur de la stratégie de sonification s'est fait avec la même modification de hauteur lors du déplacement de l'OV vers le haut ou vers le bas.

Enfin, pour la modification de l'intensité de ventilation, nous avons simplement choisi de jouer l'icône sonore principal à chaque fois qu'un quartier de la jauge circulaire apparaît ou disparaît.

Interface de gestion de la musique : rendu visuel et fonctionnement

L'interface pilotant la musique que nous avons créée s'inspire fortement de l'interface de ventilation. Ce choix, combiné au regroupement de fonctions en familles, vise à respecter les critères d'homogénéité [Bastien et Scapin, 1993] pour permettre une adaptation aisée des sujets en cas de changement de page d'interface.

Nous avons choisi de placer une jauge circulaire représentant le volume de la même manière que pour l'intensité de ventilation. Cette jauge est également pilotée par la taille de l'objet virtuel. Nous avons cependant dû opérer un changement pour le placement de la source sonore dans le véhicule. Il est en effet commun de représenter cette fonction avec une vue de dessus, ce que nous avons choisi, alors que nous avions représenté le placement de la source de ventilation avec une vue de côté. Cependant, le réglage du positionnement de la source dans l'habitacle se fait cette fois dans le plan défini par les axes \vec{x} et \vec{y} , pour qu'il y ait une correspondance entre le mouvement de la main, et le ressenti des occupants du véhicule. Faute d'avoir trouvé une représentation plus adaptée, nous avons choisi une vue du dessus. Ce problème de correspondance entre les fonctions peut nuire à l'homogénéité des interfaces. Le même principe de placement est conservé par rapport à l'interface de ventilation : lorsque la main de l'utilisateur se ferme, sa position (x_0, y_0, z_0) est enregistrée et le placement de la source se fait à partir de cette position. Par exemple, si le conducteur place la source sonore sur sa personne en bougeant son poing fermé vers l'avant et vers la gauche, son siège s'illumine.

La possibilité de choisir la piste de musique écoutée est offerte aux utilisateurs. Il est possible de changer de piste dans la liste de lecture que nous avons constitué en poussant l'objet virtuel d'une forte amplitude de balayage, dans la direction voulue. Les balayages de faible amplitude permettent quant à eux d'avancer ou de reculer de 10 secondes dans la piste de musique.

Interface de gestion de la musique : retours sonores

SONS CONTINUS

Lors de l'utilisation de cette interface, nous avons fait le choix de lire automatiquement une piste, pour que les informations relatives aux gestes puissent être transmises. La matière sonore diffusée est en effet directement utilisée pour donner un retour d'information aux conducteurs.

Comme avec l'interface de ventilation, nous avons choisi de rattacher le son de roulement à l'objet virtuel, pour notamment pouvoir percevoir son déplacement lorsque des balayages de la main sont effectués. Cette situation peut en revanche être désagréable en mêlant le son de l'objet virtuel avec la piste écoutée par le conducteur. Ce conflit est cependant de courte durée, puisque le son de l'objet virtuel ne sera pas toujours présent, comme nous le verrons en section 6.2.2. Pour le positionnement de la source sonore, nous n'avons pas jugé opportun d'ajouter un retour d'information par le son en plus des informations données par la fonction elle-même : les utilisateurs entendent la source se déplacer autour d'eux, et ils la positionnent à leur convenance. Le son de l'OV est donc masqué pendant ce réglage. Enfin, nous avons également pensé que le volume sonore véhicule suffisamment d'information pour qu'il n'y ait pas besoin d'ajouter de sons illustrant l'OV. La stratégie de sonification est donc également absente.

ICÔNES SONORES

Avec l'idée de proposer une interface similaire à l'interface de ventilation, nous avons utilisé un earcon principal, joué à différentes vitesses. Les earcons d'entrée et de sortie sont toujours présents, pour indiquer quand l'utilisateur entre en interaction avec les fonctions de positionnement et d'intensité.

Pour les balayages de la main, nous avons choisi de jouer l'earcon principal de la même manière qu'avec l'interface de ventilation dans un souci de correspondance entre les deux interfaces. Cet earcon est donc joué à une vitesse normale pour les balayages de faible amplitude, et moins vite ou plus vite si les balayages sont de grande amplitude vers la gauche ou la droite respectivement. La spatialisation des earcons est également proposée. Le positionnement de la source de la musique ne semble pas nécessiter d'icône sonore, puisque les utilisateurs de l'interface peuvent entendre la musique se déplacer autour d'eux. Nous avons choisi de laisser uniquement le retour d'information continu de la piste jouée. De la même manière, le retour sonore donné par le volume du contenu sonore joué semble suffisamment informatif à lui seul. Aucune icône sonore n'a donc été diffusée lors de la manipulation de cette fonction.

Cohérence des interfaces

Les deux interfaces que nous avons décrites sont voisines, et mis à part le changement de perspective adopté pour le positionnement de la source, s'utilisent de façon identique. Nous avons choisi de représenter les familles de fonctions de façon similaire, et nous les avons placées aux mêmes endroits sur la page d'interface pour suggérer que les mêmes gestes sont utilisés pour les manipuler. Ce procédé permet de créer des interfaces cohérentes, et nous pouvons penser que si un utilisateur maîtrise la première en créant un schéma cognitif spécifique, alors il n'aura pas de difficultés à l'adapter pour utiliser la seconde interface. Ces interfaces similaires peuvent donc permettre de diminuer le temps d'apprentissage d'une interface, en liant plusieurs pages autour d'une structure similaire. Il est donc possible d'imaginer bâtir d'autres pages d'interface sur le même principe, en utilisant les trois familles de fonctions que nous avons abordé, et en respectant le critère d'homogénéité développé par Bastien et Scapin [Bastien et Scapin, 1993].

Les deux interfaces cohérentes que nous venons de présenter permettent de piloter trois fonctions simples des menus de ventilation ou de gestion de la musique. Des familles de fonctions ont été déterminées, et des gestes ont été associés à ces familles, en s'inscrivant dans la métaphore de manipulations d'un objet que nous voulons mettre en place. Nous avons par la suite intégré ce système à un simulateur de conduite, dans le but d'évaluer l'apport d'une interface gestuelle sur le plan attentionnel par rapport aux interfaces tactiles existantes. Nous allons tout d'abord décrire rapidement les éléments constitutifs du simulateur que nous avons utilisé, avant d'aborder les adaptations que nous avons apportées à notre système pour qu'il soit facilement utilisable en conduite.

6.2 Intégration du système et simulateur de conduite

Nous allons à présent nous intéresser aux problèmes d'attention qui peuvent perturber une tâche de conduite. Comme nous l'avons vu en section 3.2, nous pouvons distinguer deux types de distractions lors de la conduite d'un véhicule notamment : la distraction visuelle, et la distraction cognitive. La première se réfère aux stimuli visuels, et désigne par exemple la distraction lors de l'utilisation d'une interface tactile. La seconde concerne notamment les stimuli auditifs, et notre interface gestuelle et sonore entraînera ce type de distraction.

Dans cette précédente section, nous avons également évoqué l'utilisation de simulateurs de conduite pour l'étude de phénomènes attentionnels. Ces simulateurs ont l'avantage de ne pas risquer de causer des dommages aux sujets de l'expérience ni au matériel que nous utilisons. Nous avons également vu que l'utilisation de ces simulateurs peut être d'une validité relative si les stimuli perceptifs principaux de notre expérience s'approchent de la condition réelle. Dans notre cas, la restitution des stimuli visuels, auditifs et tactiles provenant du maniement du véhicule semblent être les informations les plus importantes à restituer aux utilisateurs.

6.2.1 Simulateur de conduite

Le Groupe PSA dispose de plusieurs simulateurs de conduite, permettant de placer les utilisateurs dans des situations plus ou moins réalistes. Dans le cas de cette étude, nous allons nous intéresser particulièrement aux aspects attentionnels pendant la phase de conduite. Comme un simulateur simple semble suffisant [Santos *et al.*, 2005], nous avons choisi un simulateur de conduite statique, où des retours visuels, auditifs et tactiles sont proposés. Les retours provenant de la dynamique du véhicule ne nous paraissent pas pertinent en première approximation, ce qui explique que nous n'avons pas choisi de simulateur dynamique.

Le simulateur se compose d'une partie d'un habitacle d'un véhicule automobile, avec un siège, un volant, une boîte de vitesse séquentielle, de deux pédales d'accélérateur et de frein, et d'un frein à main comme on peut le voir en figure 6.2.



FIGURE 6.2 – Le simulateur de conduite que nous avons utilisé pour l'expérience 5.

Le capteur de gestes a été placé dans le simulateur de conduite au-dessus du levier de vitesse, et orienté vers le bas, comme il est possible de le voir sur la figure 6.3. Ce placement se justifie par la volonté que le capteur soit dans les meilleures conditions pour détecter la main, notamment lorsqu'elle est tournée avec la paume vers le haut. Cette orientation du capteur permet d'obtenir une captation robuste du paramètre d'ouverture de la main.



FIGURE 6.3 – Vue alternative du simulateur de conduite que nous avons utilisé pour l'expérience 5. Le capteur de geste est placé dans la boîte noire placée au-dessus de l'écran central.

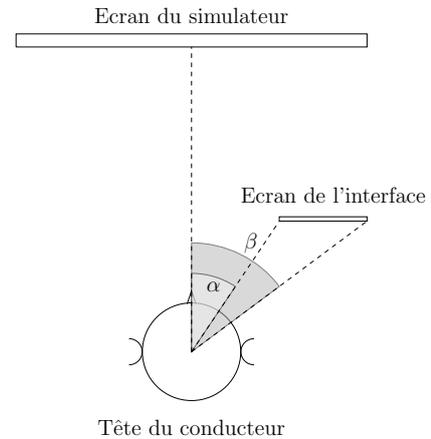


FIGURE 6.4 – Positions relatives des deux écrans dans le simulateur de conduite.

Stimuli visuels

Les stimuli visuels du simulateur sont restitués par un écran centré placé en face des sujets. Comme le montre la figure 6.2, les dimensions de l'écran ne sont pas assez importantes pour couvrir l'ensemble du champ de vision des sujets. Cette limitation peut par exemple être problématique pour la sensation de réalisme de la simulation provoquée chez les conducteurs.

Pour pouvoir restituer certains stimuli visuels de l'interface, nous avons ajouté un écran secondaire, placé au-dessus du levier de vitesse comme le montre la figure 6.3. Nous avons veillé à ce que cet écran soit à portée de main, notamment pour ne pas introduire une difficulté de manipulation en situation tactile. En prenant l'axe horizontal partant du centre des yeux de l'utilisateur et allant vers l'écran du simulateur comme référence, il y a un angle $\alpha \approx 15$ et 20° pour le bord gauche de l'écran, et $\beta \approx 40$ et 45° pour le bord droit comme le montre la figure 6.4.

Stimuli sonores

Les signaux sonores sont restitués par 6 enceintes Genelec 8050 BPM placées autour du poste de conduite, comme le montrent les figures 6.2 et 6.3. Sur ce simulateur, certaines données de conduite sont collectées puis traitées pour générer un bruit de moteur et aérodynamique cohérents avec la conduite. Le son restitué aux sujets est donc composé de bruits engendré grâce aux paramètres de conduite, et des sons de l'interface que nous avons mise au point.

Le simulateur que nous avons utilisé ne permet pas de raccorder notre interface à la ventilation à la partie d'habitacle dont nous disposons. Pour illustrer le retour d'information sonore de la ventilation, nous avons ajouté un son de ventilation artificiel, dont le volume est modulé par la taille de l'OV. Il n'a en revanche pas été possible de simuler le retour d'information tactile de l'air expulsé par la ventilation.

6.2.2 Verrouillage de l'interface

Pour que l'interface gestuelle soit utilisable en conduite de manière agréable, nous avons ajouté un système de verrouillage automatique de l'interface. Ce verrouillage permet de ne pas tenir compte de tous les gestes qui ont lieu – lorsqu'un occupant du véhicule effectue des gestes coverbaux par exemple – mais seulement ceux destinés à l'interface. De cette manière, les fonctions ne sont pas modifiées involontairement. Il est à noter qu'un tel système n'est pas utile pour l'interface tactile, les gestes avec contact étant beaucoup plus rarement accomplis involontairement. Cependant, notre système de verrouillage automatique s'inspire de celui présent sur les smartphones à dalle tactile, en empêchant les manipulations non-désirées.

Le verrouillage de l'interface se fait automatiquement lorsque la main de l'utilisateur sort d'une zone prédéfinie sous le capteur de gestes. Les frontières de cette zone ont été définies arbitrairement en tenant compte des contraintes du simulateur. Par exemple, l'interface se verrouille avant que la main entre en contact avec le volant, ou avant que la main atteigne le levier de vitesse. En revanche, cette zone est relativement étendue sur la droite, pour permettre aux sujets d'effectuer des balayages vers la gauche ou vers la droite sans gêne. Si la main sort de la zone pendant un geste, le réglage est interrompu et l'interface se verrouille.

Lorsque l'interface gestuelle est verrouillée, un earcon sonore spécifique est lu pour indiquer l'état du système. La sonification des mouvements de l'objet virtuel est alors réduite au silence, et le visuel de l'interface s'assombrit. Lorsque l'interface est verrouillée, les seuls sons émis proviennent de la ventilation artificielle, ou de la musique. Pour déverrouiller l'interface, il est nécessaire d'actionner un bouton situé sous le volant. À cet instant, un second earcon indique le déverrouillage de l'IHM, l'assombrissement de l'interface disparaît et l'objet virtuel est de nouveau audible. Le système est alors à l'écoute des gestes de l'utilisateur.

Le verrouillage de l'interface permet de ne pas restituer en permanence le son de l'objet virtuel, qui peut être gênant comme nous l'avons évoqué dans l'expérience 3, mais seulement lorsqu'il est utile c'est-à-dire lorsque l'interface est déverrouillée. Deux icônes sonores ont été choisies spécifiquement pour le déverrouillage et le verrouillage de l'interface, pour que cette information soit transmise sans ambiguïté.

La mise au point d'une interface gestuelle, puis son intégration au simulateur de conduite à notre disposition va nous permettre d'aborder la question de l'attention liée à l'utilisation d'une IHM en situation de conduite. Comme nous l'avons vu en section 3.2, les phénomènes distractifs sont la cause de nombreux accidents de la route. Cette question d'attention est centrale, puisqu'elle a en partie motivé la mise en place de ces travaux de thèse. Nous proposons donc une alternative aux interfaces tactiles actuelles, mobilisant une attention visuelle, grâce à notre interface associant gestes et sons. Nous allons donc tenter de comparer ces deux interfaces dans une cinquième expérience perceptive.

6.3 Expérience 5 : comparaison d'interfaces en situation de conduite

Nous avons jusqu'à présent mené des études en condition laboratoire, sans tenir compte des aspects distractifs. Dans les dernières expériences, nous avons cependant amorcé un rapprochement vers des conditions d'utilisation réelles en véhicule. Après avoir mis au point une interface gestuelle, nous allons donc nous intéresser aux aspects distractifs de l'utilisation d'interfaces simultanément à une tâche de conduite.

6.3.1 Hypothèses initiales et motivations

Au cours de cette expérience, nous désirons tout d'abord estimer si les conducteurs parviennent à manipuler l'interface gestuelle que nous avons mise au point. Cette interface inconnue peut en effet être complexe à manipuler. Il sera donc nécessaire de prévoir une période de prise en main, où les sujets de l'expérience pourront développer les schémas cognitifs nécessaires à l'utilisation de l'interface simultanément à une tâche de conduite. Nous désirons ensuite observer si notre interface gestuelle peut apporter des gains en terme d'attention, et donc de sécurité, ce qui pourrait être intéressant pour le Groupe PSA. Pour répondre à ces interrogations, nous avons fait le choix de nous baser sur notre premier cas d'usage, qui concerne une interface de ventilation. L'interface permettant la gestion de la musique est proposée brièvement en fin d'expérience.

L'utilisation d'un simulateur de conduite permet de placer les sujets de cette expérience dans des conditions analogues à une utilisation en véhicule. Il est cependant nécessaire, pour jauger les performances de la combinaison des modalités gestuelles et sonores, de les comparer à une situation de référence. Cette procédure permet de situer l'utilisation de notre interface gestuelle par rapport à un autre type d'interface, pour en souligner les avantages et inconvénients. Nous avons choisi une interface tactile pour effectuer cette comparaison, qui est le type d'interface implanté dans les véhicules récents. Nous avons donc développé une interface tactile ayant à la fois un aspect similaire à l'interface gestuelle, et comparable aux interfaces des véhicules actuels.

Développement d'une interface tactile

Cette interface tactile ne doit pas différer dans sa construction avec l'IHM gestuelle, pour qu'il soit possible de comparer uniquement les modalités qui interviennent dans les interactions. Nous avons opté pour une construction d'interface « calquée » sur l'IHM gestuelle.

Il est possible d'interagir avec l'interface en appuyant directement sur la zone concernée par la fonction voulue à l'écran. Par exemple, l'appui sur un quartier de la jauge circulaire active l'intensité de ventilation correspondante ; ou l'appui sur la partie haute ou basse de la partie consacrée au positionnement de la source déplace la source dans la zone correspondante. L'axe de température a été doté de deux zones tactiles sous les flèches pour changer la température de 2°C. Nous avons également ajouté deux boutons aux extrémités gauche et droite, pour qu'il soit aussi possible de mettre directement la température au maximum ou au minimum avec l'interface tactile, au même titre qu'avec des balayages de la main de grande amplitude. Nous avons veillé à ce que les surfaces des différents boutons pour une même famille de boutons soient les mêmes, pour ne pas introduire de biais entre les différentes manipulations. Les positions des boutons sont cependant un autre biais : il est probablement plus facile d'appuyer sur le bouton « Min » que sur le bouton « Max », puisque le premier est plus proche du volant, et donc de la main du conducteur. Nous ne pouvons cependant pas contrebalancer ce biais.

Pour chaque manipulation de l'interface, un earcon est restitué aux utilisateurs, comme c'est couramment le cas sur les interfaces tactiles actuelles implantées en véhicule. Nous avons choisi les mêmes earcons que ceux utilisés pour l'interface gestuelle, avec les mêmes vitesses de lecture pour les mêmes commandes.

L'utilisation d'une interface tactile comme base de comparaison à notre interface gestuelle peut sembler déséquilibrée. Les interfaces tactiles, avec l'utilisation massive de smartphones et de tablettes, sont manipulées très couramment, à la différence des interfaces gestuelles. Il est donc possible que les sujets, grâce à leurs habitudes de manipulations, maîtrisent les interfaces tactiles beaucoup plus facilement que l'interface gestuelle. Pour essayer de minimiser cette différence entre les deux interfaces, nous avons suivi une procédure visant à laisser les sujets apprendre à utiliser l'objet virtuel avant d'effectuer la comparaison des deux types d'interface en conduite.

6.3.2 Méthodologie

Panel de sujets

Trente sujets droitiers travaillant pour le Groupe PSA ont participé à cette expérience, dont 10 femmes et 20 hommes. Ils sont âgés entre 20 à 53 ans – âge moyen : 33,8 ans ; écart type : 10,1 ans – sont tous titulaires du permis B, et n'ont pas déclaré de problèmes moteurs ou auditifs. Aucun des sujets n'a participé aux premières expériences, de sorte qu'aucun d'entre eux n'a manipulé l'objet virtuel avant l'expérience.

Dispositif expérimental

Les sujets de cette expérience sont assis tout le long de cette expérience dans le simulateur de conduite statique que nous avons présenté en section 6.2. Le capteur de geste est placé au-dessus du levier de vitesse, et permet de capter la main droite des sujets.

Comme présenté dans la section précédente, un écran placé au-dessus du levier de vitesse permet aux sujets de voir l'interface, comme le montre la figure 6.3. Cet écran permet une captation tactile grâce à une technologie capacitive, ce qui permet de l'utiliser lors des passations avec la technologie tactile.

Scénario d'utilisation de l'interface en conduite

Pour comparer les interfaces tactiles et gestuelles que nous avons mises au point, nous avons choisi de nous placer dans un scénario précis. Nous imaginons que le conducteur est au volant de son véhicule et suit une direction donnée par son GPS, sur l'interface centrale. Dans notre scénario, le conducteur a alors envie de modifier une fonction de l'interface de ventilation. Nous proposons alors deux façons différentes d'accomplir ce réglage.

INTERFACE GESTUELLE

Avec l'interface gestuelle, le conducteur peut déverrouiller son interface grâce au bouton situé sous le volant, un carcon lui confirmant le déverrouillage. Il effectue alors le geste correspondant au réglage voulu, les retours sonores permettent à l'utilisateur de savoir si son geste a bien été pris en compte, et si le réglage a été correctement effectué. Lorsque la main du conducteur revient sur le volant, une icône sonore indique le verrouillage de l'interface, et la conduite peut alors se poursuivre. Il est à noter que nous avons fait en sorte que l'ensemble des manipulations détaillées ici ne requièrent pas la vision du conducteur. Nous avons donc uniquement mis à disposition des stimuli sonores relatifs à l'IHM dans ce cas, en masquant toute information visuelle de l'interface de ventilation et en proposant uniquement le visuel du GPS. Visuellement, l'interface reste sur la page de GPS lors de manipulations gestuelles.

INTERFACE TACTILE

Il est tout d'abord nécessaire de changer de page de menu avec l'interface tactile, pour avoir accès aux réglages de ventilation. Ce changement de page peut être accompli en appuyant sur le bouton virtuel sur la gauche de la figure. Les conducteurs peuvent alors effectuer le réglage voulu en appuyant sur l'écran, un retour visuel et sonore leur confirmant leur manipulation. Il est ensuite nécessaire de revenir au menu de GPS pour poursuivre la navigation GPS en changeant une nouvelle fois de menu grâce à un second bouton virtuel situé sur la gauche de l'interface.

COMPARAISON DES INTERACTIONS

Nous pouvons noter que pour un même réglage de l'interface, la procédure à effectuer est différente entre les deux interfaces : déverrouillage, geste, puis verrouillage dans un cas ; changement de menu, réglage tactile, puis changement de menu dans l'autre. Nous n'avons pas cherché à comparer exclusivement l'interaction permettant le changement de paramètre de l'interface, mais l'ensemble des changements qui peuvent s'opérer dans une situation donnée lorsque les modalités d'interactions avec l'interface sont modifiées. L'interface gestuelle peut par exemple permettre de se passer de retour visuel, ce qui supprime l'étape de changement de menu, nécessaire en tactile. En revanche, nous pouvons raisonnablement penser que l'interaction est plus facile à exécuter en tactile – appuyer sur une zone de l'écran – qu'en geste – effectuer un geste précis dans une zone précise. Le point de vue que nous avons adopté d'utiliser un scénario permet une comparaison de l'ensemble des changements apportés par l'interface gestuelle.

Les manipulations d'interfaces tactiles sont courantes, il est donc possible que les utilisateurs effectuent facilement les étapes que nous venons de décrire. En revanche, l'utilisation de gestes pour piloter une interface étant peu courante, il est nécessaire de laisser le temps aux sujets pour qu'ils développent les schémas cognitifs nécessaires au bon maniement de l'objet virtuel. Nous avons donc mis au point un protocole se déroulant sur deux jours consécutifs, où une grande partie du temps est consacrée à la prise en main de l'objet virtuel, puis à un entraînement à l'arrêt pour maîtriser les réglages associés.

Stimuli et procédure

Pour résumer l'ensemble des gestes qu'il est possible de réaliser, ainsi que les retours sonores proposés aux utilisateurs, nous avons annoté la figure 6.1 pour en proposer une version plus complète appliquée à l'interface de ventilation, en figure 6.5. Les notes ♪ indiquent le type de son lié à l'interaction correspondante.

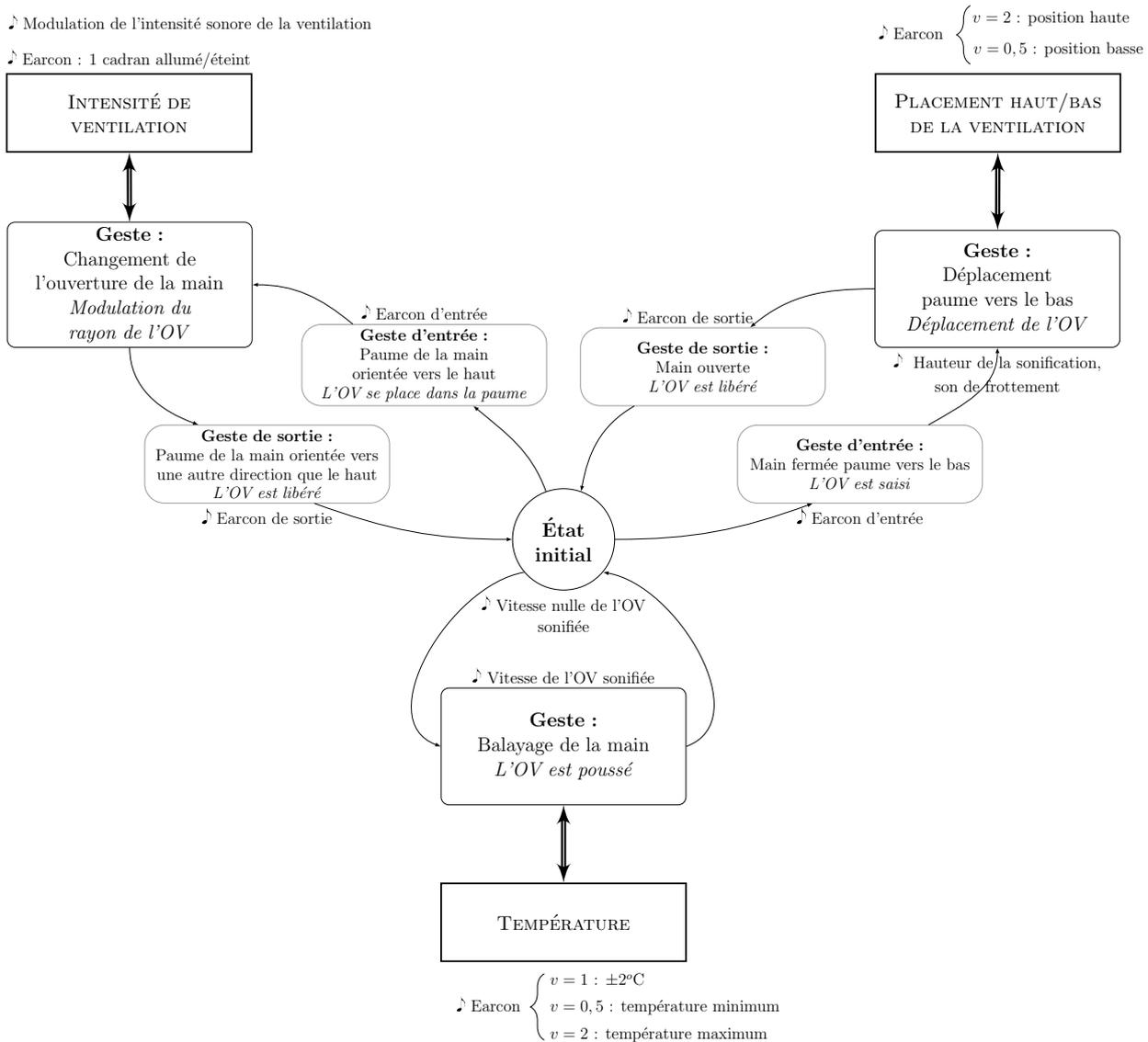


FIGURE 6.5 – Schéma récapitulatif des liens entre l'OV, les gestes, les fonctions de ventilation, et les sons. Les descriptions en italique indiquent les actions subies par l'OV dans le cadre de la métaphore que nous avons mis en place. Les familles de fonctions sont liées aux gestes par les doubles flèches. Les notes ♪ décrivent le retour sonore proposé pour chaque geste ou fonction.

L'ensemble des trente sujets utilisent les deux interfaces, quinze commençant par l'interface gestuelle, et les quinze autres utilisant d'abord l'interface tactile. Comme nous l'avons abordé précédemment, il est hautement probable que les sujets aient une plus grande habitude d'utiliser les interfaces tactiles que les interfaces utilisant des gestes. Nous avons donc choisi de faire l'expérience sur 2 jours consécutifs.

Le premier jour est consacré à la prise en main de l'interface, où les sujets doivent développer des schémas cognitifs pour maîtriser l'objet virtuel. Plusieurs études ont en effet souligné l'apport du sommeil dans l'acquisition de connaissances [Stickgold *et al.*, 2000] [Maquet, 2001] [Walker *et al.*, 2002]. La séance est divisée en deux phases distinctes : une première phase où les sujets découvrent le fonctionnement de l'interface, et une seconde phase au cours de laquelle un entraînement leur est proposé.

Au début du deuxième jour, consacré aux manipulations en conduite, une prise en main du simulateur est tout d'abord proposée, pour permettre aux sujets de s'adapter à cette conduite particulière. Par la suite, les sujets s'entraînent une nouvelle fois sur l'interface, pour solliciter de nouveau les schémas cognitifs mis en place la veille. Consécutivement à cette phase, les sujets passent à la phase de test de l'interface en situation de conduite, sur le même circuit que lors de la prise en main du simulateur. À l'issue de la seconde séance, il est proposé aux sujets une manipulation libre de l'interface de gestion de la musique présentée précédemment, avant de répondre à un questionnaire.

Pour un sujet commençant par l'interface gestuelle, les étapes suivies pendant cette expérience sur les deux jours consécutifs – notés jour 1 et jour 2 – sont donc les suivantes :

- Jour 1 :
 - phase de découverte de l'interface gestuelle,
 - entraînement 1 sur l'interface gestuelle,
 - phase de découverte de l'interface tactile,
 - entraînement 1 sur l'interface tactile,
- Jour 2 :
 - prise en main du simulateur de conduite,
 - entraînement 2 sur l'interface gestuelle,
 - test de l'interface gestuelle en conduite,
 - entraînement 2 sur l'interface tactile,
 - test de l'interface tactile en conduite,
 - manipulation libre de l'interface de gestion de la musique,
 - questionnaires.

Nous allons maintenant nous attacher à décrire chacune de ces étapes plus précisément, en détaillant les différents stimuli utilisés.

Jour 1 : phase de découverte de l'interface gestuelle

En ce qui concerne la découverte de l'objet virtuel, nous avons voulu tirer les enseignements des conclusions de l'expérience 4. Dans cette précédente expérience, il semble que l'attention des sujets se soit focalisée sur la représentation visuelle de l'objet virtuel, ce qui a potentiellement réduit l'attention portée aux stratégies sonores que nous avons mises au point. Dans cette expérience, nous visons à utiliser uniquement des retours d'informations sonores lors de l'utilisation de l'interface gestuelle en conduite. Nous allons donc faire en sorte que l'attention des sujets soit focalisée sur les sons émis lors des manipulations de l'objet virtuel ou de l'interface.

Nous avons donc choisi de montrer l'objet virtuel seulement au début de la première séance. Les sujets peuvent ainsi avoir une représentation mentale de l'objet virtuel, et associer les variations de la stratégie sonore à des déplacements de l'objet virtuel. Après cette période de découverte, l'OV est caché pour que l'attention des sujets se porte sur la stratégie de sonification.

Nous avons découpé cette phase de découverte en plusieurs étapes étudiées en série, pour permettre une compréhension de plusieurs concepts de manière isolée sans provoquer de surcharge cognitive. Nous avons par la suite expliqué les différentes interactions entre éléments progressivement. Une présentation globale du système peut en effet être un frein à la compréhension des sujets, demandant une mémoire de travail trop importante comme nous l'avons vu en section 3.1.2 [Pollock *et al.*, 2002] [Van Merriënboer et Sweller, 2005]. Chaque étape débute par l'explication d'un concept, avant de laisser les sujets interagir avec le système pour mettre en pratique la notion abordée. La création de schémas cognitifs peut être facilitée par cette procédure progressive, avec des manipulations interactives de différentes parties de l'interface.

Nous avons résumé les étapes de cette phase de découverte dans le tableau 6.1. Une présentation détaillée est proposée en annexe C. Il est important de noter que les sujets peuvent voir leur main virtuelle et l'OV des étapes 1 à 5, mais ces retours visuels ne sont plus disponibles à partir de l'étape 6 incluse, où le visuel de l'interface est proposé.

N° d'étape	Stimuli visuels	Stimuli sonores	Objectif
1	Main seule	-	Découverte de la reconstruction de la main
2	Main + OV	Sonification de l'OV	Découvrir l'OV, et les sons qui lui sont rattachés
3	Main + OV	Sonification de l'OV	Faire des balayages de main à gauche et à droite
4	Main + OV	Sonification de l'OV + earcons (prise en main de l'OV)	Saisie de l'OV
5	Main + OV	Sonification de l'OV + earcons (prise en main de l'OV)	Modification de la taille de l'OV
6	Interface de température	Sonification de l'OV + earcons	Modifier la température de ventilation
7	Interface de température + placement de la ventilation	Sonification de l'OV + earcons	Modifier la localisation de la ventilation
8	Interface de ventilation complète	Sonification de l'OV + earcons	Modifier l'intensité de ventilation
9	Interface de ventilation complète + assombrissement (verrouillage)	Sonification de l'OV + earcons	Maîtriser le verrouillage de l'interface

TABLEAU 6.1 – Résumé des étapes de la phase de découverte de l'OV.

Jour 1 : entraînement 1 sur l'interface gestuelle

À la suite de cette phase de découverte de l'interface, les sujets sont invités à s'entraîner à utiliser l'interface, à l'arrêt dans un premier temps. Nous les plaçons donc dans le scénario que nous avons choisi pour tester l'interface en conduite : avec le visuel du GPS, et avec des retours d'informations sonores uniquement. Les sujets doivent donc déverrouiller l'interface avec le bouton proche du volant, changer la fonction de la manière demandée, puis sortir la main de la zone de détection pour verrouiller l'interface.

Lors de cette phase d'entraînement, il est demandé aux sujets par des enregistrements de voix d'effectuer un réglage de l'interface. Lorsque le bon réglage est effectué, même après plusieurs gestes, une nouvelle manipulation de l'interface est demandée au sujet. L'entraînement se poursuit ainsi jusqu'à la fin des 54 manipulations de l'entraînement, ce qui termine la première séance.

Les manipulations dictées aux sujets sont les suivantes :

1. augmentez fortement la température,
2. augmentez faiblement la température,
3. baissez faiblement la température,
4. baissez fortement la température,
5. mettez l'intensité de ventilation au minimum,
6. mettez l'intensité de ventilation environ à la moitié,
7. mettez l'intensité de ventilation au maximum,
8. mettez la source de ventilation sur votre tête,
9. mettez la source de ventilation sur vos pieds.

Quatre manipulations correspondant aux quatre zones de balayages de la main ont été choisies. Concernant l'intensité de ventilation, nous avons décidé de proposer trois manipulations correspondant au minimum, maximum, mais également un réglage entre ces deux extrêmes. Ce réglage plus fin ne demande cependant pas une extrême précision, puisque nous considérons ce réglage comme réussi lorsqu'il y a entre trois et six cadrants inclus allumés sur les huit possibles. Enfin, nous avons proposé deux réglages pour le positionnement de la source de ventilation, correspondant au maximum dans le sens vertical ascendant et descendant. Les 54 réglages correspondent à 6 essais de chaque manipulation, tirés dans un ordre aléatoire.

À l'issue de ces 54 manipulations, nous avons proposé aux sujets deux mesures de leur entraînement. 18 essais, 2 de chaque instruction, leur sont demandés dans une condition sans aucun retour d'information, pour observer si la première manipulation exécutée est bonne. De cette manière, il est possible de mesurer le degré de maîtrise des gestes seuls, pour observer les manipulations problématiques et celles facilement exécutées.

Une seconde série de 18 essais est ensuite attendue, proposés cette fois-ci avec retours d'informations. À la différence de l'entraînement où l'enchaînement des manipulations est automatique, les sujets doivent ici indiquer à l'expérimentateur quand ils pensent avoir fait le réglage demandé, quelque soit le nombre de gestes nécessaires. Il est ainsi possible de mesurer si les sujets ont bien compris le fonctionnement de l'interface, et en particulier s'ils interprètent correctement les sons de validation des réglages.

Jour 1 : phase de découverte de l'interface tactile

La technologie tactile étant en général bien maîtrisée par les sujets de notre expérience, cette étape de découverte a été relativement courte. Nous avons commencé directement par montrer l'axe de température aux sujets, en les invitant à modifier les réglages et à écouter les sons émis – ce qui correspond à l'étape 6 de la phase de découverte de l'interface gestuelle.

Nous avons ensuite présenté le positionnement de la source, puis la jauge circulaire de l'intensité de ventilation, tout en demandant aux sujets d'essayer plusieurs réglages, et de porter leur attention sur les sons émis. Le verrouillage d'interface n'étant cependant pas nécessaire avec la technologie tactile, nous sommes donc ensuite passés directement à la phase d'entraînement.

Jour 1 : entraînement 1 sur l'interface tactile

L'entraînement avec l'interface tactile est proche de l'entraînement avec l'interface gestuelle. 54 réglages correspondant à 6 essais de chacune des 9 manipulations sont proposés. Les sujets doivent alors changer le menu pour afficher la page concernant la ventilation, faire le réglage demandé, puis revenir au menu de GPS.

De la même manière qu'avec l'interface gestuelle, nous avons effectué deux mesures de l'entraînement. Lors des 18 essais de la première mesure, les sujets doivent appuyer sur la zone correspondant au bon réglage, alors que les retours sonores et visuels sont absents – seul le fond de l'interface est visible, pas les éléments en surbrillance. Pour la seconde série de 18 essais, les sujets doivent faire le bon réglage et l'indiquer à l'expérimentateur, qui passe à l'instruction suivante. Ces deux mesures peuvent paraître inutiles car les tâches semblent évidentes : il est cependant nécessaire de faire rigoureusement la même procédure pour les deux technologies que nous utilisons dans ce test.

Jour 2 : prise en main du simulateur de conduite

Le début de la deuxième journée est consacré à la prise en main du simulateur de conduite. Cette étape a pour but d'habituer les sujets à la conduite sur simulateur, qui est légèrement différente de la conduite en véhicule. Avec cette phase d'adaptation nous supposons que, si certaines différences sont constatées sur les données de conduite entre les conduites avec interfaces gestuelle ou tactile, alors elles seront dues à la différence d'interface et non pas à un manque d'habitude d'utilisation du simulateur.

Nous avons choisi de faire conduire les sujets sur le même circuit utilisé pour les manipulations des interfaces, pour que la découverte du circuit n'influence pas les résultats. Le parcours de l'ensemble du circuit dure environ 10 minutes. Ce circuit représente une route de campagne où la vitesse est limitée à 90 km.h^{-1} , avec plusieurs intersections. Sur ce circuit, des véhicules ont été ajoutés pour simuler un trafic avec une densité subjectivement jugée normale. Comme nous l'avons vu précédemment, le simulateur permet seulement de voir la scène située en face du conducteur, nous avons donc fait en sorte qu'il n'y ait pas de trafic arrivant sur les côtés du véhicule du conducteur aux intersections, et nous l'en avons informé.

Lors de la conduite, nous avons donné plusieurs consignes aux conducteurs. La première consiste à maîtriser son véhicule et à respecter le code de la route comme lors d'une conduite sur route. Nous avons également ajouté une tâche à l'utilisateur, pour ajouter une sollicitation visuelle : suivre un véhicule *lièvre* tout en maintenant avec lui une distance minimum. Ce véhicule représente un conducteur dont la conduite varie selon les conditions de conduite. Nous avons choisi d'assigner à ce véhicule une conduite fluide sur la première moitié du circuit, puis une conduite plus chaotique sur la seconde moitié. Sur cette seconde moitié du circuit, la vitesse moyenne du véhicule lièvre est augmentée de 10 km.h^{-1} , et nous avons ajouté trois freinages et accélérations d'intensités plus ou moins importantes, qui sont renseignés sur la figure 6.6.



FIGURE 6.6 – Vue d'ensemble du circuit proposé. Les endroits où les instructions ont été proposées sont signalés par des croix vertes, avec la notation I_x , où x est le numéro de l'instruction.

La distance à laquelle les sujets doivent suivre le véhicule lièvre est indiquée par une imitation de vision tête haute – ou *VTH* – affiché sur l'écran du simulateur, où un rappel de la vitesse est également indiqué. La couleur de cette *VTH* passe à l'orange si le conducteur est à plus de 40 mètres du véhicule lièvre, et au rouge si la distance est supérieure à 60 mètres, comme nous pouvons le voir sur les figures D.1, D.2 et D.3 en annexe. Nous avons indiqué aux conducteurs que la distance avec le véhicule lièvre doit rester inférieure à 40 mètres, et que la *VTH* ne doit donc pas être en orange ni en rouge. Si ce cas se produit, les sujets sont invités à accélérer pour que la distance inter-véhiculaire passe en dessous de la limite des 40 mètres que nous avons fixée.

Nous supposons que cette tâche de conduite sollicite suffisamment la modalité visuelle des conducteurs pour simuler une véritable situation de conduite, où le trafic et les événements extérieurs requièrent une vigilance visuelle soutenue. À la suite de cette étape, nous avons considéré que les sujets maîtrisent la conduite sur le simulateur.

Jour 2 : entraînement 2 sur l'interface gestuelle

Avant de proposer un deuxième entraînement, nous avons choisi de procéder à nouveau à deux mesures de la maîtrise de l'interface. De cette manière, nous pourrions comparer les résultats à la fin du jour 1 avec ceux en début du jour 2, et évaluer ainsi les différences.

L'entraînement est ensuite proposé aux sujets, pour solliciter de nouveau les schémas cognitifs mis en place la veille. Le déroulement est similaire à celui du jour 1, si ce n'est que la durée n'est que de 18 manipulations pour ne pas faire apparaître de phénomène de fatigue, ce qui pourrait dégrader la conduite ultérieure.

Cette phase est clôturée par une troisième série de deux mesures de la maîtrise de l'interface, pour mesurer les effets de l'entraînement.

Jour 2 : test de l'interface gestuelle en conduite

Lors de la phase de test, les sujets sont invités à piloter le simulateur de conduite sur le circuit précédemment parcouru. Dans le même temps, des manipulations de l'interface sont demandées oralement à différents endroits du circuit. Les sujets ont pour consigne d'accomplir les changements d'interface demandés, tout en assurant en priorité une conduite sécuritaire. Ils peuvent donc attendre quelques secondes avant de manipuler l'interface s'ils le jugent nécessaire.

Vingt-quatre instructions ont été disposées régulièrement sur le circuit, sur des portions où les manipulations de l'interface semblent possibles – pas en virage par exemple. La figure 6.6 montre l'ensemble du circuit parcouru, ainsi que les endroits où les instructions ont été proposées. Le départ, l'arrivée ont été indiqués, ainsi que le début des deux conditions de conduite du véhicule lièvre, notées condition lièvre 1 ou 2. Les zones de freinage et d'accélération du véhicule lièvre sont également indiquées. Il est à noter que nous avons choisi de fixer l'ordre des instructions : tous les sujets vont exécuter la même séquence d'instructions.

En résumé, lors de cette phase de test, les sujets doivent à la fois maîtriser leur véhicule, suivre le véhicule lièvre en étant suffisamment proche de lui, et exécuter les manipulations de l'interface demandées.

Jour 2 : entraînement 2 sur l'interface tactile

De la même manière qu'avec l'interface gestuelle, il est proposé aux sujets un second entraînement sur l'interface tactile avec 18 manipulations présentées dans un ordre aléatoire. Cet entraînement est également précédé et suivi par deux mesures de la maîtrise de l'interface.

Jour 2 : test de l'interface tactile en conduite

La phase de test avec l'interface tactile se déroule de la même manière qu'avec l'interface gestuelle, dans les mêmes conditions. Les vingt-quatre instructions sont disposées de la même manière sur le circuit, et il est également demandé de suivre le véhicule lièvre de manière à ce que la distance inter-véhiculaire soit inférieure à 40 mètres. Les sujets sont invités à donner la priorité à une conduite sécuritaire par rapport aux manipulations de l'interface.

Il est important de rappeler que l'ordre de présentation des interfaces tactile et gestuelle est contrebalancé sur l'ensemble des sujets, pour ne pas influencer les résultats.

Jour 2 : manipulation libre de l'interface de gestion de la musique

À la suite du test des interfaces gestuelle et tactile, nous avons souhaité proposer aux sujets de manipuler grâce à leurs gestes l'interface de gestion de la musique que nous avons mise au point. Le but est ici d'observer les impressions des sujets quant à ce type d'interface à l'arrêt. Nous avons prévu ensuite de mesurer le ressenti des sujets au travers d'un questionnaire qui concerne d'abord cette interface, puis l'ensemble de l'expérience.

Analyse des données

Pour analyser les données, nous avons procédé à différents tests, selon le type de données dont nous disposons. Nous avons utilisé des tests de Cochran, des tests de McNemar, des tests du χ^2 , des ACP suivies d'ANOVA et des tests de Student.

6.3.3 Résultats

Nous allons à présent détailler les résultats par condition, en distinguant les manipulations effectuées à l'arrêt ou en conduite. Pour certains résultats, un détail est donné par manipulation. Nous en avons précédemment donné la liste, mais nous l'indiquons de nouveau accompagnée d'une notation pour plus de lisibilité :

1. T_{\max} : augmentez fortement la température,
2. T_+ : augmentez faiblement la température,
3. T_- : baissez faiblement la température,
4. T_{\min} : baissez fortement la température,
5. I_{\min} : mettez l'intensité de ventilation au minimum,
6. I_{moit} : mettez l'intensité de ventilation environ à la moitié,
7. I_{\max} : mettez l'intensité de ventilation au maximum,
8. S_T : mettez la source de ventilation sur votre tête,
9. S_P : mettez la source de ventilation sur vos pieds.

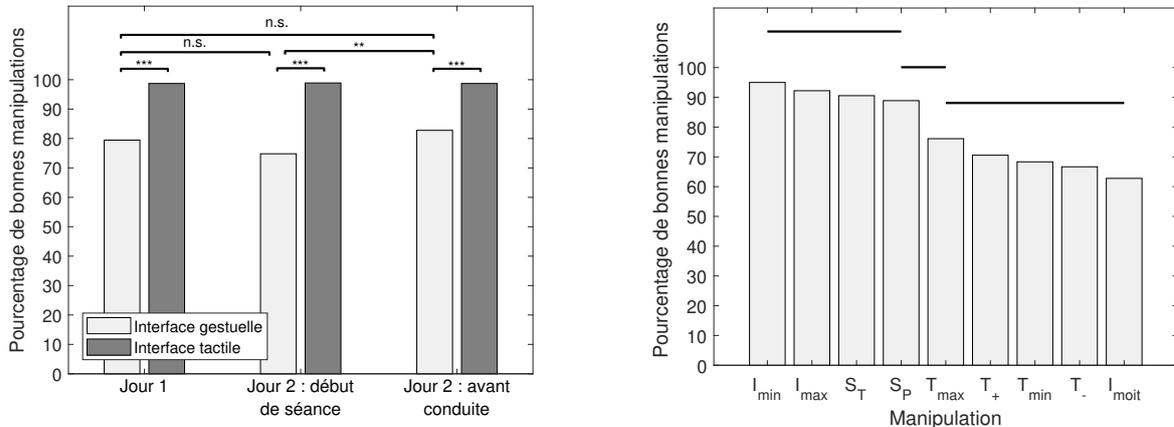
Les manipulations en conduite sont quant à elles données par numéro d'instruction, allant de 1 à 24, comme nous l'avons présenté sur la figure 6.6.

Manipulations de l'interface à l'arrêt sans retour d'information

Nous allons tout d'abord nous intéresser à l'une des deux mesures de l'entraînement, qui s'effectue à l'arrêt, hors d'une phase de conduite. Cette première évaluation consiste à demander aux sujets de faire des manipulations de l'interface sans retour d'information, et elle a lieu à trois instants du déroulement de l'expérience : après l'entraînement du premier jour, avant puis après l'entraînement du second jour. À chaque fois, les sujets doivent faire 18 manipulations correspondant à 2 répétitions des 9 manipulations de fonctions proposées avec cette interface.

Les résultats sont présentés en figure 6.7. La figure 6.7a montre les pourcentages de réussite des manipulations effectuées avec l'interface gestuelle ou tactile. Pour comparer ces résultats, nous avons effectué un test de Cochran qui a montré des différences significatives – $Q = 358,090$; $N = 540$; $p < 0,001$. Des tests de McNemar ont été effectués en test post-hoc, en prenant en compte une correction de Bonferroni [Bland et Altman, 1995].

Ces résultats montrent une facilité d'utilisation du tactile par rapport au gestuel, quelque soit l'instant où la mesure a été effectuée. Nous pouvons observer une diminution non significative entre les mesures en jour 1 et en début de jour 2, et une augmentation significative entre les deux mesures du jour 2. Les résultats entre le jour 1 et la fin du jour 2 ne sont en revanche pas significativement différents. L'évolution des performances pour l'interface tactile n'est pas significative, mais les résultats sont proches de 100% dans tous les cas.



(a) Pourcentage de bonnes manipulations de l'interface sans retour d'information.
n.s. = non significatif; ** = $p < 0,01$; *** = $p < 0,001$.

(b) Détail par type de manipulation gestuelle sans retour d'information, moyennes sur les 3 mesures. Les barres indiquent les résultats des manipulations appartenant au même groupe, n'étant pas significativement différents.

FIGURE 6.7 – Résultats des manipulations des interfaces tactile ou gestuelle sans retour d'information (a) à 3 différents instants de l'expérience ou (b) selon le type de manipulation en ce qui concerne l'interface gestuelle.

Concernant l'interface gestuelle, nous avons choisi de détailler les résultats globaux par manipulation, que nous pouvons observer en figure 6.7b. Nous avons effectué un test de Cochran qui a révélé la présence de différences significatives – $Q = 138,794$; $N = 180$; $p < 0,001$.

Plusieurs tests de McNemar accompagnés d'une correction de Bonferroni ont permis de détailler ce résultat, en classant les manipulations en deux groupes principaux, comme on peut le voir sur la figure 6.7b : les manipulations I_{\min} , I_{\max} , S_T et S_P d'un côté, et les manipulations T_{\max} , T_+ , T_{\min} , T_- et I_{moit} de l'autre. Il est important de noter que les différences entre les manipulations S_P et T_{\max} ne sont pas significatives.

Globalement, les manipulations ayant les moins bons résultats concernent les balayages de la main ainsi que le réglage de la ventilation à la moitié. Les balayages semblent donc difficiles à réaliser, même à l'arrêt sans tâche distractive. Les mauvais résultats de la manipulation de l'intensité de ventilation peuvent s'expliquer par l'absence de barrière physique à la réalisation du geste : pour les manipulations I_{\min} ou I_{\max} , les sujets ont simplement à ouvrir ou fermer complètement la main, et les résultats sont respectivement aux alentours de 95% et 92% de réussite. Les performances pour la manipulation I_{moit} semblent donc pâtir de l'absence de cette limite physique.

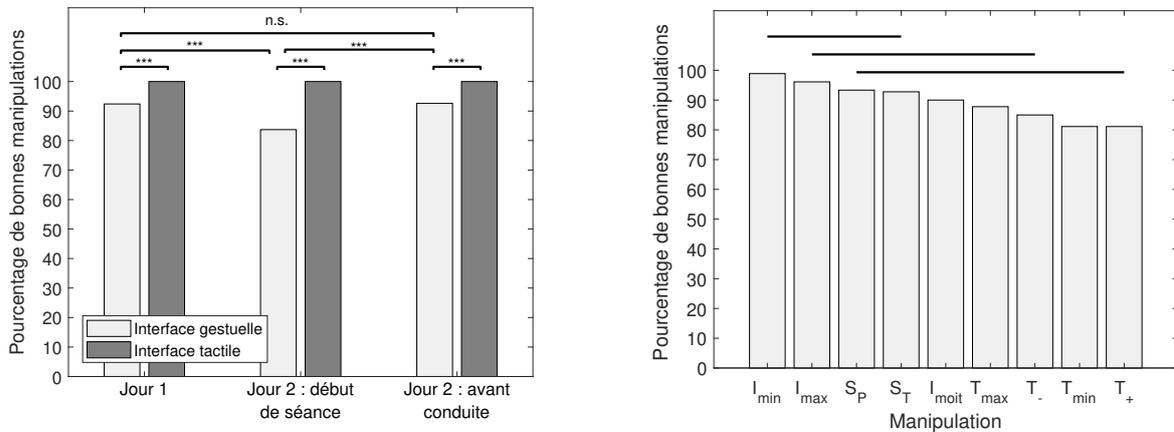
Le groupe des manipulations I_{\min} , I_{\max} , S_T et S_P obtient des résultats prometteurs, avec une moyenne de 91,67%. Ce taux de réussite, même s'il reste plus faible qu'avec l'interface tactile, est globalement satisfaisant. Nous pouvons penser qu'en tirant les leçons de cette expérience, et en approfondissant la reconnaissance de gestes, nous pourrions arriver à des taux de réussite suffisants pour une application industrielle. Le second groupe de manipulation semble en revanche poser problème, et d'importantes modifications devront être apportées pour les conserver.

Manipulations de l'interface à l'arrêt avec retours d'informations

À la suite des manipulations sans retour d'information, nous avons effectué le même type d'évaluation en proposant cette fois des retours d'informations visuelles et/ou sonores. Le déroulement est similaire à la section précédente, à la différence que le nombre de gestes n'est pas limité : ce sont les sujets qui indiquent à l'expérimentateur lorsqu'ils pensent avoir effectué la bonne manipulation. Le but n'est donc plus d'évaluer les gestes, mais d'observer si l'ensemble gestes – manipulations de l'OV – sons est correctement compris.

La figure 6.8 présente les différents résultats obtenus, où un détail par jour et par type d'interface est proposé en figure 6.8a. De la même manière que dans le précédent paragraphe, un test de Cochran a révélé des différences significatives – $Q = 243,768$; $N = 540$; $p < 0,001$, et des tests post-hoc de McNemar effectués avec une correction de Bonferroni ont permis de détailler les différences entre chaque condition. Une nouvelle fois, la condition avec interface tactile obtient des résultats significativement plus élevés qu'avec l'interface gestuelle. En se focalisant sur les manipulations de l'interface gestuelle, les résultats baissent significativement entre la fin du jour 1 et le début du jour 2, et augmentent significativement entre le début et la fin du jour 2. Aucune amélioration n'est en revanche constatée entre la fin du jour 1 et la fin du jour 2. Concernant l'interface tactile, aucune évolution n'a été constatée, le taux de réussite étant à 100%.

Nous avons également détaillé les résultats par manipulation de l'interface gestuelle en figure 6.8b. Les tests de Cochran puis McNemar ont révélé des différences significatives – $Q = 66,482$; $N = 180$; $p < 0,001$. Globalement, les différences sont moins nettes que pour la figure 6.7b, avec de meilleurs taux de réussite. Les résultats ne sont pour autant pas tout à fait satisfaisants : les retours sonores sur les réglages ne sont pas bien compris par l'ensemble des sujets. Les différences entre les figures 6.7b et 6.8b ne peuvent cependant pas découler uniquement des retours d'informations, car les sujets ne pouvaient effectuer qu'un seul geste dans la précédente phase.



(a) Pourcentage de bonnes manipulations de l'interface avec retours d'informations.

n.s. = non significatif; ** = $p < 0,01$; *** = $p < 0,001$.

(b) Détail par type de manipulation gestuelle avec retours d'informations, moyennes sur les 3 mesures. Les barres indiquent les résultats des manipulations appartenant au même groupe, n'étant pas significativement différents.

FIGURE 6.8 – Résultats des manipulations des interfaces tactile ou gestuelle avec retours d'informations (a) à 3 différents instants de l'expérience ou (b) selon le type de manipulation en ce qui concerne l'interface gestuelle.

Nous pouvons constater que les manipulations qui ont obtenu les meilleurs performances sans retour d'information sont de nouveau parmi les meilleurs résultats. Une nouvelle fois, les gestes de balayages sont les manipulations avec le plus faible taux de réussite. Il semblerait donc que la difficulté du geste ne soit pas totalement compensée par les retours d'informations et la possibilité de faire plusieurs gestes. Une différence notable entre les figures 6.7b et 6.8b est la relative progression de la manipulation I_{moit} .

Manipulations de l'interface en situation de conduite

Nous allons à présent nous intéresser aux manipulations des interfaces gestuelle et tactile effectuées conjointement à la tâche de conduite. Dans ce paragraphe, nous allons reprendre les numérotations de la figure 6.6, en détaillant les résultats par instruction. Nous avons choisi de ne pas regrouper les résultats selon les 9 manipulations demandées, puisque les conditions de conduite sont à chaque fois différentes. Il est important de noter que certaines manipulations gestuelles ont été menées avec difficulté à cause d'une mauvaise détection de la main de la part du capteur. Ces difficultés ont pu compliquer les tâches de manipulation de l'IHM et de conduite. Dans ces cas, les sujets étaient invités par l'expérimentateur à bouger la main au-dessus du capteur pour tenter une nouvelle fois les gestes de manipulation de l'OV.

Les résultats présentés en figure 6.9 donnent les performances des sujets en fonction des instructions et du type d'interface manipulée. Sur cette figure, un essai est considéré comme correct si le bon réglage est effectué même après plusieurs tentatives. Les numéros d'instruction sont donnés en abscisse, ainsi que les manipulations sur la seconde ligne. Nous avons effectué des tests du χ^2 pour déterminer si les différences de performance avec les interfaces gestuelle et tactile sont significatives ou non. Les résultats significatifs sont indiqués au-dessus des barres de la figure 6.9.

Les résultats montrent que, pour une majorité de cas, les différences entre les deux types d'interface ne sont pas significatives. Les instructions I05, I07, I11, I17, I20 et I23 mènent cependant à des performances significativement plus élevées avec l'interface tactile qu'avec l'interface gestuelle. Nous avons conscience qu'une expérience menée avec plus de sujets aurait pu donner des résultats encore plus contrastés. Nous pouvons cependant observer que les 6 instructions ayant mené à des différences significatives concernent les balayages de la main en ce qui concerne l'interface gestuelle : les manipulations T_+ , T_- et T_{\min} sont en cause par 2 fois chacune.

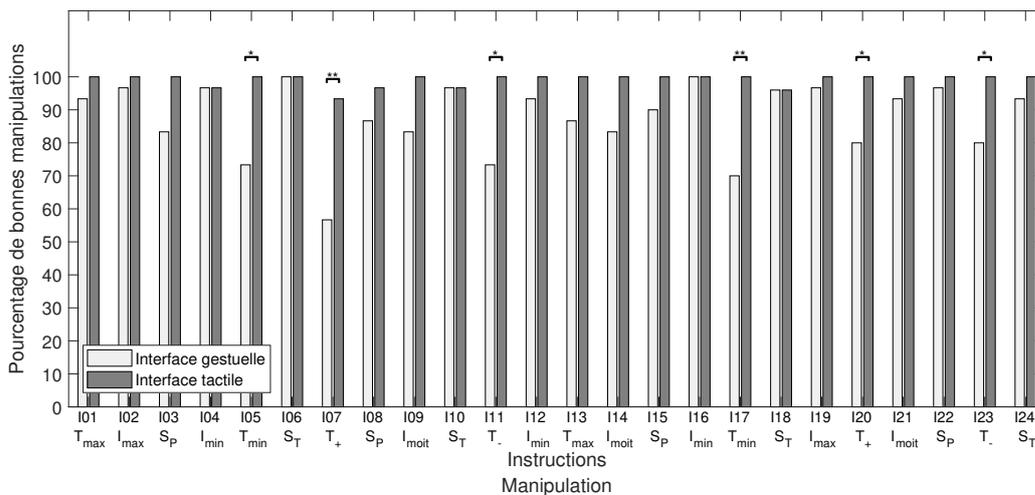


FIGURE 6.9 – Résultats des manipulations des interfaces tactile et gestuelle en conduite. Une manipulation est jugée correcte si le bon réglage est effectué même après plusieurs essais. En abscisse de la figure, les instructions I_n sont notées sur la première ligne, où n représente le numéro de l'instruction. La notation correspondant à la manipulation est donnée sur la seconde ligne.

Nous avons également choisi de présenter des résultats voisins, s'intéressant cette fois au taux de réussite des manipulations dès le premier essai. La figure 6.10 détaille les performances suivant les instructions et les interfaces gestuelle et tactile. Le même type de notation est utilisé que pour la figure 6.9. De la même manière que précédemment, des tests du χ^2 ont permis de déterminer si les différences entre les performances des deux types d'interface sont significatives.

Sur les 24 instructions proposées, 5 conduisent à des différences significatives entre les interfaces, avec de meilleurs résultats pour l'IHM tactile. Une nouvelle fois, les manipulations gestuelles ayant posé problème concernent les balayages de la main, avec les manipulations T_{\max} , T_+ , et T_{\min} . Les différences avec la figure précédente se situent aux instructions I11, I13 et I23.

Sur les deux figures, les différences significatives proviennent du type d'interface, mais la tâche de conduite peut également interférer avec les manipulations des IHM. Nous pourrions par exemple penser que les différences de performances pour l'instruction I17 proviennent de la combinaison des tâches de manipulations d'interface et de conduite, à cause de l'enchaînement de virages par exemple. Cependant, les différences se produisant uniquement pour des gestes de balayages, nous pouvons penser que c'est plutôt le geste qui pose problème, comme le montre également la figure 6.7b.

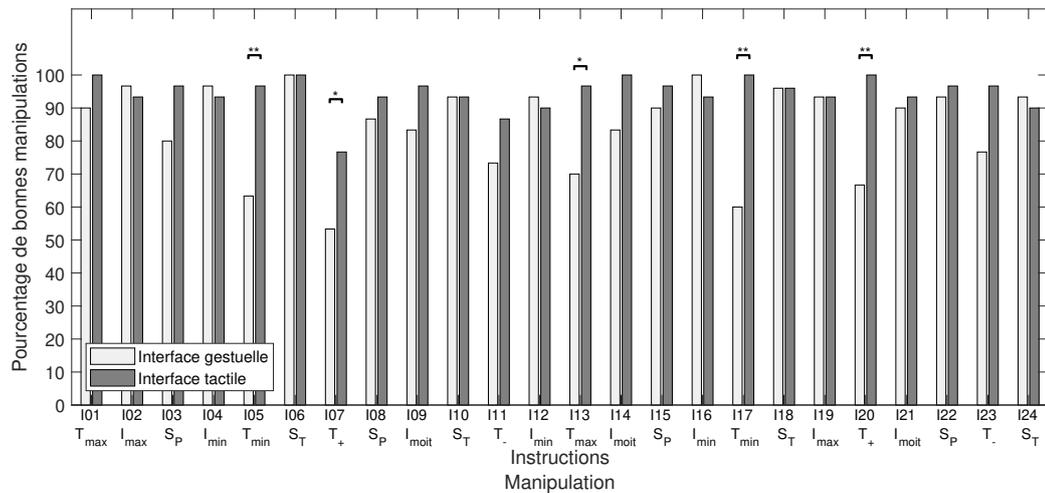


FIGURE 6.10 – Résultats des manipulations des interfaces tactile ou gestuelle en conduite. Une manipulation est jugée correcte si le bon réglage est effectué dès le premier essai. En abscisse de la figure, les instructions I_n sont notées sur la première ligne, où n représente le numéro de l'instruction. La notation correspondant à la manipulation est donnée sur la seconde ligne.

Données de conduite

Nous avons enregistré plusieurs paramètres de conduite pendant l'utilisation des deux types d'interfaces sur le simulateur : l'angle au volant, les enfoncements des pédales d'accélérateur et de frein, l'écart latéral, la vitesse et l'accélération. Nous avons voulu observer d'éventuelles différences sur ces paramètres entre les deux types d'interface. Nous avons donc fait la moyenne de ces paramètres sur les 24 tronçons¹ correspondant aux 24 instructions.

Comme exposé précédemment, les mains des sujets étaient parfois mal captées par le capteur, et plusieurs tentatives de gestes pouvaient être nécessaires pour effectuer le réglage de l'interface demandé. Cette demande additionnelle a pu perturber les sujets, mais également leur conduite.

Nous avons tout d'abord effectué des analyses en composantes principales pour chaque paramètre, mais aucune d'entre elles n'a révélé de groupes de sujets. Nous avons donc effectué des analyses de variance sur l'ensemble des sujets, avec un facteur sujet – 29 ddl, un facteur du type de conduite du véhicule livreur – 1 ddl, un facteur du type d'interface – gestuelle et sonore ou tactile, 1 ddl, un facteur du numéro d'instruction proposé imbriqué dans le facteur du type de conduite du véhicule livreur – 22 ddl, et des facteurs d'interaction entre le facteur sujet et le facteur du type d'interface – 29 ddl, entre le facteur du type d'interface et celui du numéro d'instruction proposé imbriqué dans le facteur du type de conduite du véhicule livreur – 22 ddl, et entre ce dernier facteur et le facteur du type de conduite du véhicule livreur – 1 ddl. Les résultats sont détaillés sur le tableau 6.2. Ces résultats montrent qu'il n'existe pas de différences significatives pour ces paramètres entre les interfaces tactiles et gestuelles. Nous avons également effectué les calculs sur l'écart-type des paramètres de conduite sur chacun des tronçons, et une nouvelle fois aucune différence n'est à relever.

1. Le tronçon n est délimité par la portion de route entre les instructions I_n et I_{n+1} .

Paramètres	Moyenne avec IHM gestuelle	Moyenne avec IHM tactile	p-value du facteur <i>type d'interface</i>
Angle au volant (en rad)	$1,90 \times 10^{-2}$	$1,78 \times 10^{-2}$	0,209
Enfoncement de l'accélérateur (en %)	$2,92 \times 10^1$	$2,89 \times 10^1$	0,534
Enfoncement du frein (en %)	$3,65 \times 10^{-2}$	$3,56 \times 10^{-2}$	0,478
Écart latéral (en m)	$-1,30 \times 10^{-3}$	$1,37 \times 10^{-2}$	0,136
Vitesse (en m.s ⁻¹)	63,3	63,1	0,263
Accélération (en m.s ⁻²)	$-4,22 \times 10^{-2}$	$-4,54 \times 10^{-2}$	0,676

TABLEAU 6.2 – Moyennes des paramètres de conduite sur les 24 tronçons pour les interfaces gestuelles ou tactiles. La p-value du facteur *type d'interface* est également donnée.

Il semble donc que les distractions visuelles de l'interface tactile et cognitives de l'interface gestuelle ont eu la même influence sur ces paramètres de conduite. Nous pouvons alors nous interroger : est-ce que les interfaces n'ont pas perturbé les sujets pendant leur conduite, ou les ont-elles perturbées de la même façon ? Cette question aurait pu être tranchée par un enregistrement d'une phase de conduite sans manipulation d'interface. Cependant, le protocole de l'expérience étant déjà conséquent, nous n'avons pas choisi d'ajouter cette condition, qui aurait, de plus, pu favoriser un phénomène de fatigue chez les sujets.

Nous allons maintenant voir que le type d'interface a bien joué un rôle sur la conduite de sujets, en particulier sur la tâche symbolisant le trafic routier.

Suivi du véhicule lièvre

Durant la phase de conduite, les sujets doivent maîtriser leur véhicule, suivre les instructions données concernant la manipulation de l'interface, mais également suivre le véhicule lièvre en respectant une consigne de distance que nous avons proposée. Cette tâche a pour but de solliciter visuellement les sujets, pour simuler la demande visuelle d'attention sur le trafic routier. Pour rappel, la vision tête haute se colore si la distance avec le lièvre est supérieure à 40 mètres.

Nous avons calculé la valeur moyenne de la distance au-delà de la limite de 40 mètres sur chacun des 24 tronçons correspondant aux 24 instructions. L'ACP que nous avons réalisé n'a pas révélé de groupes de sujets ayant des comportements différents. Nous avons donc réalisé une analyse de variance avec les mêmes facteurs que le paragraphe précédent, qui révèle cette fois-ci une différence significative du facteur *type d'interface* – $F(1,1320) = 12,365$; $p < 0,001$. La figure 6.11 montre le détail des résultats selon les tronçons et les deux interfaces. Nous pouvons voir que la distance entre les sujets et le véhicule lièvre est plus faible dans le cas de l'interface gestuelle, où les conducteurs respectent mieux la consigne de distance que dans la configuration avec l'interface tactile. Cependant, il est à noter que la consigne n'est bien respectée pour aucune des deux interfaces.

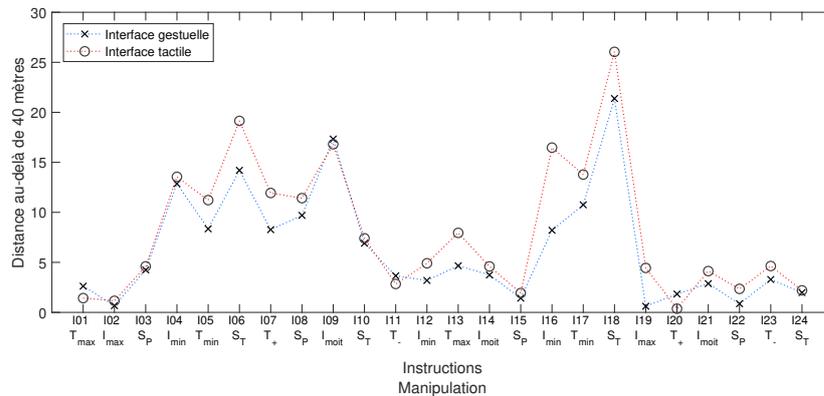


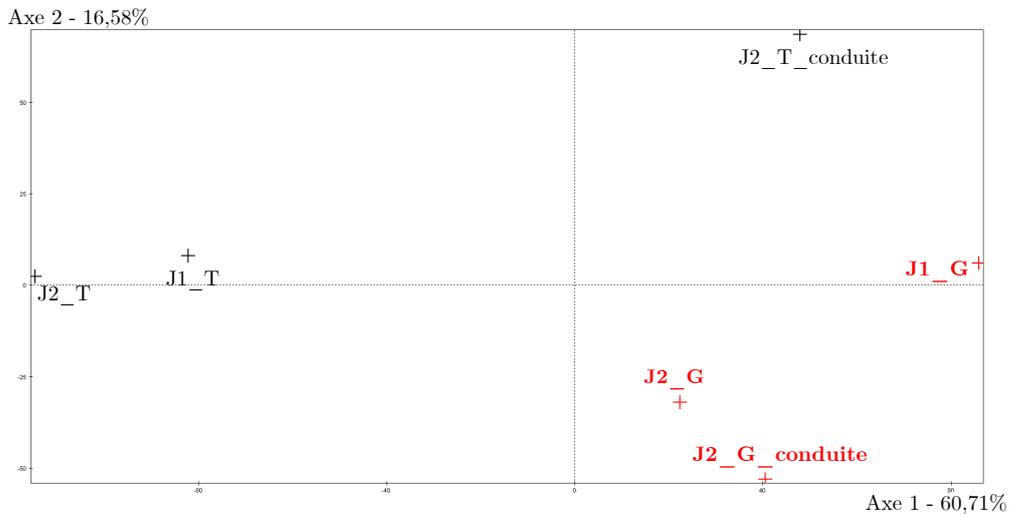
FIGURE 6.11 – Distance au-delà de la limite de 40 mètres par type d'interface.

Mesure de la charge cognitive

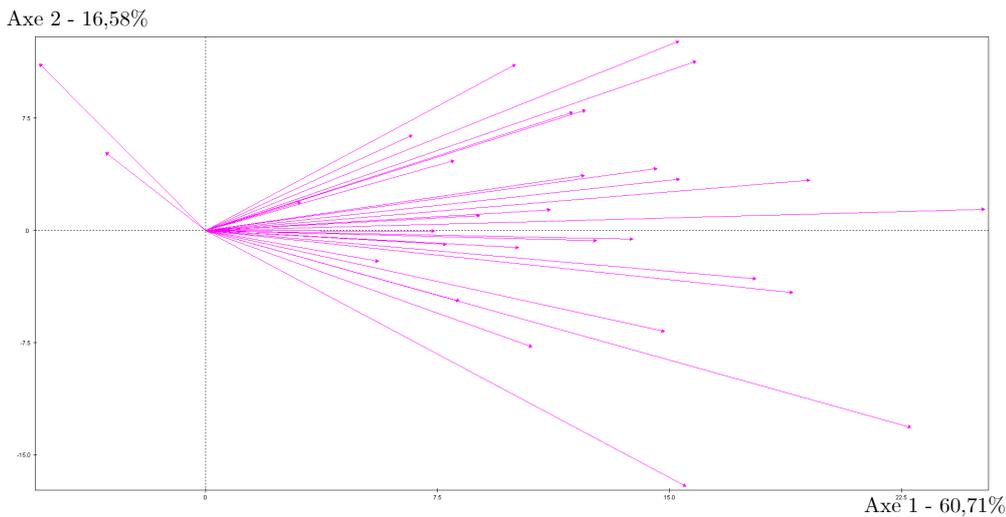
À différents moments de l'expérience, nous avons demandé aux sujets d'estimer leur charge cognitive avec l'échelle RSME – pour *Rating Scale of Mental Effort* – que nous avons évoqué au chapitre 3. L'échelle que nous avons utilisée est visible en figure E.1 en annexe. L'estimation est réalisée juste après les manipulations des IHM, pour avoir une représentation la plus précise possible de la concentration mentale des sujets. Il nous a semblé intéressant d'avoir cette mesure après les entraînements à l'arrêt des jours 1 et 2, puis après la conduite du jour 2 pour observer l'évolution suivant le jour et la présence ou non de tâche de conduite.

Une ACP a été réalisée pour observer la répartition des sujets suivant le score de RSME mesuré à différents instants. Les résultats de l'ACP sont visibles sur la figure 6.11. La première observation concerne les axes des analyses en composantes principales, qui représentent 53,32% et 17,66% de la variabilité pour les axes 1 et 2 respectivement : l'interprétation doit donc se concentrer en premier lieu sur l'axe horizontal. La figure 6.12a montre les résultats suivant les différentes conditions de l'expérience. La notation J1 ou J2 indique les mesures effectuées les jours 1 et 2 respectivement, pour les interfaces gestuelles ou tactiles – notées G ou T. Nous avons précisé la mesure effectuée après la conduite du jour 2 par la notation *conduite*. Enfin, pour plus de lisibilité, les points concernant l'interface gestuelle sont notés en rouge et en gras.

Nous pouvons observer que les points de l'IHM tactile hors situation de conduite sont plutôt sur la gauche de la figure, alors que ceux de l'interface gestuelle se situent sur la droite. Nous pouvons alors supposer des différences significatives entre les deux types d'interface pour les conditions à l'arrêt. En revanche, les scores obtenus en conduite pour les deux types d'interfaces sont proches, témoignant de potentielles faibles différences de RSME. Sur la figure 6.11b, chaque flèche représente le score de RSME d'un sujet. Nous pouvons observer un faisceau vers la droite, témoignant d'un consensus entre les sujets, malgré deux flèches pointant vers la gauche. Le comportement de ces deux sujets est jugé anecdotique, et nous avons choisi de faire une analyse de variance sur l'ensemble des sujets.



(a) Évolution du score de RSME selon les conditions.



(b) Évolution du score de RSME par sujet.

FIGURE 6.11 – ACP réalisée sur les scores de RSME.

Les résultats de l'ANOVA sont présentés en figure 6.12 et cette analyse a été effectuée avec un facteur sujet – 29 ddl, un facteur du type d'interface utilisée – 1 ddl, un facteur lié au moment de la mesure dans l'avancement de l'expérience – 2 ddl, et un facteur d'interaction entre ces deux derniers – 2 ddl. Les différences significatives sont indiquées par les barres au-dessus des scores de RSME. Nous pouvons observer que la concentration mentale des sujets est jugée plus faible lors de l'utilisation de l'interface tactile pour les conditions à l'arrêt des jours 1 et 2. Une diminution significative peut être notée pour les deux types d'interface entre les deux jours. Après avoir effectué les manipulations des interfaces en conduite, les sujets jugent en revanche une concentration équivalente entre les interfaces tactile et gestuelle. Les scores de RSME sont en effet restés constant entre les conditions à l'arrêt et en conduite pour l'interface gestuelle, alors qu'une augmentation significative est à noter pour l'interface tactile.

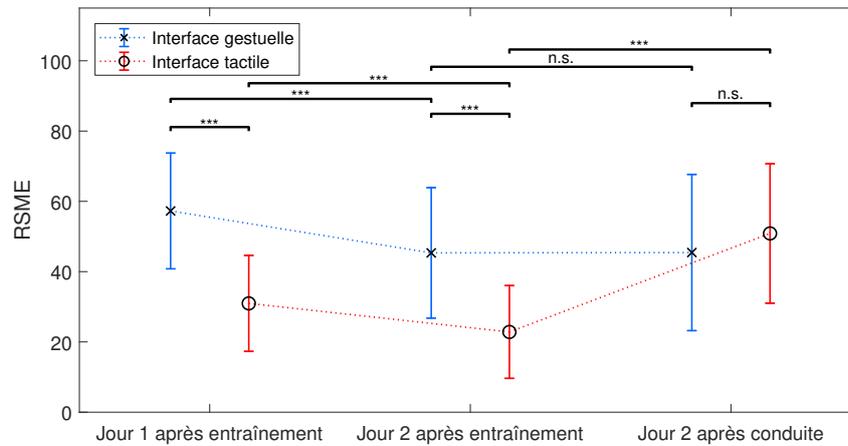


FIGURE 6.12 – Mesures de la charge cognitive à différents moments de l'expérience. La moyenne et l'écart type des scores de RSME sont indiqués. n.s. = non significatif; *** = $p < 0,001$.

Maîtrise de l'interface

Pour essayer d'évaluer d'une manière différente la concentration des sujets, nous avons proposé aux sujets de répondre, immédiatement après l'utilisation de l'échelle de RSME, à la question suivante :

« Avez-vous confiance en votre maîtrise de l'interface ? »

Nous voulions ici interroger les sujets sur leur confiance en leur bonne utilisation de l'IHM. Pour répondre, les sujets doivent cocher une case de l'échelle de Likert en 7 points labellisés allant de 1 pour *pas confiant* à 7 pour *très confiant*. Nous avons tout d'abord effectuée une ACP révélant un comportement homogène sur l'ensemble des sujets – voir la figure F.1 en annexe. Nous avons donc par la suite effectué une ANOVA, dont les résultats sont présentés en figure 6.13. Sur cette figure, nous pouvons observer des résultats pouvant être rapprochés des scores de RSME présentés précédemment. Pour les mesures effectuées à l'arrêt, l'interface gestuelle est significativement moins bien notée que l'interface tactile. Après la phase de conduite, les sujets ont significativement augmenté la note de l'interface gestuelle et baissée celle de l'interface tactile, si bien que la différence entre les deux n'est pas significative.

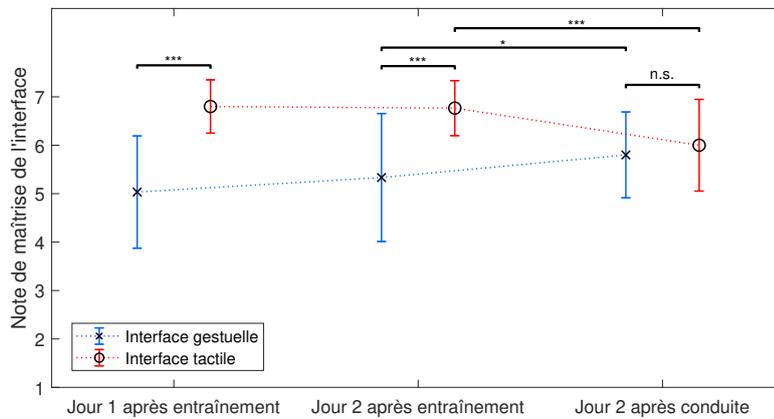


FIGURE 6.13 – Réponses à la question portant sur la confiance des sujets en leur maîtrise de l'interface à différents moments de l'expérience. La moyenne et l'écart type des notes sont indiqués. n.s. = non significatif; * = $p < 0,05$; *** = $p < 0,001$.

Questionnaire concernant l'interface de ventilation

À la suite de l'expérience de conduite avec les interfaces tactile et gestuelle, nous avons demandé aux sujets de remplir un questionnaire comparant les deux interfaces selon différents critères. Pour répondre, nous leur avons fourni une échelle de Likert en 7 points allant de *tactile* à *gestuelle*, permettant d'exprimer une opinion plus ou moins tranchée en faveur de l'une ou l'autre des deux interfaces. La figure 6.14 montre un exemple de réponse donnée par les sujets.

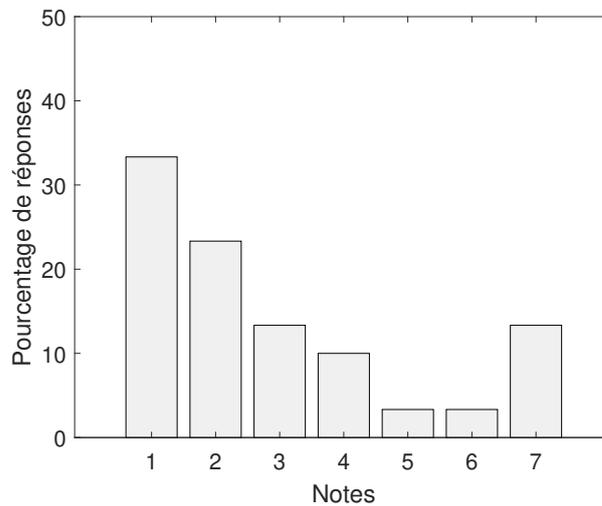


FIGURE 6.14 – Répartition des réponses des sujets à *designez l'interface paraissant la plus facile à utiliser à l'arrêt*. 1 : interface tactile, 7 : interface gestuelle

Pour déterminer si les réponses des sujets sont significativement en faveur de l'une des deux interfaces, nous avons utilisé un test de Student de comparaison à un standard, en comparant la moyenne des réponses obtenues avec une moyenne de 4 – correspondant à des réponses ne privilégiant aucune des deux IHM. Dans un souci de clarté, nous allons présenter les différentes réponses des sujets sous forme d'un tableau récapitulatif, voir tableau 6.3. Le détail des réponses des sujets au questionnaire est disponible en annexe F, figures F.2, F.3, F.4, F.5 et F.6, où les tendances globales peuvent paraître plus nettement qu'avec le tableau 6.3.

Pour l'ensemble des réponses formulées, il est nécessaire de tenir compte d'un biais lié à la bonne volonté des sujets : cette étude portant sur une technologie maîtrisée et une technologie gestuelle nouvelle, les sujets ont pu se douter du but de l'étude, et donc répondre aux questions de manière clémente concernant l'interface gestuelle. En l'état actuel des choses, nous ne pouvons pas éviter ce biais, il est donc nécessaire d'être prudent sur les conclusions pouvant être tirées du questionnaire.

Questionnaire : Désignez la technologie paraissant :	Moyenne des réponses	p-value calculée	Interface privilegiée
- la plus facile à utiliser à l'arrêt	2,90	$7,36 \times 10^{-3}$	Tactile
- la plus facile à utiliser en conduite	5,60	$7,36 \times 10^{-3}$	Gestuelle
- la plus agréable utiliser à l'arrêt	3,47	0,14	-
- la plus agréable utiliser en conduite	5,60	$3,15 \times 10^{-7}$	Gestuelle
- la plus sécuritaire en conduite	6,13	$1,88 \times 10^{-10}$	Gestuelle

TABLEAU 6.3 – Tableau récapitulatif des réponses au questionnaire comparant les interfaces tactile et gestuelle.

Les sujets ont tendance à privilégier l'interface tactile concernant la facilité d'utilisation à l'arrêt, mais préfèrent l'interface gestuelle en conduite. Aucune des deux interfaces n'a semblé plus agréable que l'autre pour une utilisation à l'arrêt, comme en témoigne la moyenne des réponses de 3,47. L'opinion des volontaires de l'expérience est beaucoup plus tranchée en conduite, où l'interface gestuelle est plébiscitée. Les réponses les plus nettes sont obtenues concernant l'aspect sécuritaire en conduite, où l'interface gestuelle est nettement préférée, la moyenne des réponses étant de 6,13.

Questionnaire spécifique à l'interface gestuelle de ventilation

Une partie du questionnaire s'est focalisée plus spécifiquement sur les différents appariements que nous avons mis en place entre les gestes, les fonctions et les sons – voir figure 6.1 – pour l'interface de ventilation. Les questions portaient par exemple sur la facilité d'exécution des gestes, ou la pertinence des associations proposées. Nous avons alors pu comparer les réponses des sujets suivant les trois types de fonctions de défilement, d'intensité et de placement que nous avons utilisés, spécifiquement pour notre interface de ventilation. Pour répondre, les sujets ont utilisé une échelle de Likert en 7 points allant de *non pas du tout* à *oui tout à fait*. Les moyennes des réponses aux questions sont indiquées dans le tableau 6.4, et une analyse de variance a été effectuée pour comparer les moyennes, avec un test post-hoc de Duncan.

Questions :	Fonction de défilement	Fonction d'intensité	Fonction de placement
Avez-vous trouvé que le geste est facile à réaliser pour :	4,00	6,30	6,30
Avez-vous trouvé que le geste est adapté à la fonction pour :	5,07	5,96	6,19
Avez-vous trouvé que les sons sont adaptés aux gestes pour :	4,81	6,41	6,04
Avez-vous trouvé l'utilisation de la modalité gestuelle pertinente pour la fonction :	4,93	6,00	6,04

TABLEAU 6.4 – Tableau récapitulatif des réponses au questionnaire comparant les fonctions de l'interface de ventilation, 1 : *non pas du tout*, 7 : *oui tout à fait*. Les notes indiquées en gras font parti du même groupe, et ne sont donc pas significativement différentes.

Pour l'ensemble des questions, nous pouvons observer des notes relativement élevées pour les fonctions d'intensité et de placement, aux alentours de 6. En revanche, les notes de la fonction de défilement sont significativement plus basses que pour les deux autres types de fonction. L'ANOVA a révélé des différences significatives pour le facteur *fonction* pour chacune des questions.

Il semble que les sujets de l'expérience ont particulièrement rejeté la fonction de défilement, la difficulté à réaliser les balayages de la main pouvant en être la cause, voir figure 6.7b. Nous pensons que cette difficulté de réalisation des gestes a influencé l'ensemble des critères que nous voulions évaluer avec le questionnaire, ce qui peut expliquer des réponses aux quatre questions significativement moins bonnes pour cette fonction.

Questionnaire concernant l'interface gestuelle de gestion de la musique

À la suite des manipulations de l'interface de ventilation, nous avons proposé aux volontaires de l'expérience de manipuler l'interface de gestion de la musique que nous avons construite sur la base de l'interface de ventilation. Un questionnaire a ensuite été proposé, pour recueillir leurs impressions, où il était une nouvelle fois possible de répondre grâce à une échelle de Likert en 7 points allant de *non pas du tout* à *oui tout à fait*. Le tableau 6.5 résume les réponses données par les sujets, un détail est donné en annexe G, figures G.1, G.2, G.3 et G.4. Nous avons également utilisé une ANOVA pour comparer les réponses pour chaque fonction, avec un test post-hoc de Duncan.

Questions :	le changement de musique ?	avancer/reculer dans un morceau ?	placer la source de musique ?	manipuler le volume ?
Avez-vous trouvé pertinente l'utilisation des gestes pour :	6,10	5,26	6,23	6,47

TABLEAU 6.5 – Tableau récapitulatif des réponses au questions concernant l'interface des médias, 1 : *non pas du tout*, 7 : *oui tout à fait*. Les notes indiquées en gras font parti du même groupe, et ne sont donc pas significativement différentes.

L'ensemble des réponses données par les sujets sont positives, la note la plus basse étant de 5,26. Seule la fonction d'avance ou de recul dans un morceau, correspondant à des balayages de la main de faible intensité, a été significativement moins bien notée que les autres fonctions. Il est intéressant de constater que les balayages de la main permettant le changement du morceau ont quant à eux été plutôt appréciés. Il semble que les sujets ont tout particulièrement plébiscité l'utilisation des gestes pour manipuler le volume et la position de la source sonore, les notes correspondant à ces fonctions étant les plus hautes.

Questionnaire général concernant l'interface gestuelle

Le questionnaire de fin d'expérience a également proposé des questions plus générales concernant l'interface gestuelle. Une échelle de Likert en 7 points allant de *non pas du tout* à *oui tout à fait* a de nouveau été utilisée. Le tableau 6.6 donne les réponses des sujets aux questions, et un détail des réponses est donné en annexe H, sur les figures H.1, H.2, H.3, H.4, H.5. Les seuils de significativité calculés par rapport à une réponse neutre sont obtenus une nouvelle fois par un test de Student.

Questions :	Moyenne des réponses	p-value calculée
Avez-vous réussi à imaginer le comportement de l'objet virtuel ?	5,59	$4,51 \times 10^{-05}$
Avez-vous trouvé que le nombre de gestes à retenir est convenable ?	6,19	$2,51 \times 10^{-10}$
La combinaison du tactile et des gestes vous paraît-elle intéressante ?	6,07	$4,41 \times 10^{-11}$
Seriez-vous intéressé-e pour avoir ce genre de technologie dans votre véhicule pour la gestion de la ventilation ?	5,60	$4,23 \times 10^{-8}$
Seriez-vous intéressé-e pour avoir ce genre de technologie dans votre véhicule pour la gestion des médias ?	6,07	$5,24 \times 10^{-12}$

TABLEAU 6.6 – Tableau récapitulatif des réponses au questionnaire concernant l'interface gestuelle, 1 : *non pas du tout*, 7 : *oui tout à fait*.

La métaphore de l'objet virtuel a été bien imaginée par les sujets, avec une moyenne de 5,59 : il semble donc que nos stratégies de sonification ont été utiles de ce point de vue. Même si le nombre de gestes à retenir pour manipuler les trois fonctions pouvait paraître important – allant de 7 à 10 selon la manière de compter, les sujets n'ont pas paru gênés par cet aspect, avec un nombre de gestes à retenir perçu comme convenable, avec une moyenne de 6,19. La combinaison des gestes et du tactile semble bien envisagée par les sujets, avec une moyenne de 6,07. Enfin, les sujets paraissent intéressés pour avoir ce type de technologie dans leur véhicule malgré les difficultés à utiliser le balayage de la main, avec des notes de 5,60 et 6,07 respectivement pour l'interface de ventilation et l'interface des médias.

Manipulations de l'interface gestuelle sur plusieurs séances

Nous avons prolongé l'expérience avec 3 sujets – dont deux encadrants de cette thèse travaillant pour le Groupe PSA et l'auteur de ce document – pour explorer l'évolution des performances sur plusieurs séances de conduite. Nous avons proposé le même protocole que lors de la précédente expérience, mais en effectuant plusieurs fois l'étape de conduite sur simulateur uniquement avec les manipulations gestuelles de l'interface. Les séances étaient espacées de plusieurs jours, et le parcours de conduite était identique. Les résultats des performances de manipulations de ces utilisateurs *experts* sont données sur la figure 6.15.

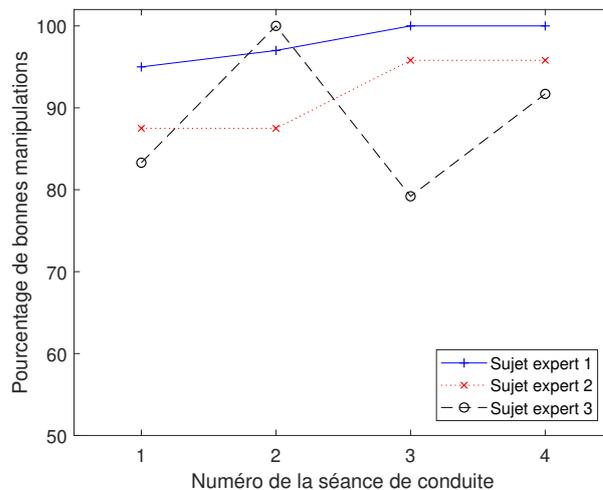


FIGURE 6.15 – Performances d'utilisation de l'IHM gestuelle sur plusieurs séances de conduite pour 3 sujets.

Ces résultats, issus d'un panel de 3 sujets, ne peuvent en aucun cas permettre de tirer des conclusions. Il est cependant possible de discuter des évolutions de l'utilisation de l'IHM sur le long terme à la lumière de ces mesures. Cette figure montre une amélioration globale des performances, même si les résultats du sujet expert 3 sont chaotiques. Les 2 autres sujets experts s'améliorent au fil des séances. Ces progressions peuvent provenir de la connaissance du scénario de conduite, mais également de l'habitude d'utilisation de l'interface. Il serait donc intéressant de varier les parcours de conduite, pour supprimer le biais lié au circuit.

6.3.4 Discussion

Dans cette expérience, nous avons voulu tester l'utilisation de l'interface gestuelle que nous avons mis au point en situation de conduite, sur un simulateur. Pour évaluer son potentiel, une comparaison avec une interface tactile a été mise en place, ce type d'interface étant présent dans de nombreux véhicules récents. Nous avons mis au point un scénario de conduite où des manipulations de l'IHM sont dictées à 24 endroits différents du parcours. Il est ici essentiel de noter que l'utilisation d'une véritable interface dans la vie courante n'est pas la conséquence d'instructions dictées. Elle répond plutôt à une perception de la température de l'habitacle par l'utilisateur, qui a ensuite la volonté de la réguler. Une limite de cette expérience est donc que les manipulations des utilisateurs sont la réponse à une injonction extérieure, qui demande de plus une certaine précision. Dans une utilis-

tion normale, cette précision est déterminée par l'utilisateur qui peut savoir par exemple à quelle intensité il veut régler la ventilation grâce à des stimuli perceptifs sonores ou tactiles – provenant du flux d'air, dans ce cas. Le simulateur de conduite permet de mettre les sujets dans une situation attentionnelle similaire à celle d'un véhicule réel [Santos *et al.*, 2005], même si la perception du danger ou les interactions avec le trafic routier sont différents. Par exemple, la tâche de suivre un véhicule lièvre à une certaine distance que nous avons choisie permet de refléter une sollicitation visuelle due au trafic routier présent en conduite réelle, même s'il est plutôt déconseillé de suivre un véhicule de manière trop rapprochée. Nous pouvons donc observer certaines tendances, mais il est nécessaire de rester prudent quant à la transposition de nos résultats à une situation réelle de conduite.

Nous allons discuter les résultats en évoquant tout d'abord ceux concernant la tâche de manipulation de l'interface, avant de passer à la tâche de conduite. Nous prendrons ensuite du recul pour essayer de décrypter l'ensemble des résultats.

Tâche de manipulation de l'interface

Lors de l'entraînement à l'arrêt, plusieurs manipulations ont été demandées pour estimer la difficulté des sujets à réaliser les différentes manipulations. Les figures 6.7a et 6.7b montrent respectivement les résultats et le détail des manipulations gestuelles. La première figure montre d'un côté l'expertise concernant l'IHM tactile des sujets développée dans leur vie courante, au travers de l'utilisation de smartphone ou tablettes; et de l'autre une difficulté à accomplir les manipulations gestuelles. Plus précisément, les balayages de la main semblent difficiles à effectuer, ainsi que le réglage de la ventilation à la moitié de la jauge. Nous demandons cependant des balayages de la main de plusieurs intensités, cet objectif peut paraître ambitieux et une remise en question de cet aspect peut permettre une amélioration des résultats. Les autres gestes semblent en revanche plutôt maîtrisés, et nous pouvons penser que leur facilité d'exécution augmentera avec l'habitude du maniement de l'IHM. Les figures 6.8a et 6.8b montrent les performances de manipulations des interfaces avec retours informationnels. Ces mesures visent à déterminer si les sujets saisissent bien le fonctionnement de l'IHM, et notamment les retours d'informations leur permettant de faire les bons réglages. Les résultats montrent toujours des différences significatives entre les interfaces gestuelle et tactile, avec cette fois des différences importantes selon l'instant de la mesure pour l'interface gestuelle. Il semble donc que les sujets ont eu plus de difficultés à interpréter les retours au début de la deuxième séance qu'à la fin de la première séance, et l'entraînement à l'arrêt du deuxième jour leur a permis de retrouver le niveau de performance de la veille. En détaillant les performances de l'interface gestuelle par manipulation, nous pouvons constater que les balayages de la main restent les gestes les plus difficiles à réaliser de l'interface. Même si les sujets ont pu reprendre plusieurs fois leurs gestes, ces manipulations sont les moins bien réussies : nous pouvons penser que certains sujets n'ont pas compris les retours sonores, ou ont délaissé ces manipulations à cause de leur difficulté.

Nous pouvons faire le parallèle entre ces observations et les résultats obtenus sur les figures 6.9 et 6.10 : les manipulations ayant conduit à des différences significatives entre les résultats des interfaces tactile et gestuelle concernent dans tous les cas des manipulations de la température, correspondant à des balayages de la main pour l'interface gestuelle. Les difficultés éprouvées à l'arrêt se sont donc parfois retrouvées lors des manipulations en conduite. L'ensemble de ces observations nous pousse à penser que les balayages de la main ne sont pas adaptés à la manipulation de l'interface gestuelle dans leur forme actuelle, car ils sont trop difficiles à réaliser. Des améliorations peuvent

être apportées en définissant par exemple une limite entre les zones personnalisée à chaque sujet, ou en proposant une unique zone pour un seul type de balayage.

De la même manière, nous pouvons observer que les manipulations I_{\min} , I_{\max} , S_T et S_P , faisant parti des meilleures manipulations gestuelles indiquées par la figure 6.7b, ont obtenu de bonnes performances en conduite. Plusieurs de ces manipulations ont obtenu des taux de réussite égaux à l'interface tactile, voir figure 6.9. De plus, certaines ont conduit à de meilleurs résultats si nous observons les performances à la première tentative sur la figure 6.9 – les instructions I02, I04, I12, I16 ou I24, même si les différences ne sont pas significatives. La manipulation de l'intensité de ventilation à la moitié semble être un peu plus difficile que les manipulations que nous venons d'évoquer, sans toutefois poser autant de difficultés que les balayages de la main. Les difficultés de cette manipulation ne se retrouveront probablement pas lors d'une véritable utilisation, puisque l'utilisateur pourra lui-même définir la précision qu'il désire. Nous pouvons donc penser que les gestes liés aux fonctions d'intensité et de placement sont adaptées à une interface gestuelle utilisée en conduite, comme le montre les réponses au questionnaire de fin d'expérience données dans le tableau 6.4. Ces fonctions peuvent donc être conservées dans la suite des études portant sur cette IHM gestuelle.

Tâche de maîtrise du véhicule et de suivi du véhicule lièvre

Lors des passages sur le simulateur de conduite, les sujets ont semble-t-il maîtrisé le véhicule de la même façon pour l'interface gestuelle ou tactile, comme le montre le tableau 6.2. Ces résultats sont cependant à prendre avec précautions, les données étant bruitées et très variables d'un individu à l'autre. Des études comparant des distractions portant sur les différences de distractions visuelles et cognitives que nous avons abordées au chapitre 3 ont révélé que certains paramètres pouvaient être différents entre les deux conditions [Engström *et al.*, 2005] [Hamish Jamson et Merat, 2005] [Lamble *et al.*, 1999]. Dans nos résultats, il n'existe pas de différences significatives entre les vitesses des deux conditions, ce qui peut être expliqué par la présence du véhicule lièvre ainsi que le tracé du parcours. Nous avons en effet demandé aux sujets de suivre un véhicule pour proposer une tâche de conduite demandant une concentration soutenue, comme dans des conditions de conduite normale dans un trafic classique. De plus, les éventuelles différences de vitesse entre les deux IHM ont pu être masquées par la présence d'arrêts à certains carrefours – entre les instructions I8 et I9, et I17 et I18. De plus, aucune différence significative n'a été détectée concernant l'écart type de l'écart latéral entre les deux interfaces gestuelle ou tactile. Enfin, les temps-avant-collision et les temps de réaction au freinage que nous avons également vu au chapitre 3 ne nous ont pas semblé pertinent, même si le véhicule lièvre ralentissait parfois brusquement : ces freinages ont parfois été utilisés par les sujets pour réduire leur distance par rapport au lièvre, qui était trop grande avant le freinage. Ils peuvent donc avoir des temps-avant-collision et des temps de réaction au freinage convenables, tout en ne respectant pas l'exigence de distance par rapport au véhicule lièvre.

Malgré l'absence de différences dans les paramètres de conduite, nous pouvons cependant proposer plusieurs hypothèses expliquant cette absence de différence entre ces paramètres de conduite. Nous pouvons tout d'abord penser que les sujets n'ont pas été perturbés par les distractions visuelle ou cognitive des deux interfaces. Même si cette supposition est peu probable, il aurait été nécessaire de proposer une condition sans manipulations d'interface pour la vérifier. Nous pensons plutôt que les perturbations introduites par les deux interfaces ont été similaires. Lors des manipulations de l'interface tactile, la modalité visuelle est fortement sollicitée ce qui peut perturber la conduite des utilisateurs. Avec l'interface gestuelle, les problèmes de captation de la main ont pu gêner les conducteurs et donc influencer leur conduite. De plus, la difficulté à réaliser les balayages de la main

demandés – et donc la multiplication des tentatives – a pu être un second facteur perturbateur. Nous pouvons alors nous interroger sur les paramètres de conduite que nous aurions pu obtenir dans une étude avec une captation fiable de la main des sujets et avec des balayages plus faciles à effectuer.

La distance avec le véhicule lièvre comptée au-delà de la limite autorisée conduit à des différences significatives entre les interfaces tactile et gestuelle, malgré les arrêts aux carrefours permettant de rattraper le véhicule en cas de distance trop importante. Plus spécifiquement, la figure 6.11 montre que l'IHM gestuelle entraîne une distance plus faible que l'interface tactile sur la presque totalité des instructions. En particulier, une différence importante entre les deux interfaces est à noter pour les instructions I05 à I08, I12 à I14, et I16 à I19. La première série d'instructions correspond à un enchaînement de petits virages, comme nous pouvons le voir sur la figure 6.6. Nous pouvons alors penser que les conducteurs ont eu des difficultés à effectuer les manipulations de l'interface tactile conjointes à un changement de direction du véhicule, demandant également de l'attention visuelle. Les manipulations ont pu être assurées, mais la consigne de distance n'a pu être respectée. La deuxième et la troisième série d'instructions correspondent quant à elles à des comportements particuliers du véhicule lièvre, avec une accélération entre les instructions I12 et I13, et des enchaînements d'un freinage et d'une accélération. Nous pouvons également penser que ces différences de distance proviennent du cumul d'informations provenant du visuel dans le cas de l'interface tactile. Même si les manipulations de l'IHM et la maîtrise du véhicule sont assurées, les sujets n'ont pas réussi à assurer le suivi du véhicule. En particulier, l'instruction I16 suit un freinage du véhicule lièvre – voir freinage 1 en figure 6.6, nous pourrions alors penser qu'il est aisé de maintenir la distance voulue avec le véhicule. Cependant une manipulation de l'interface est demandée à la suite de ce freinage, et nous pouvons alors envisager que l'attention visuelle dans le cas tactile a été concentrée sur cette tâche, délaissant le suivi du véhicule lièvre. L'instruction I16 mène d'ailleurs aux seules différences significatives selon le facteur *type d'interface* concernant la distance détaillée par instruction – $F(1,29) = 7,061$; $p = 0,013$. Dans le cas gestuel, il est possible que l'attention soit divisée efficacement entre les deux tâches grâce aux sollicitations multimodales, comme l'ont constaté plusieurs études que nous avons abordé [Penney, 1989] [Mousavi *et al.*, 1995] [Tindall-Ford *et al.*, 1997] [Rosati *et al.*, 2012]. Ainsi, les performances de manipulation de l'interface ont pu être de 100% pour les deux types d'interface, et la tâche de suivi a pu être significativement mieux respectée avec l'IHM gestuelle.

Selon notre interprétation, la différence significative de la distance au lièvre selon le facteur du type d'interface suggère une meilleure répartition de l'attention des conducteurs entre les tâches de maîtrise du véhicule, de manipulation de l'interface, et de suivi du véhicule. Nous pensons que ces différences d'attention n'ont pas pu se traduire sur des paramètres de conduite comme dans des études que nous avons évoqué [Engström *et al.*, 2005] [Hamish Jamson et Merat, 2005] [Lamble *et al.*, 1999], mais sur la tâche de suivi du véhicule lièvre. De plus, les difficultés de manipulations de l'interface gestuelle ont pu perturber les conducteurs dans leur conduite, comme nous l'avons évoqué. Les résultats présentés précédemment portant sur la concentration des sujets vont permettre de nous renseigner sur l'évolution de la charge mentale des sujets au cours de l'expérience. Ces évaluations subjectives vont être à mettre en regard des résultats que nous venons de discuter.

Charge mentale lors de l'expérience

Les résultats présentés en figures 6.12 et 6.13 nous renseignent sur l'évaluation subjective des sujets concernant leur concentration et leur maîtrise de l'interface respectivement. Ces mesures ont été effectuées à plusieurs moments de l'expérience : à la fin de l'entraînement du jour 1, le jour 2 à la suite de l'entraînement à l'arrêt, et à la suite de la conduite du jour 2. Les résultats des deux figures montrent que l'interface tactile demande moins de concentration et entraîne une plus grande maîtrise que l'interface gestuelle à l'arrêt. En particulier, l'écart type des réponses semble plus faible dans le cas tactile, témoignant du relatif consensus entre les sujets. Nous pensons que cet effet provient notamment de l'expertise des utilisateurs dans la manipulation de ce type d'interface – même si cette expertise est variable d'une personne à l'autre : ils ont pu développer les schémas cognitifs relatifs à l'utilisation d'une IHM tactile. Cette hypothèse est appuyée par les réponses au questionnaire présentées en figures 6.14 et F.4, montrant que les sujets ont tendance à privilégier le tactile à l'arrêt. D'autre part, la facilité et la simplicité d'utilisation des écrans tactiles peut amener les sujets à préférer cette technologie à l'arrêt.

En revanche, la situation est tout à fait différente après la conduite du jour 2 : la concentration ou la maîtrise de l'interface ne sont plus jugées significativement différentes entre les deux types d'interface. Plus spécifiquement, la concentration est jugée significativement plus importante en conduite comparée à l'arrêt dans le cas tactile, et la maîtrise est jugée significativement moins bonne en conduite. Nous pouvons également observer une augmentation de l'écart type dans les réponses concernant l'interface tactile. Nous pensons que ces résultats témoignent de la difficulté des sujets à utiliser leur modalité visuelle à la fois pour les tâches de maîtrise du véhicule et de suivi du véhicule, et pour les manipulations de l'interface. Cette surcharge visuelle pourrait expliquer les résultats présentés en figure 6.11. Cette interprétation est également appuyée par les réponses aux questions présentées en figures F.3, F.5 et F.6.

Par ailleurs, il est intéressant de noter que la concentration pour la manipulation de l'interface gestuelle n'est pas significativement différente au jour 2 utilisée à l'arrêt et en conduite. La maîtrise de l'IHM gestuelle est même jugée significativement plus importante après la conduite qu'après l'entraînement. Nous pensons que ces résultats témoignent de deux phénomènes. Tout d'abord, les tâches de conduite et de manipulations de l'interface gestuelle peuvent être complémentaires d'un point de vue sensoriel, grâce à leur caractère multimodal, à opposer à la concentration d'informations visuelles dans le cas de l'IHM tactile. D'autre part, nous pensons que les sujets sont encore dans une phase de constitution des schémas cognitifs relatifs à l'OV et à l'interface gestuelle, et que la prise en main de l'interface augmente toujours pendant la phase de conduite. Au contraire, nous pouvons supposer que les sujets sont experts en ce qui concerne l'IHM tactile et ne s'améliorent plus avec le temps. C'est ce phénomène d'apprentissage qui peut expliquer la hausse significative de la note de maîtrise de l'interface gestuelle entre les conditions à l'arrêt et en conduite du jour 2.

Interface de gestion de la musique et remarques générales

En fin d'expérience, nous avons proposé aux sujets de manipuler une interface permettant de contrôler plusieurs fonctions liées à l'écoute de la musique. Les sujets semblent avoir plébiscité ce type d'interface, avec une note moyenne de 6,07 sur un maximum de 7, voir tableau 6.6. D'après les impressions de plusieurs sujets, la manipulation directe de la musique a été particulièrement appréciée. Le tableau 6.5 permet d'avoir un détail plus précis suivant la fonction manipulée, et il est à noter que les balayages de la main permettant de changer un morceau de musique ont obtenu une note convenable. Ce résultat est à mettre en regard du rejet des balayages dans le cas de l'interface de ventilation. Ces différences d'appréciation peuvent provenir de la différence de contexte lors de l'utilisation des IHM : l'interface de ventilation était manipulée dans un contexte de conduite plutôt stressant, alors que l'interface de gestion de la musique était utilisée à l'arrêt. Le placement de la source et la manipulation du volume ont semble-t-il été les fonctions préférées des sujets. Ces résultats peuvent nous permettre de penser qu'une transposition de l'interface gestuelle développée sur plusieurs types de fonctions – GPS, musique, ventilation, téléphone, ou autre – peut être bien acceptée par les utilisateurs.

Enfin, nous pouvons observer sur le tableau 6.6 que le nombre de gestes à retenir a paru convenable aux utilisateurs, avec une note de 6,19 sur un maximum de 7. Nous pouvons alors penser que notre supposition développée en section 6.1.3 était plutôt bonne : les sujets doivent plutôt retenir le nombre de manipulations de l'OV qui peuvent être réalisées, et non le nombre de gestes. Il peut donc être possible de proposer une ou plusieurs manipulations de l'objet virtuel additionnelles, pour augmenter les possibilités d'interactions avec des fonctions de l'interface grâce à notre objet virtuel.

6.3.5 Conclusion

Dans cette expérience perceptive accomplie en simulateur de conduite, nous avons désiré évaluer notre interface gestuelle et la comparer à une interface tactile, dont la technologie est utilisée dans certains véhicules récents. Nous avons choisi de proposer cette expérience sur deux jours consécutifs, pour permettre aux sujets d'avoir le temps d'apprendre à utiliser l'OV et l'interface gestuelle. Le premier jour est donc consacré à la prise en main de l'interface et à un entraînement de plusieurs manipulations. Un rappel des manipulations est proposé au début du deuxième jour, avant d'effectuer un parcours sur le simulateur de conduite, où les manipulations de l'interface sont régulièrement demandées.

Les résultats principaux de cette expérience montrent que l'interface tactile conduit à de meilleures manipulations que l'interface gestuelle dans 6 manipulations sur les 24 proposées. Nous avons vu que ces manipulations concernent les balayages de la main du côté de l'IHM gestuelle, révélant un geste particulièrement difficile à exécuter, à l'arrêt comme en conduite. Les paramètres de conduite que nous avons enregistrés sur le simulateur de conduite n'ont pas révélé de différences significatives entre les deux types d'interface. En particulier, aucune différence concernant l'écart type de l'écart latéral ou la vitesse n'a été constatée, qui se sont révélés être des paramètres pouvant différencier des distractions visuelles ou cognitives dans de précédentes études. La tâche de suivi d'un véhicule lièvre, pour simuler une demande visuelle liée au trafic routier, a en revanche conduit à des différences significatives. En particulier, l'interface gestuelle a permis de mieux accomplir la tâche demandée.

De manière générale, la concentration des sujets a semblé plus importante avec l'interface gestuelle lors de son utilisation comme tâche principale à l'arrêt. En revanche, il semble que la charge mentale des sujets a augmenté avec l'IHM tactile entre les conditions à l'arrêt et en conduite, ce qui n'est pas le cas avec l'interface gestuelle. Nous pensons que ces résultats témoignent d'une meilleure répartition des informations entre les différentes modalités du conducteur dans le cas de l'interface gestuelle. L'interface tactile concentre les informations sur la modalité visuelle : une surcharge visuelle peut entraîner une dégradation des performances – ce que nous avons observé sur la tâche de suivi, et une augmentation de la concentration mentale nécessaire. Plusieurs réponses du questionnaire vont dans le sens de cette supposition.

L'interface gestuelle peut donc sembler une piste intéressante comme complément à l'interface tactile, pour des manipulations simultanées à la conduite. L'interface tactile reste cependant privilégiée lorsque le conducteur n'est pas occupé, à l'arrêt. Plusieurs améliorations peuvent être apportées à l'interface gestuelle pour augmenter sa facilité de prise en main, comme la redéfinition de balayages de la main plus faciles à exécuter. Ces pistes seront explorées dans les perspectives exposées à la fin de ce document.

Les résultats de cette expérience sont encourageants, mais ils doivent être confirmés par rapport à plusieurs critères. Tout d'abord, le simulateur de conduite ne permet pas de recréer une situation réelle de conduite. En particulier, la perception du danger peut être minimisée en simulateur de conduite. Nous pouvons alors supposer qu'une étude sur route pourrait étayer les résultats de l'expérience. Ensuite, les instructions sont ici dictées pour mesurer si les conducteurs arrivent à réaliser certaines manipulations précises. Cette situation ne correspond pas à la réalité, où le réglage à atteindre est directement déterminé par le conducteur, grâce aux retours perceptifs donnés par la modification de la fonction – l'intensité du flux d'air ressenti sur la peau et par le son par exemple. Enfin, il serait intéressant de prolonger cette étude en s'intéressant à l'évolution des performances sur le long terme. En effet, les sujets n'ont pu manipuler qu'une seule fois l'IHM gestuelle tout en conduisant dans cette expérience. Nous pouvons imaginer que l'expertise de manipulation de l'IHM gestuelle peut augmenter avec le temps, ce qui peut conduire à une baisse de la concentration nécessaire au maniement de l'interface, grâce à la constitution de schémas cognitifs. Nous avons amorcé une telle étude avec notre panel d'experts qui a effectué plusieurs parcours de conduite de la dernière expérience. Nous avons pu observer des tendances encourageantes, il est néanmoins nécessaire de mettre au point une expérience perceptive centrée sur cette acquisition de schémas cognitifs relatifs à l'objet virtuel. Nous pouvons alors espérer qu'après plusieurs séances de conduite, l'interface gestuelle conduise à de meilleurs résultats en terme de manipulation d'interface, de conduite ou de concentration des conducteurs.

Conclusion du chapitre

Dans ce chapitre, nous avons choisi de mettre en application les enseignements des quatre précédentes expériences perceptives dans un contexte automobile. Notre réflexion s'est tout d'abord basée sur les cas d'usages d'IHM gérant la ventilation ou la musique que nous avons évoqués dans le précédent chapitre. Plusieurs fonctions ont été ciblées, dont les manipulations grâce à des gestes pouvaient être prometteuses. Ces fonctions ont été regroupées en familles évoquant des concepts proches : le réglage d'une intensité, d'un placement dans l'espace, ou d'un défilement.

Trois manipulations de l'objet virtuel ont été proposées, et elles ont été liées à trois gestes différents pouvant être exécutés en série. Ces manipulations ont été rapprochées des familles de fonctions, et une attention particulière a été portée à la cohérence entre la métaphore de manipulation de l'objet virtuel, les gestes qui y sont liés et la fonction modifiée. Nous avons donc mis au point une interface de ventilation, et une interface de gestion de la musique possédant une construction similaire. Il est important de noter que l'interface de gestion de la musique a la particularité de se servir directement du contenu sonore écouté par l'utilisateur pour lui transmettre un retour d'information.

Cette interface gestuelle a par la suite été intégrée à un simulateur de conduite, pour permettre de réaliser une expérience perceptive, ayant pour but de tester les effets attentionnels liés à la manipulation d'une interface simultanément à une tâche de conduite. Une comparaison avec une interface tactile, utilisée couramment dans les véhicules récents, a été proposée pour évaluer la pertinence de notre interface gestuelle. Une confrontation a été choisie pour cette expérience, mais il pourrait être intéressant de combiner ces deux interactions possibles, pour offrir un panel de possibilités aux futurs conducteurs.

Les résultats montrent que certaines manipulations en conduite sont mieux effectuées avec l'interface tactile, alors que pour d'autres les différences entre les deux interfaces ne sont pas significatives. Nous avons pu constater que les manipulations posant problème correspondent aux balayages de la main, qui semblent difficiles à réaliser. Des modifications peuvent permettre de faciliter l'exécution de ces balayages, comme la définition de gestes personnalisés ou la définition d'une seule zone, permettant un unique réglage.

Les données de conduite ne révèlent pas de comportement particulier lié à l'interface en ce qui concernent la maîtrise du véhicule. En revanche, la tâche de suivi d'un véhicule livreur a été mieux réalisée lors de l'utilisation de l'interface gestuelle, comparé à l'usage de l'interface tactile. Nous pensons que ces différences témoignent de la difficulté qu'ont eu les sujets à accomplir convenablement les tâches demandées pour l'interface tactile, à cause d'une surcharge de la modalité visuelle. Au contraire, l'interface gestuelle, en diversifiant les sollicitations sensorielles, peut permettre de mieux accomplir certaines tâches.

Ces suppositions sont appuyées par des mesures subjectives de la charge mentale des sujets que nous avons mesurés grâce à l'échelle RSME – ou *Rating Scale of Mental Effort*. L'interface gestuelle utilisée conjointement à une tâche de conduite ne demande pas de plus grande concentration que l'interface tactile. En revanche, l'interface tactile était perçue significativement plus facile à utiliser du point de vue de la charge mentale à l'arrêt. Il semble donc que ces deux interfaces puissent être complémentaires : les sujets peuvent préférer l'interface tactile à l'arrêt, mais ils utiliseront l'interface gestuelle en conduite. Plusieurs réponses au questionnaire appuient cette supposition.

Les résultats de cette étude ouvrent donc des perspectives d'études prometteuses d'un point de vue académique sur la manipulation d'interfaces gestuelles sur le temps long, mais également pour le Groupe PSA qui peut exploiter ces résultats pour mettre au point un complément gestuel à leurs interfaces tactiles actuelles.

Ces travaux de thèse ont porté sur l'association de gestes et de sons pour permettre le contrôle d'une interface utilisée dans un contexte de conduite. L'objectif était de proposer une interface pilotée par des gestes, avec des retours d'informations sonores pour permettre une manipulation en aveugle. Comme nous l'avons vu, les écrans tactiles implantés actuellement en véhicule peuvent entraîner une perte d'attention [Tsimhoni *et al.*, 2004] [Crandall et Chaparro, 2012], avec des conséquences potentiellement néfastes sur la conduite [Dingus *et al.*, 2006]. En effet, la modalité visuelle est essentielle pour assurer la conduite sécuritaire du véhicule dans le flot de véhicules du trafic, mais elle est également sollicitée pour manipuler l'interface tactile sur certaines automobiles récentes. Le réglage d'une fonction sur un écran tactile nécessite au préalable de repérer visuellement la position du bouton virtuel à actionner, puis de contrôler si cette opération a été correctement exécutée. L'ensemble de ces sollicitations de la vision des conducteurs peut entraîner une saturation de ce canal sensoriel, et demander une concentration trop importante au conducteur pour assurer l'ensemble des tâches simultanément.

Approche multisensorielle

Notre approche a consisté à exploiter d'autres modalités que la vision pour interagir avec l'IHM. Nous avons proposé d'utiliser des gestes sans contact pour donner des instructions au système central. Pour assurer un retour d'information à l'utilisateur, la modalité auditive a semblé pertinente en situation de conduite. Nous pensons que cette association multisensorielle peut permettre de diminuer la charge cognitive des conducteurs par rapport aux manipulations des interfaces tactiles, par une meilleure répartition des informations vers différents canaux sensoriels. Pour mettre en place un couplage pertinent, nous avons introduit un objet virtuel comme médiateur entre les gestes de l'utilisateur et les sons. Nous avons imaginé un objet virtuel sphérique manipulable par des gestes, qui serait le support d'une stratégie de sonification et des interactions avec l'IHM. Cet objet peut permettre l'agrégation de plusieurs gestes autour d'un même concept, en proposant des métaphores de manipulations d'objets réels. Une interface gestuelle a été mise au point en suivant ce principe, où nous avons pu redéfinir un système de menu en liant trois fonctions à trois manipulations de l'objet virtuel.

Ces travaux de thèse ont été menés de manière séquentielle, en construisant tout d'abord l'objet virtuel et les concepts liés, avant de progressivement mettre au point une interface gestuelle, pour finalement proposer des manipulations de l'interface gestuelle à des sujets compatibles avec une situation de conduite. Nous avons fait appel à différents champs scientifiques : la perception du son et la proprioception, la perception multisensorielle et attentionnelle, certains aspects cognitifs liés à la mémoire et à la création de schémas cognitifs, la psychologie expérimentale lors de nos tests perceptifs, et les statistiques lors de nos analyses de données.

Synthèse des principaux résultats

Nous avons donc choisi d'étudier différents phénomènes, pour en tirer des informations et progresser vers l'objectif de la manipulation de l'interface gestuelle en aveugle, simultanément à une tâche de conduite.

Les stratégies de sonification

Après avoir défini le concept de notre objet virtuel, nous nous sommes attachés à construire une stratégie de sonification transmettant des informations pertinentes sur sa dynamique. Nous voulons en effet qu'il soit manipulable en aveugle, et donc que la stratégie sonore permette aux utilisateurs de ressentir ses mouvements.

Nous avons choisi deux paramètres à sonifier – la distance main-OV et la vitesse de l'Ov – et trois paramètres sonores pouvant transmettre ces informations – la hauteur, la modulation d'amplitude et la brillance, qui nous ont permis de former six stratégies de sonification. Une première expérience perceptive nous a conduit à choisir un lien ou *mapping* pour chaque stratégie sonore entre le paramètre de l'Ov à sonifier et le paramètre sonore. Six mappings différents étaient proposés, composés de trois courbes différentes – linéaire, exponentielle ou logarithmique – combinées à deux polarités – croissante ou décroissante. Le choix du mapping pour chaque stratégie sonore a découlé d'un test de comparaison par paire, où deux liens étaient proposés pour une même stratégie. Les sujets devaient déterminer quel mapping leur semblait le plus naturel en prenant en compte des stimuli audio-visuels.

Ces résultats nous ont permis de mettre en place une seconde expérience visant à hiérarchiser les stratégies sonores, pour ne sélectionner que celles transmettant les informations les plus riches sur la dynamique de l'objet virtuel. Une tâche d'envoi de l'Ov dans des zones a été proposée pour discriminer les stratégies. Nous avons fait l'hypothèse que la stratégie permettant le mieux d'accomplir cette tâche véhicule des informations pertinentes pour percevoir la dynamique de l'objet virtuel. Deux stratégies de sonification ont été conservées à la suite de cette étude : une stratégie illustrant la distance main-Ov par une modification de la hauteur du son, et une sonification de la vitesse de l'Ov par le paramètre de brillance. Nous avons qualifié la première stratégie de sonification *abstraite* par opposition à la seconde, qui semble proche d'un son de roulement et qui a été caractérisée comme stratégie sonore *naturelle*. Il paraissait intéressant de conserver deux stratégies, pour étudier plus précisément l'étape de prise en main de l'Ov sur une période longue.

Dans la première expérience, nous avons fait le choix de limiter les bornes d'évolution des paramètres sonores, et nous avons évalué trois mappings croissants et trois mappings décroissants. Ces décisions que nous avons prises arbitrairement peuvent être discutables compte tenu du grand nombre d'autres possibilités, notamment au niveau des bornes fréquentielles. Nous avons dû trancher pour mettre en place notre expérience, et il est possible que d'autres limites de paramètres mènent à des résultats différents. Des études complémentaires pourraient être menées sur la base de notre expérience de comparaison des mappings, mais en proposant plusieurs limites minimales et maximales à l'évolution du paramètre sonore pour un même mapping et une même stratégie sonore. Nous pensons cependant que notre étude, en faisant le choix de paramètres sonores courants et en proposant une diversité de mapping différents, a pu répondre à nos interrogations sur le lien entre paramètres de l'OV et paramètres sonores, avec parfois des résultats nets.

La discrimination des stratégies sonores que nous avons mis en place par la suite aurait pu tenir compte de l'aspect attentionnel, qui semble important pour la perception [Mack et Rock, 1999], puisque nous désirons proposer des manipulations de l'OV dans un contexte de conduite. Ainsi, une prochaine expérience perceptive pourrait proposer une tâche attentionnelle en plus de la tâche d'envoi de l'objet dans les zones, et la discrimination des stratégies de sonification pourrait être différente. Nous pensons cependant que les résultats seraient similaires, avec de bonnes performances pour la stratégie *abstraite* – liée à son adéquation avec la tâche demandée – et pour la stratégie *naturelle* – grâce à sa similarité avec des sons de roulement.

Prise en main de l'objet virtuel

Nous nous sommes ensuite rapprochés des conditions véhicule, en mettant en place une première version d'interface gestuelle, permettant le réglage d'une fonction de température. Nous avons également revu le design des retours sonores, en se basant sur des textures sonores dont nous avons travaillé l'esthétique. Nous avons cependant à l'esprit que les stratégies de sonification devaient rester informatives. En particulier, nous avons combiné une texture à un son de roulement pour la stratégie de sonification liant la vitesse de l'OV au paramètre de brillance. Nous avons qualifié cette stratégie sonore de *naturelle* après avoir discuté des résultats des deux premières expériences avec les résultats de Houben [Houben *et al.*, 2001] [Houben, 2002] [Houben *et al.*, 2004], car elle pouvait se rapprocher de sons de roulement. La géométrie sphérique que nous avons choisie pour l'OV permet ici de créer une métaphore complète en associant de manière cohérente le visuel de l'OV et la stratégie de sonification.

Plusieurs questions concernant la prise en main de l'OV se sont alors posées. Dans un premier temps, nous nous sommes intéressés aux évocations pouvant être véhiculées par les manipulations gestuelles de notre objet virtuel sonifié lors de sa première prise en main : les liens entre les gestes exécutés et les sons perçus permettent-ils d'imaginer la présence d'un objet virtuel ? Les concepts liés à l'OV peuvent-ils être correctement compris par des sujets naïfs ? Une troisième expérience a été mise en place pour répondre à ces questions, où des sujets naïfs découvraient l'objet virtuel en aveugle, uniquement grâce à leurs interactions gestuelles et aux retours sonores. Lors de cette expérience, une analyse thématique a permis de faire ressortir certaines images mentales formulées de manières répétées par les sujets. En particulier, la sonification de la vitesse de l'OV par le son de roulement de synthèse a évoqué à plusieurs sujets des concepts proches des choix que nous avons fait concernant l'objet virtuel : le caractère sphérique, la présence d'un ressort ou la restriction à des mouvements horizontaux. Nous pensons que les évocations peuvent aider les sujets à avoir une représentation mentale de l'objet virtuel, et donc participer à une meilleure constitution de sché-

mas cognitifs à partir des stimuli multisensoriels gestuels, visuels et sonores. En nous projetant en véhicule, ces évocations peuvent être intéressantes lors d'une prise en main de l'OV à l'arrêt, mais également pendant une utilisation de l'OV en aveugle tout en conduisant.

Nous nous sommes ensuite interrogés sur l'intégration multisensorielle lors des premières manipulations de l'OV pour favoriser une utilisation en aveugle : faut-il proposer des stimuli audio-visuels en premier lieu ? L'une des deux stratégies de sonification conservées jusqu'à présent permet-elle une meilleure intégration et une meilleure utilisation de l'OV en aveugle ? Dans une quatrième expérience perceptive, nous avons fait le choix de présenter des stimuli audio-visuels dans une première phase, avant de poursuivre par une phase en masquant le visuel de l'OV. La première partie peut correspondre à une prise en main de l'OV à l'arrêt, lors de la première utilisation du système grâce à un didacticiel par exemple, alors que la seconde représente l'utilisation du système à l'aveugle en conduite. Les résultats ont montré que les sujets semblent concentrer uniquement leur attention sur le visuel de l'OV lors de la première phase, probablement à cause de la forte influence de cette modalité au sein du système perceptif humain [De Gelder et Vroomen, 2000] [Colavita, 1974] [Spence, 2009] [Koppen et Spence, 2007] [Koppen *et al.*, 2009]. Nous pensons alors qu'ils n'ont pas développé de schéma cognitif liant le visuel de l'OV, les stratégies de sonification, et leurs gestes. Pour une application en aveugle, il paraît donc essentiel de limiter l'apport d'informations visuelles, pouvant entraîner un phénomène de « capture attentionnelle ».

Dans cette expérience, nous avons également fait le choix de ne pas proposer de tâche attentionnelle, en particulier dans la seconde phase. Une nouvelle étude pourrait proposer une distraction visuelle – la détection d'une cible par exemple – pour préciser cet aspect. La première phase pourrait également être l'objet d'un travail supplémentaire, en proposant trois conditions à trois groupes de sujets différents : une condition auditive seule, une condition visuelle seule, et une condition audio-visuelle. Dans ce cas, il serait possible de confirmer ou non le phénomène de « capture attentionnelle » opéré par la modalité visuelle. Nous pouvons penser, compte tenu des réactions et commentaires des sujets aux expériences 4 et 5, que cette capture a bien lieu.

Interface gestuelle et conduite

Nous avons ensuite voulu évaluer les effets attentionnels de la manipulation de l'interface simultanément à une tâche de conduite. Une interface gestuelle a tout d'abord été mise au point, liant trois manipulations de l'OV à trois fonctions d'une interface de ventilation, avant d'être intégrée dans un simulateur de conduite. Nous avons alors voulu évaluer l'utilisation de cette IHM gestuelle en conduite. En particulier, est-il possible que les conducteurs développent des schémas cognitifs nécessaires à la manipulation de l'interface gestuelle ? L'attention des sujets portée aux manipulations de l'IHM peut-elle être réduite avec notre interface gestuelle comparée à une interface tactile ? Une cinquième expérience perceptive a été mise en place, pour répondre à ces questions. Cette expérience, réalisée sur deux jours consécutifs, consistait dans un premier temps à prendre en main l'interface à l'arrêt. Pour l'interface gestuelle, cette étape doit permettre de construire les schémas cognitifs nécessaires à la manipulation de l'OV en conduite. Pour éviter un phénomène de capture attentionnelle, les stimuli visuels de l'OV n'ont été disponibles que brièvement, uniquement pour que les sujets se constituent une image mentale. L'ensemble de l'entraînement aux manipulations de l'interface gestuelle s'est donc déroulé uniquement avec des stimuli auditifs. Des retours visuels et auditifs étaient proposés pour l'IHM tactile. Le deuxième jour, à la suite d'un rappel des manipulations possibles de l'interface, les sujets ont effectué des manipulations des interfaces tout en conduisant sur le simulateur de conduite. Ils devaient à la fois assurer une conduite sécuritaire,

accomplir les manipulations de l'interface demandées, et suivre un véhicule lièvre en respectant une distance maximum. Cette tâche avait pour but de simuler la demande visuelle liée au trafic de véhicules.

Les résultats montrent de meilleures manipulations de l'IHM avec l'interface tactile, en particulier pour la fonction de température, où les gestes de balayage ont semblé particulièrement difficiles à réaliser. Aucun effet significatif n'a été observé sur les données de conduite. En revanche, l'interface gestuelle donne de meilleures performances concernant la tâche de suivi du véhicule. Nous pensons que ces résultats s'expliquent par la possibilité offerte par l'interface gestuelle d'accomplir des interactions en aveugle, les conducteurs ont donc pu concentrer leur regard sur la route et donc accomplir plus facilement la tâche de suivi du véhicule. Les mesures subjectives de la concentration mentale des sujets sont également prometteuses : l'interface gestuelle demande significativement plus de concentration pour la manipuler à l'arrêt par rapport à l'IHM tactile, mais la charge cognitive mobilisée est comparable en conduite entre les deux types d'interface. Il est important de noter que les sujets peuvent être considérés comme experts sur la technologie tactile, puisqu'elle est implantée dans de nombreux objets utilisés quotidiennement – smartphones ou tablettes par exemple. Au contraire, l'association de gestes et de sons pour manipuler une interface est nouvelle. Nous pouvons alors faire l'hypothèse qu'avec l'IHM gestuelle, la concentration va baisser avec l'augmentation de l'expertise des conducteurs, alors que ce phénomène n'aura pas lieu avec l'interface tactile. Il pourrait alors être possible de montrer que l'IHM gestuelle est intéressante d'un point de vue cognitif, et permet aux conducteurs de mobiliser leur attention sur la tâche de conduite. Les manipulations proposées par ce type d'interface pourraient donc être plus sécuritaires qu'avec les écrans tactiles, ce que semble déjà ressentir la majorité des sujets de cette expérience.

Cette dernière expérience s'est déroulée en simulateur de conduite, pour des raisons de sécurité vis-à-vis des sujets de l'expérience, mais aussi car les stimuli perceptifs liés à la conduite peuvent être facilement contrôlés. Malgré une validité relative du simulateur, il pourrait être intéressant de mettre en place cette expérience en conduite réelle. La perception est en effet différente lors d'une tâche de conduite réelle, la perception du danger est en particulier altérée en simulateur de conduite. Les conducteurs pourraient avoir plus de difficultés à effectuer un réglage de l'interface dans une situation de conduite réelle demandant une attention soutenue. Les résultats obtenus pour les IHM gestuelle et tactile pourraient être différents.

Critères de Bastien et Scapin

Lors du développement de notre IHM gestuelle, nous avons tenté d'être fidèle au maximum aux critères de Bastien et Scapin, permettant de construire des interfaces ergonomiques [Bastien et Scapin, 1993]. Nous allons ici discuter de chacun de ces points par rapport à notre interface.

- **Guidage** : le guidage est assuré en partie par notre stratégie de sonification, qui permet aux utilisateurs de se rendre compte en temps réel des déplacements latéraux de l'objet virtuel et d'adapter leurs gestes. Le déplacement vertical de la source de ventilation est accompagné par une modification de la hauteur du son selon la position, ainsi que par des earcons lorsque la source est exclusivement au niveau de la tête ou les pieds du conducteur. Un guidage est également proposé avec le réglage de l'intensité par le volume du son de ventilation qui est modulé en temps réel, ainsi que par les earcons illustrant l'apparition ou la disparition de chaque quartier de la jauge d'intensité. Ces remarques sont également applicables à l'interface de musique, où la matière sonore écoutée assure le retour d'information en temps réel, et guide les utilisateurs pour atteindre le réglage qu'ils souhaitent.
- **Charge de travail** : la mesure de la concentration mentale des sujets par l'échelle RSME a montré le potentiel de l'interface gestuelle par rapport à une interface tactile. L'expérience 5 a montré que la concentration nécessaire à des manipulations de l'interface gestuelle est équivalente à celle requise pour manier l'IHM tactile. Nous pensons que de futures études concernant le développement de l'expertise dans les manipulations gestuelles de l'OV et de l'IHM peuvent permettre de significativement mettre en évidence un intérêt attentionnel de l'interface gestuelle par rapport à l'IHM tactile.
- **Contrôle explicite** : ce critère concerne d'une part la relation entre les actions de l'utilisateur et les réglages de l'interface ; et d'autre part le contrôle du déroulement des traitements informatiques en cours, avec par exemple la possibilité de les interrompre ou de les reprendre. Les gestes étant directement liés à des manipulations de l'OV et à des fonctions de l'IHM, nous avons bien donné le contrôle aux utilisateurs sur les réglages de l'interface. Le système de verrouillage de l'interface permet d'interrompre la manipulation d'une fonction, par exemple si une réaction rapide sur le contrôle du véhicule est nécessaire à cause d'une situation dangereuse. La reprise du réglage est alors possible en déverrouillant l'IHM.
- **Adaptabilité** : ce critère regroupe à la fois la flexibilité de l'interface et la prise en compte de l'expérience de l'utilisateur. La première version de l'interface gestuelle que nous avons mise au point ne remplit pas les conditions de ce critère, mais plusieurs améliorations peuvent être apportées comme nous le verrons dans les perspectives, avec de l'apprentissage automatique ou en adaptant les stratégies de sonification à l'expertise des sujets.
- **Gestion des erreurs** : comme nous avons pu le mentionner, la captation des gestes des utilisateurs a été parfois inconstante, ce qui a pu mettre en difficulté les sujets. L'utilisation d'un capteur de gestes plus précis peut par exemple pallier ce problème. D'autre part, proposer une seule zone de chaque côté pour les balayages de la main peut permettre de réduire les erreurs de manipulation. Enfin, l'utilisation d'algorithmes d'apprentissage automatique des gestes peut faciliter la prise en main de l'interface, et diminuer les erreurs qui ont pu être observées en figure 6.7b.

- Homogénéité/cohérence : l’homogénéité entre les interfaces de ventilation et de gestion de la musique a été développée dans un paragraphe de la section 6.1.4. Nous avons construit ces deux interfaces en regroupant les fonctions que nous voulions pouvoir contrôler en familles, auxquelles nous avons associés un geste et une position similaire sur l’interface visuelle. De cette manière, les sujets peuvent passer d’une interface à l’autre sans avoir besoin d’apprendre de nouvelles manipulations. Seule l’interaction avec la fonction de placement de source n’est pas cohérente entre ces deux interfaces, à cause de la différence intrinsèque au réglage – placement vertical ou dans le plan horizontal. En conséquence, la représentation graphique du réglage n’est également pas homogène.
- Signifiante des codes et dénominations : ce critère concerne « l’adéquation entre l’objet ou l’information affichée ou entrée, et son référent ». En ce qui concerne notre interface gestuelle, ce critère semble concerner l’adéquation entre les gestes, les manipulations de l’objet virtuel et les fonctions. En choisissant une métaphore de manipulations d’un objet simple du quotidien, il semble que ce critère ergonomique soit rempli, puisqu’une correspondance existe entre ces trois parties de l’IHM. La concordance entre gestes, manipulations de l’OV et fonctions a par exemple permis aux sujets de se rappeler facilement du fonctionnement de l’interface grâce à leurs gestes, comme le montre la figure H.2. La mémorisation des codes de l’interface semble donc correcte.
- Compatibilité : le critère de compatibilité concerne l’accord entre la mémoire, la perception, les habitudes de l’utilisateur et les tâches d’une part ; et « l’organisation des sorties, des entrées et du dialogue d’une application donnée » d’autre part. Comme nous l’avons évoqué précédemment, nous avons essayé de lier au maximum des gestes courants à des manipulations habituelles de véritables objets, pour interagir avec des fonctions qui peuvent naturellement être liées à ces gestes. D’autre part, l’utilisation d’un son de roulement, permet d’exploiter des schémas cognitifs existants de manipulations d’objets pour les adapter à la manipulation de notre interface gestuelle. Enfin, la coordination entre les entrées et les sorties de l’interface, les gestes et les sons dans notre cas, ont été au cœur de ces travaux de thèse. Les résultats de l’expérience 5 montrent que le dialogue entre IHM et conducteur peut être compris et mémorisé par les sujets ; et que des manipulations de l’interface peuvent être accomplies simultanément à une tâche de conduite.

Il semble donc que, même si tous les critères ne sont pas respectés par la première version de l’interface que nous avons mis au point, plusieurs améliorations peuvent permettre de remplir au mieux les critères de Bastien et Scapin. Ces modifications peuvent permettre de mettre au point une interface gestuelle ergonomique et, nous l’espérons, intuitive.

Perspectives

Coordination des gestes

Lors de l'utilisation d'une interface – tactile ou gestuelle – simultanément à une tâche de conduite, une main reste sur le volant pour assurer la maîtrise de la direction du véhicule. Dans le cas de notre interface gestuelle, et dans la perspective de multiplication des gestes à réaliser, nous pouvons nous interroger sur la facilité de réalisation de gestes simultanés. En effet, certains d'entre eux permettant la manipulation de l'interface pourraient introduire une distraction motrice. Il est donc question de savoir si des gestes peuvent interférer avec les manipulations du volant, ce qui pourrait conduire à des situations dangereuses. Une étude pourrait donc être menée, en demandant à des conducteurs d'effectuer différents gestes de la main droite – du haut vers le bas, des cercles ou des rotations de la main par exemple – tout en proposant des mouvements de volant correspondant à des trajectoires variées – ligne droite, virages serrés ou non. Il serait ainsi possible de voir, par la précision d'accomplissement des gestes et l'écart latéral du véhicule par exemple, si l'accomplissement simultané de deux gestes peut poser problème.

De nouvelles manipulations

L'interface gestuelle que nous avons mise au point permet la manipulation de trois fonctions, par l'intermédiaire de trois manipulations de l'objet virtuel. D'après la réponse au questionnaire présenté en figure H.2, les sujets de l'expérience menée en simulateur ont réussi à retenir les gestes nécessaires pour les différents réglages de l'interface. Comme nous l'avons précédemment évoqué, la limite de sept éléments pouvant être simultanément retenus en mémoire de travail ne semble donc pas s'appliquer ici aux gestes, qui sont au nombre de dix [Miller, 1956]. Nous pensons donc que cette limite s'applique plutôt aux manipulations possibles de l'objet virtuel, qui sont cognitivement représentées comme un seul élément.

Il pourrait donc être possible d'ajouter d'autres manipulations de l'OV, pour que les acteurs industriels puissent proposer des réglages supplémentaires de fonctions par des gestes. En prenant notre exemple de l'interface de ventilation, il pourrait être intéressant de pouvoir manipuler la fonction de dégivrage avant ou arrière, ou encore de modifier la température des sièges avant de manière individualisée. Nous pensons qu'il est alors nécessaire que le lien entre les gestes, les manipulations de l'OV et les sons respecte la métaphore que nous avons introduite grâce à l'OV pour la création de schémas cognitifs dans les meilleures conditions. La question est alors de savoir s'il est possible de proposer des fonctions supplémentaires, basées sur des manipulations de l'OV, sans provoquer de surcharge cognitive chez les utilisateurs.

Vers un objet virtuel libre en deux ou trois dimensions ?

Pour permettre des manipulations en aveugle, nous avons choisi de contraindre les mouvements de l'objet virtuel à des déplacements en une dimension horizontale. Nous avons en effet pensé que les mouvements d'un objet virtuel libre peuvent être difficiles à contrôler de manière générale, et à fortiori sans retour visuel. Cette contrainte permet par exemple d'éviter que l'objet ne tourne autour de la main, dans le cas où il aurait une accélération centrifuge suffisamment importante pour contrebalancer la force de rappel du ressort.

Il pourrait être intéressant, par exemple dans le but de permettre de nouvelles interactions avec l'OV, d'autoriser des déplacements de l'objet en deux ou trois dimensions. Il est alors essentiel de réfléchir à des moyens pour le contrôler facilement, d'autant plus s'il n'est pas possible de le voir. Par exemple, il pourrait être intéressant de proposer aux conducteurs de pouvoir lancer l'OV vers un point d'intérêt, selon un azimuth particulier. Le système du véhicule, grâce à sa position GPS et à la direction du lancer, pourrait alors trouver le lieu visé par l'utilisateur, et lui en donner une description. Cette situation peut par exemple se produire sur l'autoroute, où certains lieux particuliers peuvent attirer l'attention des conducteurs ou des passagers.

Des gestes personnalisés

Dans la première version de l'interface gestuelle, certains gestes précis sont attendus pour modifier une fonction, et en particulier certains gestes *d'entrée* pour s'adresser à une fonction particulière, voir figure 6.1. Si l'utilisateur ne positionne pas correctement sa main, certains gestes peuvent ne pas être interprétés par le système, pouvant entraîner une expérience désagréable. Sur le long terme, de mauvaises détections de la main répétées peuvent amener les conducteurs à se détourner de l'interface gestuelle.

L'enregistrement de gestes personnalisés peut alors être une solution. Lors de sa familiarisation avec le système par le biais d'un didacticiel par exemple, il serait demandé à l'utilisateur de réaliser plusieurs fois chaque manipulation de l'OV. Les paramètres pertinents de ces gestes seraient enregistrés, et des algorithmes d'apprentissage automatique permettraient de reconnaître les manipulations de l'objet personnelles à chaque utilisateur. Il pourrait alors être intéressant d'inverser le processus : chaque utilisateur a-t-il des manières de manier l'objet virtuel qui lui sont propres ? Quels sont les paramètres invariants et variants dans les gestes des conducteurs ? Peut-on reconnaître un utilisateur uniquement avec sa manière de manier des objets ? Peut-on le reconnaître par ses manipulations de l'objet virtuel ?

Expertise de manipulation de l'interface

L'expérience 5 que nous avons présenté dans le chapitre 6 compare une interface gestuelle et une interface tactile en simulateur de conduite. Nous avons voulu confronter ces deux interfaces pour démontrer le potentiel des interfaces gestuelles, mais nous avons conscience que ce duel n'était pas à armes égales. En effet, l'ensemble des sujets ayant participé à cette expérience a déjà été confronté de nombreuses fois à des interfaces tactiles, lors de l'utilisation de tablettes ou smartphones par exemple. Ils ont alors pu développer une expertise dans ces manipulations. À l'inverse, très peu de systèmes ou d'interfaces sont utilisés avec des gestes sans contact. Les interactions en aveugle que nous avons proposées avec notre interface sont donc nouvelles pour les utilisateurs, qui ont dû se familiariser avec l'OV, les retours sonores et l'IHM en deux jours, pour les besoins de l'expérience.

Nous pouvons alors nous interroger sur l'utilisation à long terme de l'interface gestuelle. Nous avons amorcé cette réflexion sur la dernière expérience, avec trois sujets *experts* qui ont accompli quatre fois le parcours de conduite en manipulant simultanément l'interface gestuelle. Même si les données sont trop peu nombreuses pour tirer des conclusions, il semble que les résultats s'améliorent avec le temps : les sujets peuvent s'habituer au système, et maîtrisent de plus en plus l'interface gestuelle.

Pour approfondir cet aspect d'expertise, il pourrait par exemple être intéressant de proposer à des conducteurs d'utiliser l'interface gestuelle sur plusieurs semaines ou plusieurs mois, pour leur permettre de développer des schémas cognitifs relatifs à la manipulation de l'interface, et les habituer à une utilisation simultanée à la tâche de conduite. Un questionnaire permettrait également d'interroger les conducteurs sur leurs usages ou non de l'interface gestuelle, pour recueillir leurs critiques négatives ou positives. Ces retours d'informations peuvent permettre aux développeurs de mettre au point une interface créée pour répondre au mieux aux exigences des utilisateurs.

L'approfondissement de nos études sur l'interface gestuelle sur le long terme peut ensuite permettre une véritable « confrontation » avec l'interface tactile, où les deux interfaces seraient complètement maîtrisées par les sujets. Nous pensons que les performances de manipulation de l'IHM gestuelle peuvent être significativement améliorées avec le temps, et la charge mentale des sujets diminuée en conséquence. Cette piste d'étude peut être la prochaine étape importante à réaliser par le Groupe PSA pour ouvrir une voie à une implantation en véhicule.

Expertise et stratégies sonores

Cette expertise développée dans le précédent paragraphe peut mener à une interrogation sur une évolution à apporter au système, pour accompagner la prise en main des conducteurs. Nous avons vu au chapitre 3 qu'un effet d'inversion d'expertise pouvait avoir lieu [Kalyuga *et al.*, 2003] : les aides mises en place pour permettre aux utilisateurs de constituer des schémas cognitifs peuvent se révéler neutres voir handicapantes pour des utilisateurs habitués.

Une évolution de l'interface gestuelle pourrait par exemple être proposée : la présence des stratégies de sonification diminuerait au fil du temps à partir d'un certain nombre d'utilisations du système à déterminer. Ce système de guidage des gestes par le biais des sons serait en effet moins nécessaire quand l'habitude de manipulation de l'IHM augmente. À partir d'un moment, l'interface gestuelle pourrait même uniquement proposer une validation des modifications de fonctions par l'intermédiaire d'icônes sonores, en ayant supprimé toute stratégie de sonification continue.

Nous pouvons ici constater que la stratégie de sonification que nous avons mise en place dans cette thèse n'est pas un but à atteindre, mais seulement un moyen : elle permet d'accompagner la création d'un schéma cognitif, et guide les sujets pour leur apprendre à utiliser l'interface gestuelle. Par la suite ce guidage peut être supprimé lorsqu'il n'est plus jugé nécessaire.

Le geste comme complément

L'expérience 5 a proposé la manipulation de certaines fonctions par les gestes. Des manipulations de fonctions de ventilation ont été demandées, tout en ayant le visuel d'une interface de GPS. Nous pouvons en effet imaginer l'interaction gestuelle comme un raccourci vers certaines fonctions particulièrement sollicitées pendant la conduite.

Comme nous l'avons évoqué, la modalité gestuelle peut également être complémentaire d'autres modes d'interaction. Son association avec la technologie tactile a déjà été évoquée et peut être prometteuse, pour proposer des interactions complémentaires : l'une adaptée aux situations à l'arrêt, et l'autre aux manipulations de l'interface en situations de conduite. Les sujets de l'expérience 5 ont paru réagir positivement à cette idée, comme le montre la figure H.3.

Vers des interfaces pentasensorielles ?

Le chapitre 2 a été l'occasion d'aborder la problématique de l'intégration multisensorielle, et nous nous sommes particulièrement concentrés sur la proprioception, l'audition et leur association. L'être humain interagit cependant avec l'ensemble de ses canaux sensoriels, véritables ponts entre lui-même et l'environnement. Les interfaces homme-machine peuvent être vues comme le reflet de ces sens, permettant aux *machines* de s'adresser aux utilisateurs en leur donnant des informations au travers de leurs différentes modalités.

Une innovation dans le domaine des interfaces a été d'avoir la possibilité de donner des instructions par l'intermédiaire de la voix. Ce mode d'interaction est en effet naturel pour la communication d'humain à humain, et son intégration aux IHM automobiles permet d'offrir une manière naturelle de dialoguer avec l'interface. L'ajout de l'association de gestes et de sons permet d'étendre cette palette d'interactions possibles, et le conducteur pourra choisir l'interaction qu'il préfère selon sa situation.

Nous pouvons alors nous interroger sur les prochaines évolutions des interfaces. En particulier, allons-nous pouvoir nous affranchir des traditionnelles interfaces WIMP – pour *Windows, Icons, Menus and Pointing* ou fenêtres, icônes, menus et pointage [Beaudouin-Lafon, 2004] – pour interagir avec les IHM, ou va t-on progresser vers une plus grande interactivité [Beaudouin-Lafon, 2016a] [Beaudouin-Lafon, 2016b] ? Avec cet objectif, est-il possible d'utiliser d'autres sens ou d'autres combinaisons de modalités ? Nous avons vu au cours de cette thèse que l'association de modalités permet d'obtenir de nouvelles façons d'interagir, de manière plus naturelle et plus fluide. Dans notre cas, l'association de gestes et de sons peut permettre d'améliorer les interactions avec l'interface simultanément à une situation de conduite. Il est alors possible d'imaginer que de nouveaux rapprochements entre des modalités sensorielles permettront de nouveaux usages, et de meilleures adaptations à des situations particulières.

1. ADHIKARLA, V., SODNIK, J., SZOLGAY, P. et JAKUS, G. (2015). Exploring Direct 3d Interaction for Full Horizontal Parallax Light Field Displays Using Leap Motion Controller. *Sensors*, 15(4):8642–8663. (Cité en page 81.)
2. AKYOL, S., CANZLER, U., BENGLER, K. et HAHN, W. (2000). Gesture Control for Use in Automobiles. *In IAPR WMVA 2000 Workshop on Machine Vision Applications*, pages 349–352, Tokyo, Japon. (Cité en page 23.)
3. ALPERN, M. et MINARDO, K. (2003). Developing a car gesture interface for use as a secondary task. *In CHI'03 extended abstracts on Human factors in computing systems*, pages 932–933. ACM. (Cité en page 23.)
4. ARAMAKI, M., GONDRE, C., KRONLAND-MARTINET, R., VOINIER, T. et YSTAD, S. (2009). Thinking the sounds: an intuitive control of an impact sound synthesizer. *In Proceedings of the International Conference on Auditory Display*, pages 119–124, Copenhagen, Danemark. (Cité en page 49.)
5. ARAMAKI, M., GONDRE, C., KRONLAND-MARTINET, R., VOINIER, T. et YSTAD, S. (2010). Imagine the Sounds: An Intuitive Control of an Impact Sound Synthesizer. *In YSTAD, S., ARAMAKI, M., KRONLAND-MARTINET, R. et JENSEN, K., éditeurs : Auditory Display, Lecture Notes in Computer Science*, pages 408–421. Springer Berlin Heidelberg. (Cité en page 49.)
6. ARAMAKI, M. et KRONLAND-MARTINET, R. (2006). Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech and Language Processing*, 14(2):695–705. (Cité en pages 46 et 49.)
7. AVANZINI, F. et ROCCHESO, D. (2001). Controlling Material Properties in Physical Models of Sounding Objects. *In Proceedings of the International Computer Music Conference*, volume 2001, pages 91–94, La Habana, Cuba. (Cité en pages 46 et 127.)
8. BADDELEY, A. (1992). Working Memory. *Science*, 255(5044):556–559. (Cité en pages 61 et 130.)

9. BARRETT, N. et NYMOEN, K. (2015). Investigations in coarticulated performance gestures using interactive parameter-mapping 3d sonification. *In Proceedings of the International Conference on Auditory Display*, pages 304–305, Graz, Autriche. (Cité en page 53.)
10. BASTIEN, J. C. et SCAPIN, D. L. (1993). Ergonomic criteria for the evaluation of human-computer interfaces. Rapport technique 156, Institut National de Recherche en Informatique et en Automatique, Rocquencourt, France. (Cité en pages 153, 155, 159, 160, 161, et 204.)
11. BATTAGLIA, P. W., JACOBS, R. A. et ASLIN, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, 20(7):1391–1397. (Cité en page 40.)
12. BEATTY, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276–292. (Cité en page 66.)
13. BEATTY, J. et LUCERO-WAGONER, B. (2000). The pupillary system. *In Handbook of psychophysiology*, volume 2, pages 142–162. Cambridge University Press, New York, États-Unis. (Cité en page 66.)
14. BEAUCHAMP, J. W. (1982). Synthesis by Spectral Amplitude and “Brightness” Matching of Analyzed Musical Instrument Tones. *Journal of the Audio Engineering Society*, 30(6):396–406. (Cité en pages 84 et 112.)
15. BEAUDOUIN-LAFON, M. (2000). Instrumental interaction: an interaction model for designing post-WIMP user interfaces. *In Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 446–453. ACM. (Cité en page 6.)
16. BEAUDOUIN-LAFON, M. (2004). Designing interaction, not interfaces. *In Proceedings of the working conference on Advanced visual interfaces*, pages 15–22, Gallipoli, Italie. ACM. (Cité en page 209.)
17. BEAUDOUIN-LAFON, M. (2016a). 50 ans d’interaction homme-machine : retours vers le futur. *Interstices*. (Cité en page 209.)
18. BEAUDOUIN-LAFON, M. (2016b). Mieux penser les interfaces informatiques. *Journal du CNRS*, 285:61. (Cité en page 209.)
19. BERMANT, R. I. et WELCH, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, 42(43):487–493. (Cité en page 39.)
20. BERNARD, C., MONNOYER, J. et WIERTLEWSKI, M. (2018). Harmonious Textures: The Perceptual Dimensions of Synthetic Sinusoidal Gratings. *In PRATTICIZZO, D., SHINODA, H., TAN, H. Z., RUFFALDI, E. et FRISOLI, A., éditeurs : Haptics: Science, Technology, and Applications*, Lecture Notes in Computer Science, pages 685–695, Pise, Italie. Springer International Publishing. (Cité en page 20.)
21. BEZAT, M.-C. (2007). *Perception des bruits d’impact Application au bruit de fermeture de porte automobile*. Thèse de doctorat, Université de Provence-Aix-Marseille I. (Cité en page 46.)

22. BEZAT, M.-C., KRONLAND-MARTINET, R., ROUSSARIE, V. et YSTAD, S. (2014). From acoustic descriptors to evoked quality of car door sounds. *The Journal of the Acoustical Society of America*, 136(1):226–241. (Cité en page 46.)
23. BLAAUW, G. J. (1982). Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors*, 24(4):473–486. (Cité en page 70.)
24. BLAKE, A., BÜLTHOFF, H. H. et SHEINBERG, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, 33(12):1723–1737. (Cité en page 40.)
25. BLAND, J. M. et ALTMAN, D. G. (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal*, 310(6973):170. (Cité en pages 98 et 177.)
26. BLATTNER, M. M., SUMIKAWA, D. A. et GREENBERG, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1):11–44. (Cité en pages vii, 25, et 26.)
27. BLAUERT, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press. (Cité en page 36.)
28. BOLT, R. A. (1980). ‘Put That There’: Voice and Gesture at the Graphics Interface. *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270. (Cité en page 17.)
29. BORDONNÉ, T., DIAS-ALVES, M., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2017). Assessing sound perception through vocal imitations of sounds that evoke movements and materials. In *Proceedings of the 13th International Symposium on Computer Music Modeling and Retrieval*, pages 272 – 282, Porto, Portugal. (Cité en pages 10 et 49.)
30. BOSC, L., DIAZ, E., ROUSSARIE, V., ARAMAKI, M., VELAY, J. L., KRONLAND-MARTINET, R., CHAMARD, J. C. et COUTURIER, J. (2014). Rôle du contenu évocatoire des sons dans l’efficacité de la sonification extérieure de véhicules silencieux. In *Actes du 12ème Congrès Français d’Acoustique*, pages 657–663, Poitiers, France. (Cité en page 116.)
31. BOYATZIS, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications. (Cité en pages 119 et 120.)
32. BOYER, E. O. (2015). *Continuous auditory feedback for sensorimotor learning*. Thèse de doctorat, Pierre et Marie Curie - Paris VI. (Cité en pages 53 et 104.)
33. BOYER, E. O., BEVILACQUA, F., SUSINI, P. et HANNETON, S. (2017). Investigating three types of continuous auditory feedback in visuo-manual trackin. *Experimental Brain Research*, 235(3):691–701. (Cité en page 104.)
34. BOYER, E. O., PYANET, Q., HANNETON, S. et BEVILACQUA, F. (2013). Learning movement kinematics with a targeted sound. In *Proceedings of the 10th International Symposium on Computer Music Modeling and Retrieval*, pages 218–233. Springer. (Cité en pages 28, 53, 54, 109, et 110.)
35. BRAUN, V. et CLARKE, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101. (Cité en page 119.)

36. BRESSOLETTE, B., DENJEAN, S., ROUSSARIE, V., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2017). Sound Selection for Gesture Sonification and Manipulation of Virtual Objects. *World Academy of Science, Engineering and Technology, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 11(1):147–152. (Cité en page 10.)
37. BRESSOLETTE, B., DENJEAN, S., ROUSSARIE, V., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2018). Harnessing Audio in Auto Control: The Challenge of Sonifying Virtual Objects for Gesture Control of Cars. *IEEE Consumer Electronics Magazine*, 7(2):91–100. (Cité en page 10.)
38. BREWSTER, S. A., WRIGHT, P. C. et EDWARDS, A. D. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 222–227. ACM. (Cité en page 26.)
39. BRUNKEN, R., PLASS, J. L. et LEUTNER, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational psychologist*, 38(1):53–61. (Cité en page 67.)
40. CABE, P. A. et PITTENGER, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of experimental psychology: human perception and performance*, 26(1):313–324. (Cité en page 46.)
41. CADOZ, C. (1994). Le geste canal de communication homme/machine: la communication “instrumentale”. *Technique et science informatiques*, 13(1):31 – 61. (Cité en pages 17 et 53.)
42. CADOZ, C. et WANDERLEY, M. M. (2000). Gesture - Music. In Marcelo Wanderley et MARC BATTIER, I.-C. P., éditeur : *Trends in Gestural Control of Music*. (Cité en page 53.)
43. CARAMIAUX, B., BEVILACQUA, F. et SCHNELL, N. (2010). Towards a Gesture-Sound Cross-Modal Analysis. In KOPP, S. et WACHSMUTH, I., éditeurs : *Gesture in Embodied Communication and Human-Computer Interaction*, Lecture Notes in Computer Science, pages 158–170. Springer Berlin Heidelberg. (Cité en page 53.)
44. CARBAJO, A., ROUSSARIE, V., DIAZ, E. et PARIZET, E. (2015). Analysis of vibrational comfort in car equipped with a modified 3-cylinders engine. In *Proceedings of Euronoise 2015*, pages 2423–2428, Maastricht. (Cité en page 70.)
45. CARPENTIER, T. (2015). Récents développements du spatialisateur. In *Journées d'Informatique Musicale*, Montréal, Canada. (Cité en page 81.)
46. CHASE, W. G. et SIMON, H. A. (1973). Perception in chess. *Cognitive psychology*, 4(1):55–81. (Cité en page 62.)
47. CHERRY, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25(5):975–979. (Cité en page 42.)
48. CHIN, E., NATHAN, F., PAUZIÉ, A., MANZANO, J., NODARI, E., CHERRI, C., RAMBALDINI, A., TOFFETTI, A. et MARCHITTO, M. (2006). Subjective Assessment methods for Workload. Rapport technique. (Cité en page 67.)

49. CLARK, J. J. et YUILLE, A. L. (2013). *Data fusion for sensory information processing systems*, volume 105. Springer Science & Business Media. (Cité en page 40.)
50. COLAVITA, F. B. (1974). Human sensory dominance. *Attention, Perception, & Psychophysics*, 16(2):409–412. (Cité en pages 42, 144, 147, et 202.)
51. CONAN, S. (2014). *Contrôle Intuitif de la Synthèse Sonore d’Interactions Solidiennes: vers les Métaphores Sonores*. Thèse de doctorat, École Centrale de Marseille. (Cité en page 46.)
52. CONAN, S., ARAMAKI, M., KRONLAND-MARTINET, R., THORET, E. et YSTAD, S. (2012). Perceptual differences between sounds produced by different continuous interactions. *In proceedings of the Acoustics 2012 Nantes Conference*, pages 581–586, Nantes, France. Société Française d’Acoustique. (Cité en page 49.)
53. CONAN, S., DERRIEN, O., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2014). A synthesis model with intuitive control capabilities for rolling sounds. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(8):1260–1273. (Cité en page 49.)
54. CONAN, S., THORET, E., GONDRE, C., ARAMAKI, M., KRONLAND-MARTINET, R. et YSTAD, S. (2013). An intuitive synthesizer of sustained interaction sounds. *In Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 1045–1050, Marseille, France. (Cité en page 49.)
55. CRANDALL, J. M. et CHAPARRO, A. (2012). Driver Distraction: Effects of Text Entry Methods on Driving Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1):1693–1697. (Cité en pages 73 et 199.)
56. DALY, J., KELLEHEAR, A. et GLIKSMAN, M. (1997). *The Public Health Researcher: A Methodological Guide*. Oxford University Press. (Cité en page 120.)
57. DANNA, J., PAZ-VILLAGRÁN, V., GONDRE, C., ARAMAKI, M., KRONLAND-MARTINET, R., YSTAD, S. et VELAY, J.-L. (2013a). Handwriting sonification for the diagnosis of dysgraphia. *In Proceedings of the 16th Conference of the International Graphonomics Society*, pages 123–126, Nara, Japon. (Cité en pages 27, 54, et 83.)
58. DANNA, J., VELAY, J.-L., PAZ-VILLAGRÁN, V., CAPEL, A., PETROZ, C., GONDRE, C., THORET, E., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2013b). Handwriting movement sonification for the rehabilitation of dysgraphia. *In Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 200–208, Marseille, France. (Cité en page 27.)
59. DE BOER, E. (1980). Auditory physics. Physical principles in hearing theory. I. *Physics reports*, 62(2):87–174. (Cité en page 34.)
60. DE GELDER, B. et VROOMEN, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, 14(3):289–311. (Cité en pages 39 et 202.)
61. DE GROOT, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. *Problem solving*, pages 19–50. (Cité en page 62.)

62. DE WAARD, D. (1996). *The measurement of drivers' mental workload*. Thèse de doctorat, University of Groningen, Traffic Research Centre, Haren, Pays-Bas. OCLC: 246923858. (Cité en pages 29, 66, et 67.)
63. DEL SIGNORE, M. (1997). Electromagnetic detection system for parking operation. Brevet US5682136A, États-Unis. (Cité en page 28.)
64. DENJEAN, S. (2015). *Sonification des véhicules électriques par illusions auditives: étude de l'intégration audiovisuelle de la perception du mouvement automobile en simulateur de conduite*. Thèse de doctorat, Aix-Marseille Université. (Cité en pages 29 et 70.)
65. DENJEAN, S., VELAY, J.-L., KRONLAND-MARTINET, R., ROUSSARIE, V., SCIABICA, J.-F. et YSTAD, S. (2013). Are electric and hybrid vehicles too quiet for drivers? *In InterNoise 2013*, pages 3081–3090. (Cité en page 29.)
66. DIEDERICH, A. et COLONIUS, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Attention, Perception, & Psychophysics*, 66(8):1388–1404. (Cité en page 38.)
67. DINGLER, T., LINDSAY, J. et WALKER, B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. Paris, France. Proceedings of the International Conference on Auditory Display. (Cité en pages 25 et 26.)
68. DINGUS, T. A., KLAUER, S. G., NEALE, V. L., PETERSEN, A., LEE, S. E., SUDWEEKS, J. D., PEREZ, M. A., HANKEY, J., RAMSEY, D. J. et GUPTA, S. (2006). The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment. Rapport technique. (Cité en pages 3, 4, 69, et 199.)
69. DOBASHI, Y., YAMAMOTO, T. et NISHITA, T. (2003). Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics. *ACM Transactions on Graphics (TOG)*, 22(3):732–740. (Cité en page 50.)
70. DOBASHI, Y., YAMAMOTO, T. et NISHITA, T. (2004). Synthesizing sound from turbulent field using sound textures for interactive fluid simulation. *In Computer Graphics Forum*, volume 23, pages 539–545. Wiley Online Library. (Cité en page 50.)
71. DOEL, K. V. D. (2005). Physically based models for liquid sounds. *ACM Transactions on Applied Perception (TAP)*, 2(4):534–546. (Cité en page 50.)
72. DOYLE, M. C. et SNOWDEN, R. J. (2001). Identification of Visual Stimuli is Improved by Accompanying Auditory Stimuli: The Role of Eye Movements and Sound Location. *Perception*, 30(7):795–810. (Cité en page 38.)
73. DREWS, F. A., PASUPATHI, M. et STRAYER, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14(4):392–400. (Cité en page 74.)
74. DUBOSE, C. N., CARDELLO, A. V. et MALLER, O. (1980). Effects of Colorants and Flavorants on Identification, Perceived Flavor Intensity, and Hedonic Quality of Fruit-Flavored Beverages and Cake. *Journal of Food Science*, 45(5):1393–1399. (Cité en page 39.)

75. DUNCAN, J., MARTENS, S. et WARD, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, 387(6635):808. (Cité en pages vii, 42, 43, et 59.)
76. DUNCAN, J., WILLIAMS, P. et BROWN, I. (1991). Components of driving skill: experience does not mean expertise. *Ergonomics*, 34(7):919–937. (Cité en page 68.)
77. DYER, J., STAPLETON, P. et RODGER, M. (2016). Sonification of Movement for Motor Skill Learning in a Novel Bimanual Task: Aesthetics and Retention Strategies. pages 99–102, Canberra, Australie. Proceedings of the International Conference on Auditory Display. (Cité en page 54.)
78. EFFENBERG, A. O. (2004). Using sonification to enhance perception and reproduction accuracy of human movement patterns. *In Proceedings of the International Workshop on Interactive Sonification*, volume 2004, pages 1–5, Bielefeld, Allemagne. (Cité en page 54.)
79. EITAN, Z. et ROTHSCHILD, I. (2011). How music touches: Musical parameters and listeners’ audio-tactile metaphorical mappings. *Psychology of Music*, 39(4):449–467. (Cité en pages 51 et 116.)
80. ENGELBART, D. C. et ENGLISH, W. K. (1968). A research center for augmenting human intellect. *In Proceedings of the 1968 Fall Joint Computer Conference*, pages 395–410, San Francisco, États-Unis. ACM. (Cité en page 15.)
81. ENGSTRÖM, J., JOHANSSON, E. et ÖSTLUND, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):97–120. (Cité en pages 74, 76, 192, et 193.)
82. ERNST, M. O. et BÜLTHOFF, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–169. (Cité en pages 40 et 41.)
83. EVANS, L. et WASIELEWSKI, P. (1983). Risky driving related to driver and vehicle characteristics. *Accident Analysis & Prevention*, 15(2):121–136. (Cité en page 68.)
84. FARNELL, A. J. (2013). *Designing sound: procedural audio research based on the book by Andy Farnell*. Thèse de doctorat. (Cité en page 48.)
85. FISHER, D. L., RIZZO, M., CAIRD, J. et LEE, J. D. (2011). *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*. CRC Press. (Cité en pages 69 et 72.)
86. FITTS, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381. (Cité en page 93.)
87. FORSAN, R. (2016). Etude et développement de processus de synthèse numérique pour les sons d’origine liquide. Rapport technique, Aix-Marseille Université. (Cité en page 50.)
88. FREED, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, 87(1):311–322. (Cité en page 46.)
89. GANDEMER, L. (2016). *Son et posture : le rôle de la perception auditive spatiale dans le maintien de l’équilibre postural*. Thèse de doctorat, Aix-Marseille Université. (Cité en page 35.)

90. GANDEMER, L., PARSEIHIAN, G., BOURDIN, C. et KRONLAND-MARTINET, R. (2016). Sound and Posture: an Overview of Recent Findings. *In Proceedings of the 12th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, CMMR 2016, São Paulo, Brazil. (Cité en page 35.)
91. GANDEMER, L., PARSEIHIAN, G., BOURDIN, C. et KRONLAND-MARTINET, R. (2018). Perception of Surrounding Sound Source Trajectories in the Horizontal Plane: A Comparison of VBAP and Basic-Decoded HOA. *Acta Acustica united with Acustica*, 104(2):338–350. (Cité en page 10.)
92. GANDEMER, L., PARSEIHIAN, G., KRONLAND-MARTINET, R. et BOURDIN, C. (2017). Spatial Cues Provided by Sound Improve Postural Stabilization: Evidence of a Spatial Auditory Map? *Frontiers in Neuroscience*, 11(357). (Cité en page 10.)
93. GARZONIS, S., JONES, S., JAY, T. et O’NEILL, E. (2009). Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1513–1522, Boston, États-Unis. ACM. (Cité en page 25.)
94. GAVER, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-computer interaction*, 2(2):167–177. (Cité en page 25.)
95. GAVER, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4(1):67–94. (Cité en page 25.)
96. GAVER, W. W. (1993a). How Do We Hear in the World? Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285–313. (Cité en pages 5, 48, et 50.)
97. GAVER, W. W. (1993b). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology*, 5(1):1–29. (Cité en pages 5, 44, et 45.)
98. GAVER, W. W. et NORMAN, D. A. (1988). *Everyday listening and auditory icons*. Thèse de doctorat, University of California, San Diego, Department of Cognitive Science and Psychology. (Cité en page 46.)
99. GHEZ, C. et SAINBURG, R. (1995). Proprioceptive control of interjoint coordination. *Canadian journal of physiology and pharmacology*, 73(2):273–284. (Cité en page 33.)
100. GIARD, M. H. et PERONNET, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, 11(5):473–490. (Cité en page 38.)
101. GIBSON, J. J. (1961). Ecological optics. *Vision Research*, 1(3):253–262. (Cité en pages 44 et 45.)
102. GIBSON, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin. (Cité en pages 44, 45, et 83.)
103. GIORDANO, B. L. et MCADAMS, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119(2):1171–1181. (Cité en pages 46 et 127.)

104. GIORDANO, B. L., ROCCHESSE, D. et MCADAMS, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of experimental psychology: human perception and performance*, 36(2):462. (Cité en page 46.)
105. GLASBERG, B. R. et MOORE, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2):103–138. (Cité en page 91.)
106. GODØY, R. I., HAGA, E. et JENSENIUS, A. R. (2006). Exploring music-related gestures by sound-tracing: A preliminary study. (Cité en page 53.)
107. GOINA, M. et POLOTTI, P. (2008). Elementary Gestalts for Gesture Sonification. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 150–153, Gênes, Italie. (Cité en page 53.)
108. GOPHER, D. et BRAUNE, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, 26(5):519–532. (Cité en page 67.)
109. GORDON, J., GHILARDI, M. F. et GHEZ, C. (1995). Impairments of reaching movements in patients without proprioception. I. Spatial errors. *Journal of neurophysiology*, 73(1):347–360. (Cité en page 33.)
110. GOTTFRIED, J. A. et DOLAN, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron*, 39(2):375–386. (Cité en page 38.)
111. GREEN, A. M. et ANGELAKI, D. E. (2003). Resolution of sensory ambiguities for gaze stabilization requires a second neural integrator. *Journal of Neuroscience*, 23(28):9265–9275. (Cité en page 38.)
112. GREEN, A. M. et ANGELAKI, D. E. (2010). Multisensory integration: resolving sensory ambiguities to build novel representations. *Current opinion in neurobiology*, 20(3):353–360. (Cité en page 38.)
113. GREEN, P. (1999). Visual and task demands of driver information systems. Rapport technique UMTRI-98-16, The University of Michigan Transportation Research Institute (UMTRI). (Cité en page 72.)
114. GREEN, P. (2005). How Driving Simulator Data Quality Can Be Improved. In *Proceedings of the Driving Simulator Conference*, page 11, Orlando, États-Unis. (Cité en page 72.)
115. GUEST, G., MACQUEEN, K. M. et NAMEY, E. E. (2011). *Applied Thematic Analysis*. Sage Publications. (Cité en page 119.)
116. GUNA, J., JAKUS, G., POGAČNIK, M., TOMAŽIČ, S. et SODNIK, J. (2014). An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking. *Sensors*, 14(12):3702–3720. (Cité en page 81.)
117. HAAS, E. C. et EDWORTHY, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computing & Control Engineering Journal*, 7(4):193–198. (Cité en page 26.)

118. HAMISH JAMSON, A. et MERAT, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):79–96. (Cit  en pages 75, 192, et 193.)
119. HANCOCK, P. A., LESCH, M. et SIMMONS, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis & Prevention*, 35(4):501–514. (Cit  en page 73.)
120. HARBLUK, J. L., NOY, Y. I. et EIZENMAN, M. (2002). The impact of cognitive distraction on driver visual behaviour and vehicle control. Rapport technique. (Cit  en page 73.)
121. HART, S. G. et STAVELAND, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier. (Cit  en page 67.)
122. HAY, J. C., PICK, H. L. et IKEDA, K. (1965). Visual capture produced by prism spectacles. *Psychonomic science*, 2(1-12):215–216. (Cit  en page 39.)
123. HEMENWAY, K. (1982). Psychological issues in the use of icons in command menus. In *Proceedings of the 1982 conference on Human factors in computing systems*, pages 20–23. ACM. (Cit  en page 24.)
124. HERMANN, T. (2008). Taxonomy and definitions for sonification and auditory display. In *Proceedings of the International Conference on Auditory Display*, Paris, France. (Cit  en page 27.)
125. HERMANN, T. et HUNT, A. (2004). The discipline of interactive sonification. In *Proceedings of the International Workshop on Interactive Sonification*. (Cit  en page 27.)
126. HERMANN, T. et HUNT, A. (2005). An Introduction to Interactive Sonification. *IEEE MultiMedia*, 12(2):20–24. (Cit  en page 27.)
127. HERSHENSON, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, 63(3):289–293. (Cit  en page 38.)
128. HORREY, W. J. et WICKENS, C. D. (2006). Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(1):196–205. (Cit  en page 73.)
129. HOUBEN, M. M., KOHLRAUSCH, A. et HERMES, D. (2001). Auditory cues determining the perception of the size and speed of rolling balls. In *Proceedings of the International Conference on Auditory Display*, pages 105–110, Helsinki, Finlande. (Cit  en pages 102 et 201.)
130. HOUBEN, M. M., KOHLRAUSCH, A. et HERMES, D. J. (2004). Perception of the size and speed of rolling balls by sound. *Speech Communication*, 43(4):331–345. (Cit  en pages 47, 91, 102, et 201.)
131. HOUBEN, M. M. J. (2002). *The sound of rolling objects: perception of size and speed*. Th se de doctorat, Technische Universiteit, Eindhoven. (Cit  en pages 47, 83, 84, et 201.)

132. HURWITZ, J. B. et WHEATLEY, D. J. (2002). Using Driver Performance Measures to Estimate Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(22):1804–1808. (Cité en page 75.)
133. JACOB, Y., MANITSARIS, S., MOUTARDE, F., LELE, G. et PRADERE, L. (2015). Hand gesture recognition for driver vehicle interaction. In *IEEE Computer Society Workshop on Observing and understanding hands in action (Hands 2015)*, Boston, États-Unis. (Cité en page 22.)
134. Jacobé de NAUROIS, C., BOURDIN, C., STRATULAT, A., DIAZ, E. et VERCHER, J.-L. (2017). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention*. (Cité en page 70.)
135. JASCHINSKI, L., DENJEAN, S., PETIOT, J. F., MARS, F. et ROUSSARIE, V. (2016). Impact of interface sonification with touchless gesture command in a car. In *Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference*, pages in–press. (Cité en pages 23 et 29.)
136. JENSENIUS, A. R., WANDERLEY, M. M., GODØY, R. I. et LEMAN, M. (2009). Concepts and Methods in Research. In *Musical gestures: Sound, movement, and meaning*, volume 12, pages 12–35. Routledge. (Cité en page 53.)
137. JOFFE, H. (2012). Thematic analysis. *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners*, 1:210–23. (Cité en page 119.)
138. JOHNSTON, E. B., CUMMING, B. G. et LANDY, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision research*, 34(17):2259–2275. (Cité en page 40.)
139. KALYUGA, S., AYRES, P., CHANDLER, P. et SWELLER, J. (2003). The expertise reversal effect. *Educational psychologist*, 38(1):23–31. (Cité en pages 63 et 208.)
140. KANE, S. K., MORRIS, M. R. et WOBROCK, J. O. (2013). Touchplates: low-cost tactile overlays for visually impaired touch screen users. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–8, Bellevue, États-Unis. ACM Press. (Cité en page 54.)
141. KAPTEIN, N., THEEUWES, J. et VAN DER HORST, R. (1996). Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board*, (1550):30–36. (Cité en pages 70 et 71.)
142. KENDALL, G. S. (1991). Visualization by ear: Auditory imagery for scientific visualization and virtual reality. *Computer Music Journal*, 15(4):70–73. (Cité en page 27.)
143. KJÆR, H. P., TAYLOR, C. C. et SERAFIN, S. (2007). Influence of Interactive Auditory Feedback on the Haptic Perception of Virtual Objects. In *Proceedings of the International Workshop on Interactive Sonification*, page 4, York, Royaume-Uni. (Cité en page 38.)
144. KLATZKY, R. L., PAI, D. K. et KROTKOV, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410. (Cité en pages 46 et 127.)

145. KOHLER, E., KEYSERS, C., UMITA, M. A., FOGASSI, L., GALLESE, V. et RIZZOLATTI, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582):846–848. (Cité en page 45.)
146. KOLERS, P. A. (1969). Some formal characteristics of pictograms. *American scientist*, 57(3): 348–363. (Cité en page 24.)
147. KOPPEN, C., LEVITAN, C. A. et SPENCE, C. (2009). A signal detection study of the Colavita visual dominance effect. *Experimental brain research*, 196(3):353–360. (Cité en pages 42, 144, 147, et 202.)
148. KOPPEN, C. et SPENCE, C. (2007). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain research*, 1186:224–232. (Cité en pages 42, 144, 147, et 202.)
149. KRAMER, G., WALKER, B., BONEBRIGHT, T., COOK, P., FLOWERS, J. H., MINER, N. et NEUHOFF, J. (1999). Sonification report: Status of the field and research agenda. In *Proceedings of the International Conference on Auditory Display*, Santa Fe, États-Unis. (Cité en pages 5, 24, 26, et 27.)
150. KRIPPENDORFF, K. (2012). *Content Analysis: An Introduction to Its Methodology*. Sage Publications. (Cité en page 119.)
151. KRUEGER, M. W., GIONFRIDDO, T. et HINRICHSEN, K. (1985). Videoplacement - an Artificial Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '85, pages 35–40, New York, États-Unis. ACM. (Cité en page 19.)
152. KURTENBACH, G. et HULTEEN, E. (1990). Gestures in Human-Computer Communications. *The Art of Human Computer Interface Design*, pages 309–317. (Cité en pages 17 et 53.)
153. LAKATOS, S., MCADAMS, S. et CAUSSÉ, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Attention, Perception, & Psychophysics*, 59(8):1180–1190. (Cité en page 46.)
154. LAMBLE, D., KAURANEN, T., LAAKSO, M. et SUMMALA, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis & Prevention*, 31(6):617–623. (Cité en pages 3, 4, 75, 192, et 193.)
155. LEMAITRE, G., HOUIX, O., MISDARIIS, N. et SUSINI, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1):16–32. (Cité en pages 44 et 127.)
156. LEMAITRE, G., SCURTO, H., FRANÇOISE, J., BEVILACQUA, F., HOUIX, O. et SUSINI, P. (2017). Rising tones and rustling noises: Metaphors in gestural depictions of sounds. *PLOS ONE*, 12(7):e0181786. (Cité en pages 53 et 159.)
157. LOEHMANN, S., KNOBEL, M., LAMARA, M. et BUTZ, A. (2013). Culturally Independent Gestures for In-Car Interactions. In KOTZÉ, P., MARSDEN, G., LINDGAARD, G., WESSON, J. et WINCKLER, M., éditeurs : *Proceedings of the IFIP International Conference*, Lecture Notes in Computer Science, pages 538–545, Cape Town, Afrique du Sud. Springer Berlin Heidelberg. (Cité en page 23.)

158. LUTFI, R. A. et LIU, C.-J. (2007). Individual differences in source identification from synthesized impact sounds. *The Journal of the Acoustical Society of America*, 122(2):1017–1028. (Cité en page 46.)
159. MACDONALD, J. et MCGURK, H. (1978). Visual influences on speech perception processes. *Attention, Perception, & Psychophysics*, 24(3):253–257. (Cité en page 39.)
160. MACK, A. et ROCK, I. (1999). Inattention blindness: an overview. In *Inattention blindness*. MIT press Cambridge, MA. (Cité en pages 2, 41, 42, et 201.)
161. MACKENZIE, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1):91–139. (Cité en page 93.)
162. MAHR, A., ENDRES, C., MÜLLER, C. et SCHNEEBERGER, T. (2011). Determining human-centered parameters of ergonomic micro-gesture interaction for drivers using the theater approach. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 151–158, Salzburg, Autriche. ACM. (Cité en pages vii et 22.)
163. MAQUET, P. (2001). The Role of Sleep in Learning and Memory. *Science*, 294(5544):1048–1052. (Cité en page 169.)
164. MARKS, D. F. et YARDLEY, L. (2004). *Research Methods for Clinical and Health Psychology*. Sage Publications. (Cité en page 119.)
165. MATHEWS, M. (1989). The Radio Drum as a Synthesizer Controller. In *Proceedings of the 1989 International Computer Music Conference*. (Cité en page 53.)
166. MATHEWS, M. et SMITH, J. O. (2003). Methods for synthesizing very high Q parametrically well behaved two pole filters. In *Proceedings of the Stockholm Musical Acoustics Conference*, Stockholm, Suède. Royal Swedish Academy of Music. (Cité en page 85.)
167. MATTHEWS, R., LEGG, S. et CHARLTON, S. (2003). The effect of cell phone type on drivers subjective workload during concurrent driving and conversing. *Accident; Analysis and Prevention*, 35(4):451–457. (Cité en page 73.)
168. MCADAMS, S. (1993). Recognition of sound sources and events. In *Thinking in Sound: The Cognitive Psychology of Human Audition*, pages 146–198. New York, États-Unis. (Cité en page 46.)
169. MCEVOY, S. P., STEVENSON, M. R., MCCARTT, A. T., WOODWARD, M., HAWORTH, C., PALAMARA, P. et CERCARELLI, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ*. (Cité en page 42.)
170. MEISTER, D. (1976). *Behavioral foundations of system development*. New York, États-Unis. (Cité en page 66.)
171. MERER, A. (2011). *Caractérisation acoustique et perceptive du mouvement évoqué par les sons pour le contrôle de la synthèse*. Thèse de doctorat, Université de Provence-Aix-Marseille I. (Cité en pages 51, 52, et 116.)

172. MERER, A., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transactions on Applied Perception (TAP)*, 10(1):1–24. (Cité en page 51.)
173. MERER, A., YSTAD, S., KRONLAND-MARTINET, R. et ARAMAKI, M. (2008). Semiotics of sounds evoking motions: categorization and acoustic features. In KRONLAND-MARTINET, R., JENSEN, K. et YSTAD, S., éditeurs : *Computer music modeling andretrieval: sense of sounds*, Lecture Notes in Computer Science, pages 139–158. Springer Berlin / Heidelberg. (Cité en page 51.)
174. MERFELD, D. M., ZUPAN, L. et PETERKA, R. J. (1999). Humans use internal models to estimate gravity and linear acceleration. *Nature*, 398(6728):615. (Cité en page 38.)
175. MICHAELS, C. F. et CARELLO, C. (1981). *Direct Perception*. Century psychology series. Prentice-Hall, Englewood Cliffs, N.J. (Cité en pages 43 et 46.)
176. MILES, M. B. et HUBERMAN, A. M. (1994). *Qualitative data analysis: An expanded source-book*. Sage Publications. (Cité en page 119.)
177. MILLER, G. (1956). *The magical number seven, plus or minus two: Some limits on our capacity to process information*, volume 63. (Cité en pages 61, 155, 156, et 206.)
178. MONNOYER, J., DIAZ, E., BOURDIN, C. et WIERTLEWSKI, M. (2016). Ultrasonic Friction Modulation While Pressing Induces a Tactile Feedback. In BELLO, F., KAJIMOTO, H. et VISELL, Y., éditeurs : *Haptics: Perception, Devices, Control, and Applications*, Lecture Notes in Computer Science, pages 171–179, Londres, Royaume-Uni. Springer International Publishing. (Cité en page 20.)
179. MOORE, B. C. (1973). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America*, 54(3):610–619. (Cité en pages 84 et 102.)
180. MORAY, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, 11(1):56–60. (Cité en page 42.)
181. MOUSAVI, S. Y., LOW, R. et SWELLER, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology*, 87(2):319. (Cité en pages 61, 79, 130, et 193.)
182. MUCKLER, F. A. et SEVEN, S. A. (1992). Selecting performance measures: “Objective” versus “subjective” measurement. *Human factors*, 34(4):441–455. (Cité en page 67.)
183. MUDD, S. (1968). Assessment of the Fidelity of Dynamic Flight Simulators. *Human Factors*, 10(4):351–358. (Cité en page 70.)
184. MULLEN, N., CHARLTON, J., DEVLIN, A. et BEDARD, M. (2011). Simulator Validity: Behaviors Observed on the Simulator and on the Road. In *Handbook of Driving Simulation for Engineering, Medicine and Psychology*, pages 1–18. CRC Press, Boca Raton, États-Unis. (Cité en page 72.)
185. MURRAY, M. M., MOLHOLM, S., MICHEL, C. M., HESLENFELD, D. J., RITTER, W., JAVITT, D. C., SCHROEDER, C. E. et FOXE, J. J. (2005). Grabbing Your Ear: Rapid Auditory–Somatosensory Multisensory Interactions in Low-level Sensory Cortices Are Not Constrained by Stimulus Alignment. *Cerebral Cortex*, 15(7):963–974. (Cité en page 38.)

186. NEISSER, U. (1979). The control of information pickup in selective looking. *Perception and its Development: A Tribute to Eleanor J Gibson*, pages 201 – 2019. (Cit  en page 42.)
187. NYMOEN, K., CARAMIAUX, B., KOZAK, M. et TORRESEN, J. (2011). Analyzing Sound Tracings: A Multimodal Approach to Music Information Retrieval. In *Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, MIRUM '11, pages 39–44, Scottsdale,  tats-Unis. ACM. (Cit  en page 159.)
188. OCCELLI, V., SPENCE, C. et ZAMPINI, M. (2009). Compatibility effects between sound frequency and tactile elevation. *NeuroReport*, 20(8):793–797. (Cit  en page 159.)
189. OHN-BAR, E., TRAN, C. et TRIVEDI, M. (2012). Hand Gesture-based Visual User Interface for Infotainment. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '12, pages 111–115, New York,  tats-Unis. ACM. (Cit  en page 23.)
190. OHN-BAR, E. et TRIVEDI, M. M. (2014). Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 15(6):2368–2377. (Cit  en page 23.)
191. PAAS, F., TUOVINEN, J. E., TABBERS, H. et VAN GERVEN, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1): 63–71. (Cit  en pages 64 et 67.)
192. PAAS, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4):429–434. (Cit  en page 67.)
193. PAAS, F. G. et VAN MERRIENBOER, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743. (Cit  en page 68.)
194. PAAS, F. G. et VAN MERRIENBOER, J. J. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review*, 6(4):351–371. (Cit  en page 61.)
195. PAAS, F. G. et VAN MERRIENBOER, J. J. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology*, 86(1):122. (Cit  en page 66.)
196. PAPANTONIOU, P., PAPADIMITRIOU, E. et YANNIS, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia*, 25: 1796–1805. (Cit  en page 67.)
197. PARSEIHIAN, G. (2012). *Sonification binaurale pour l'aide   la navigation*. Th se de doctorat, Universit  Pierre et Marie Curie-Paris VI. (Cit  en page 27.)
198. PARSEIHIAN, G., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2017). Sonification strategies for dynamic guidance tasks: example with a driving game. In *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 283–294, Matosinhos, Portugal. (Cit  en page 10.)

199. PARSEIHIAN, G., GANDEMER, L., BOURDIN, C. et KRONLAND-MARTINET, R. (2015). Design and perceptual evaluation of a fully immersive three-dimensional sound spatialization system. *In 3rd International Conference on Spatial Audio (ICSA 2015)*, Graz, Autriche. (Cit  en page 10.)
200. PARSEIHIAN, G., GONDRE, C., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2016). Comparison and Evaluation of Sonification Strategies for Guidance Tasks. *IEEE Transactions on Multimedia*, 18(4):674–686. (Cit  en pages 10, 84, 85, et 126.)
201. PARSEIHIAN, G. et KATZ, B. F. (2012). Morphocons: A new sonification concept based on morphological earcons. *Journal of the Audio Engineering Society*, 60(6):409–418. (Cit  en page 26.)
202. PATTEN, C. J. D., KIRCHER, A., OSTLUND, J. et NILSSON, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident; Analysis and Prevention*, 36(3):341–350. (Cit  en page 73.)
203. PAVANI, F., SPENCE, C. et DRIVER, J. (2000). Visual Capture of Touch: Out-of-the-Body Experiences With Rubber Gloves. *Psychological Science*, 11(5):353–359. (Cit  en page 39.)
204. PENNEY, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, 17(4):398–422. (Cit  en pages 61, 79, 130, et 193.)
205. PETIOT, J.-F., KRISTENSEN, B. G. et MAIER, A. M. (2013). How Should an Electric Vehicle Sound? User and Expert Perception. *In Proceedings of the ASME 2013 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, page V005T06A028, Portland,  tats-Unis. ASME. (Cit  en page 116.)
206. PICCOLO, A. D., ROCCHESO, D. et PAPETTI, S. (2018). Path Following in Non-Visual Conditions. *IEEE Transactions on Haptics*. (Cit  en page 27.)
207. PIZZIGHELLO, S. et BRESSAN, P. (2008). Auditory Attention Causes Visual Inattentive Blindness. *Perception*, 37(6):859–866. (Cit  en pages 42 et 144.)
208. POLLOCK, E., CHANDLER, P. et SWELLER, J. (2002). Assimilating complex information. *Learning and instruction*, 12(1):61–86. (Cit  en pages 61, 62, 63, et 170.)
209. POSNER, M. I., NISSEN, M. J. et KLEIN, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological review*, 83(2):157. (Cit  en pages 42, 144, et 147.)
210. PRUVOST, L., SCHERRER, B., ARAMAKI, M., YSTAD, S. et KRONLAND-MARTINET, R. (2015). Perception-based interactive sound synthesis of morphing solids’ interactions. *In Proceedings of the SIGGRAPH Conference*, pages 1–4, Kobe, Japon. ACM Press. (Cit  en page 10.)
211. PULKKI, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466. (Cit  en page 81.)
212. QUENAULT, S. W. et PARKER, P. M. (1973). Driver behaviour-newly qualified drivers. Rapport technique, Transport and Road Research Laboratory, Crowthorne, Royaume-Uni. (Cit  en page 68.)

213. QUIMBY, A. R. et WATTS, G. R. (1981). Human factors and driving performance. Rapport technique. (Cité en page 68.)
214. RAO, R. P. N., OLSHAUSEN, B. A. et LEWICKI, M. S. (2002). *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press. (Cité en pages 40 et 41.)
215. RATH, M. et ROCCHESO, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE MultiMedia*, 12(2):60–69. (Cité en pages 55 et 83.)
216. RATH, M. et SCHLEICHER, R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta acustica united with acustica*, 94(1):12–20. (Cité en pages xi, 55, 56, 58, 102, 104, 105, et 145.)
217. RECARTE, M. A. et NUNES, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of experimental psychology: Applied*, 6(1):31. (Cité en page 73.)
218. REGAN, M. A., LEE, J. D. et YOUNG, K. (2008). *Driver Distraction: Theory, Effects, and Mitigation*. CRC Press. (Cité en page 68.)
219. REID, G. B., SHINGLEDECKER, C. A. et EGGEMEIER, F. T. (1981). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society Annual Meeting*, volume 25, pages 522–526. Sage Publications. (Cité en page 67.)
220. REPP, B. H. (1987). The sound of two hands clapping: An exploratory study. *The Journal of the Acoustical Society of America*, 81(4):1100–1109. (Cité en page 46.)
221. RIEMANN, B. L. et LEPHART, S. M. (2002). The sensorimotor system, part I: the physiologic basis of functional joint stability. *Journal of athletic training*, 37(1):71. (Cité en page 33.)
222. RIENER, A., FERSCHA, A., BACHMAIR, F., HAGMÜLLER, P., LEMME, A., MUTTENTHALER, D., PÜHRINGER, D., ROGNER, H., TAPPE, A. et WEGER, F. (2013). Standardization of the In-car Gesture Interaction Space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 14–21, New York, États-Unis. ACM. (Cité en page 23.)
223. RIMÉ, B. et SCHIARATURA, L. (1991). Gesture and speech. In *Fundamentals of nonverbal behavior*, pages 239 – 281. New York : Cambridge University Press, r. s. feldman & b. rimé (eds.) édition. (Cité en page 17.)
224. RISSET, J.-C. (1978). Paradoxes de hauteur. Rapport technique, IRCAM, Paris, France. (Cité en page 29.)
225. RISSET, J.-C. et WESSEL, D. L. (1999). Exploration of Timbre by Analysis and Synthesis. In DEUTSCH, D., éditeur : *The Psychology of Music (Second Edition)*, Cognition and Perception, pages 113–169. Academic Press, San Diego. (Cité en page 48.)
226. ROCCHESO, D., LEMAITRE, G., SUSINI, P., TERNSTRÖM, S. et BOUSSARD, P. (2015). Sketching Sound with Voice and Gesture. *Interactions*, 22(1):38–41. (Cité en page 53.)
227. ROLFE, J. M., HAMMERTON-FRASEK, A., POULTER, R. F. et SMITH, E. M. B. (1970). Pilot response in flight and simulated flight. *Ergonomics*, 13(6):761–768. (Cité en page 70.)

228. ROSATI, G., OSCARI, F., SPAGNOL, S., AVANZINI, F. et MASIERO, S. (2012). Effect of task-related continuous auditory feedback during learning of tracking motion exercises. *Journal of Neuroengineering and Rehabilitation*, 9(79). (Cit  en pages 54, 58, et 193.)
229. ROSENBLUM, L. D., CARELLO, C. et PASTORE, R. E. (1987). Relative Effectiveness of Three Stimulus Variables for Locating a Moving Sound Source. *Perception*, 16(2):175–186. (Cit  en page 37.)
230. ROUSSARIE, V. (1999). *Analyse perceptifve de structures vibrantes simul es par mod le physique*. Th se de doctorat, Universit  du Maine. (Cit  en pages 46 et 127.)
231. ROZ , J. (2017). *L'influence des mouvements posturaux des violoncellistes sur leur expressivit  musicale*. Th se de doctorat, Aix-Marseille Universit . (Cit  en page 53.)
232. ROZ , J., KRONLAND-MARTINET, R., ARAMAKI, M., BOURDIN, C. et YSTAD, S. (2016). Investigating the effects of a postural constraint on the cellists' bowing movement and timbral quality. In *Proceedings of the 12th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 83–98, S o Paulo, Br sil. (Cit  en pages 10 et 53.)
233. SACHS, C. (2012). *The History of Musical Instruments*. Dover Publications, New York,  tats-Unis. (Cit  en pages 5 et 53.)
234. SAINBURG, R. L., POIZNER, H. et GHEZ, C. (1993). Loss of proprioception produces deficits in interjoint coordination. *Journal of neurophysiology*, 70(5):2136–2147. (Cit  en page 33.)
235. SANTOS, J., MERAT, N., MOUTA, S., BROOKHUIS, K. et de WAARD, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):135–146. (Cit  en pages 71, 72, 162, et 191.)
236. SARLEGNA, F. R., MALFAIT, N., BRINGOUX, L., BOURDIN, C. et VERCHER, J.-L. (2010). Force-field adaptation without proprioception: Can vision be used to model limb dynamics? *Neuropsychologia*, 48(1):60–67. (Cit  en page 33.)
237. SAVONA, F. (2016). *La perception des acc l rations lat rales en simulateur de conduite :  tude de l'int gration multi-sensorielle pour l'am lioration des performances de simulation*. Th se de doctorat, Aix-Marseille Universit . (Cit  en page 70.)
238. SCHAEFFER, P. (1966). *Trait  des objets musicaux*. Le Seuil, Paris, France. (Cit  en page 51.)
239. SCHAFFERT, N. et MATTES, K. (2012). Acoustic feedback training in adaptive rowing. In *Proceedings of the International Conference on Auditory Display*, Atlanta,  tats-Unis. Georgia Institute of Technology. (Cit  en pages 27 et 54.)
240. SCIABICA, J.-F. (2011). *Caract risation acoustique et perceptifve du bruit moteur dans un habitacle automobile*. Th se de doctorat, Aix-Marseille 1. (Cit  en page 70.)
241. SENDERS, J. W., KRISTOFFERSON, A. B., LEVISON, W. H., DIETRICH, C. W. et WARD, J. L. (1967). The attentional demande of automobile driving. *Highway Research Record*, (195). (Cit  en page 72.)

242. SERAFIN, S., FRANINOVIĆ, K., HERMANN, T., LEMAITRE, G., RINOTT, M. et ROCCHESO, D. (2011). Sonic Interaction Design. In *The Sonification Handbook*. Berlin, logos publishing house édition. (Cité en page 27.)
243. SHAMS, L., KAMITANI, Y. et SHIMOJO, S. (2000). Illusions: What you see is what you hear. *Nature*, 408(6814):788. (Cité en page 39.)
244. SHAMS, L., KAMITANI, Y. et SHIMOJO, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1):147–152. (Cité en page 39.)
245. SHAMS, L., KAMITANI, Y., THOMPSON, S. et SHIMOJO, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, 12(17):3849–3852. (Cité en page 39.)
246. SHECHTMAN, O., CLASSEN, S., AWADZI, K. et MANN, W. (2009). Comparison of Driving Errors Between On-the-Road and Simulated Driving Assessment: A Validation Study. *Traffic Injury Prevention*, 10(4):379–385. (Cité en page 70.)
247. SHEPARD, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, 36(12):2346–2353. (Cité en page 29.)
248. SHERRINGTON, C. (1910). *The integrative action of the nervous system*. CUP Archive. (Cité en page 33.)
249. SHOVE, P. et REPP, B. H. (1995). Musical motion and performance: theoretical and empirical perspectives. *The practice of performance: Studies in musical interpretation*, pages 55–83. (Cité en pages 51 et 116.)
250. SIMONS, D. J. et CHABRIS, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception*, 28(9):1059–1074. (Cité en page 42.)
251. SORS, F., MURGIA, M., SANTORO, I. et AGOSTINI, T. (2015). Audiobased interventions in sport. *Open Psychol. J*, 8(3):212–219. (Cité en page 54.)
252. SOTO-FARACO, S., LYONS, J., GAZZANIGA, M., SPENCE, C. et KINGSTONE, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive brain research*, 14(1):139–146. (Cité en page 39.)
253. SPENCE, C. (2009). Explaining the Colavita visual dominance effect. *Progress in brain research*, 176:245–258. (Cité en pages 42, 144, 147, et 202.)
254. SPENCE, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4):971–995. (Cité en page 159.)
255. SPENCE, C. et DRIVER, J. (1997). On measuring selective attention to an expected sensory modality. *Attention, Perception, & Psychophysics*, 59(3):389–403. (Cité en page 42.)
256. STICKGOLD, R., JAMES, L. et HOBSON, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature neuroscience*, 3(12):1237. (Cité en page 169.)
257. STOELINGA, C. et CHAIGNE, A. (2007). Time-domain modeling and simulation of rolling objects. *Acta Acustica united with Acustica*, 93(2):290–304. (Cité en page 49.)

258. STRATULAT, A. (2011). *Etude des interactions multisensorielle pour la perception des mouvements du véhicule en simulateur dynamique : contribution de l'illusion somatogravique à l'immersion en environnement virtuel*. Thèse de doctorat, Aix-Marseille 2. (Cité en page 70.)
259. STRATULAT, A., ROUSSARIE, V., VERCHER, J.-L. et BOURDIN, C. (2011). Improving the realism in motion-based driving simulators by adapting tilt-translation technique to human perception. *In Proceedings of the Virtual Reality Conference*, pages 47–50, Singapour, Singapour. IEEE. (Cité en page 70.)
260. STRAYER, D. L. et DREW, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human factors*, 46(4):640–649. (Cité en page 74.)
261. SUSINI, P., MISDARIIS, N., LEMAITRE, G. et HOUIX, O. (2012). Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal on Multimodal User Interfaces*, 5(3):175–186. (Cité en page 153.)
262. SUSINI, P., MISDARIIS, N., LEMAITRE, G., HOUIX, O., ROCCHESO, D., POLOTTI, P., FRANINOVIC, K., VISELL, Y., OBERMAYER, K. et PURWINS, H. (2006). Closing the loop of sound evaluation and design. *In Proceedings of the 2nd ISCA/DEGA Tutorial & Research Workshop on Perceptual Quality of Systems*, Berlin, Allemagne. (Cité en page 116.)
263. SWELLER, J., VAN MERRIENBOER, J. J. et PAAS, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296. (Cité en pages 61 et 62.)
264. TAGG, P. (2005). Musical meanings, classical and popular. The case of anguish. (Cité en page 116.)
265. TARDIEU, J., MISDARIIS, N., LANGLOIS, S., GAILLARD, P. et LEMERCIER, C. (2015). Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *Applied Ergonomics*, 50:41–49. (Cité en page 29.)
266. THORET, E. (2014). *Caractérisation acoustique des relations entre les mouvements biologiques et la perception sonore: application au contrôle de la synthèse et à l'apprentissage de gestes*. Thèse de doctorat, Aix-Marseille Université. (Cité en pages 39, 46, 54, 58, 83, et 102.)
267. THORET, E., ARAMAKI, M., BRINGOUX, L., YSTAD, S. et KRONLAND-MARTINET, R. (2016a). Seeing Circles and Drawing Ellipses: When Sound Biases Reproduction of Visual Motion. *PLOS ONE*, 11(4):e0154475. (Cité en pages 10, 54, et 55.)
268. THORET, E., ARAMAKI, M., BRINGOUX, L., YSTAD, S. et KRONLAND-MARTINET, R. (2016b). When eyes drive hand: Influence of non-biological motion on visuo-motor coupling. *Neuroscience Letters*, 612:225–230. (Cité en page 10.)
269. THORET, E., ARAMAKI, M., GONDRE, C., KRONLAND-MARTINET, R. et YSTAD, S. (2013). Controlling a non linear friction model for evocative sound synthesis applications. *In International Conference on Digital Audio Effects*, pages 1–7, Maynooth, Irlande. (Cité en page 49.)
270. THURLOW, W. R. et RUNGE, P. S. (1967). Effect of Induced Head Movements on Localization of Direction of Sounds. *The Journal of the Acoustical Society of America*, 42(2):480–488. (Cité en page 37.)

271. TINDALL-FORD, S., CHANDLER, P. et SWELLER, J. (1997). When two sensory modes are better than one. *Journal of experimental psychology: Applied*, 3(4):257. (Cité en pages 61, 79, 130, et 193.)
272. TÖRNROS, J. (1998). Driving behaviour in a real and a simulated road tunnel—a validation study. *Accident Analysis & Prevention*, 30(4):497–503. (Cité en page 71.)
273. TÖRNROS, J. E. B. et BOLLING, A. K. (2005). Mobile phone use—Effects of handheld and handsfree phones on driving performance. *Accident Analysis & Prevention*, 37(5):902–909. (Cité en page 75.)
274. TSIMHONI, O. et GREEN, P. (2001). Visual Demand of Driving and the Execution of Display-Intensive in-Vehicle Tasks. In *Proceedings of the Human Factors and Ergonomics Society*, volume 45, pages 1586–1590. Sage Publications. (Cité en page 72.)
275. TSIMHONI, O., SMITH, D. et GREEN, P. (2004). Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard. *Human Factors*, 46(4):600–610. (Cité en pages 73 et 199.)
276. TYLÉN, K., WEED, E., WALLENTIN, M., ROEPSTORFF, A. et FRITH, C. D. (2010). Language as a Tool for Interacting Minds. *Mind & Language*, 25(1):3–29. (Cité en page 1.)
277. VAN DEN DOEL, K., KRY, P. G. et PAI, D. K. (2001). FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 537–544, New York, États-Unis. ACM. (Cité en page 49.)
278. VAN MERRIËNBOER, J. J. G., CLARK, R. E. et de CROOCK, M. B. M. (2002). Blueprints for complex learning: The 4c/ID-model. *Educational Technology Research and Development*, 50(2):39–61. (Cité en page 62.)
279. VAN MERRIËNBOER, J. J. G. et SWELLER, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2):147–177. (Cité en pages 62, 63, et 170.)
280. VAN VUGT, F. T. et TILLMANN, B. (2015). Auditory feedback in error-based learning of motor regularity. *Brain research*, 1606:54–67. (Cité en page 54.)
281. VANDERVEER, N. J. (1979). *Ecological acoustics: human perception of environmental sounds*. Dissertation Abstracts International, University Microfilms numéro 8004002. (Cité en pages 44 et 127.)
282. VERRON, C. (2010). *Synthese immersive de sons d'environnement*. Thèse de doctorat, Aix-Marseille 1. (Cité en pages 48 et 50.)
283. VERRON, C., ARAMAKI, M., KRONLAND-MARTINET, R. et PALLONE, G. (2010). A 3-D Immersive Synthesizer for Environmental Sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1550–1561. (Cité en page 50.)
284. VICTOR, T. W., HARBLUK, J. L. et ENGSTRÖM, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):167–190. (Cité en page 74.)

285. WALKER, B. N., NANCE, A. et LINDSAY, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. *In Proceedings of the International Conference on Auditory Display*, pages 63–68, Londres, Royaume-Uni. (Cité en page 26.)
286. WALKER, B. N. et NEES, M. A. (2011). Theory of sonification. *In The sonification handbook*, pages 9–39. Berlin, logos publishing house édition. (Cité en page 27.)
287. WALKER, M. P., BRAKEFIELD, T., MORGAN, A., HOBSON, J. A. et STICKGOLD, R. (2002). Practice with Sleep Makes Perfect: Sleep-Dependent Motor Skill Learning. *Neuron*, 35(1): 205–211. (Cité en page 169.)
288. WALLACH, H. (1939). On Sound Localization. *The Journal of the Acoustical Society of America*, 10(4):270–274. (Cité en page 37.)
289. WANDERLEY, M. et DEPALLE, P. (2004). Gestural Control of Sound Synthesis. *Proceedings of the IEEE*, 92(4):632–644. (Cité en page 53.)
290. WANG, Y., MEHLER, B., REIMER, B., LAMMERS, V., D’AMBROSIO, L. A. et COUGHLIN, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, 53(3):404–420. (Cité en page 71.)
291. WARREN, W. H. et VERBRUGGE, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human perception and performance*, 10(5):704. (Cité en page 46.)
292. WATANABE, K. et SHIMOJO, S. (2001). When Sound Affects Vision: Effects of Auditory Grouping on Visual Motion Perception. *Psychological Science*, 12(2):109–116. (Cité en page 39.)
293. WEGNER, K. (1998). Surgical navigation system and method using audio feedback. *In Proceedings of the International Conference on Auditory Display*, pages 1–10, Glasgow, Royaume-Uni. (Cité en page 28.)
294. WELCH, R. B., DUTTONHURT, L. D. et WARREN, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Attention, Perception, & Psychophysics*, 39(4):294–300. (Cité en page 40.)
295. WELCH, R. B. et WARREN, D. H. (1986). Intersensory interactions. *Handbook of perception and human performance*, 1:25–1. (Cité en page 40.)
296. WIERWILLE, W. W. (1993). Visual and manual demands of in-car controls and displays. *Automotive Ergonomics*, pages 229–320. (Cité en page 72.)
297. WIERWILLE, W. W. et CASALI, J. G. (1983). A validated rating scale for global mental workload measurement applications. *In Proceedings of the Human Factors society Annual Meeting*, volume 27, pages 129–133. Sage Publications. (Cité en page 67.)
298. WOBBROCK, J. O., MORRIS, M. R. et WILSON, A. D. (2009). User-defined gestures for surface computing. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1083–1092, Boston, États-Unis. ACM. (Cité en page 19.)

299. WOLF, K., NAUMANN, A., ROHS, M. et MÜLLER, J. (2011). A Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. *In* CAMPOS, P., GRAHAM, N., JORGE, J., NUNES, N., PALANQUE, P. et WINCKLER, M., éditeurs : *Proceedings of International Conference on Human-Computer Interaction*, Lecture Notes in Computer Science, pages 559–575, Lisbonne, Portugal. Springer Berlin Heidelberg. (Cité en page 22.)
300. WUERGER, S. M., HOFBAUER, M. et MEYER, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Attention, Perception, & Psychophysics*, 65(8):1188–1196. (Cité en page 40.)
301. YEH, Y.-Y. et WICKENS, C. D. (1984). Why do performance and subjective workload measures dissociate? *In Proceedings of the Human Factors Society Annual Meeting*, volume 28, pages 504–508. Sage Publications. (Cité en page 67.)
302. YOUNG, M. J., LANDY, M. S. et MALONEY, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision research*, 33(18):2685–2696. (Cité en page 40.)
303. ZIJLSTRA, F. R. H. (1985). *The Construction of a Scale to Measure Perceived Effort*. Delft University of Technology, Delft, Pays-Bas. (Cité en page 67.)
304. ZIJLSTRA, F. R. H. (1993). *Efficiency in work behaviour: A design approach for modern tools*. Thèse de doctorat, Delft University. (Cité en page 67.)
305. ZIMMERMAN, T. G., LANIER, J., BLANCHARD, C., BRYSON, S. et HARVILL, Y. (1987). 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, États-Unis. ACM. (Cité en page 20.)
306. ZOBL, M., GEIGER, M., SCHULLER, B., LANG, M. et RIGOLL, G. (2003). A Real-time System for Hand Gesture Controlled Operation of In-car Devices. *In Proceedings of the International Conference on Multimedia and Expo*, volume 3 de ICME '03, pages 541–544, Washington, États-Unis. IEEE Computer Society. (Cité en page 23.)
307. ZUPAN, L. H. et MERFELD, D. M. (2003). Neural processing of gravito-inertial cues in humans. IV. Influence of visual rotational cues during roll optokinetic stimuli. *Journal of Neurophysiology*, 89(1):390–400. (Cité en page 38.)

Annexes

Figures des résultats de l'expérience 1

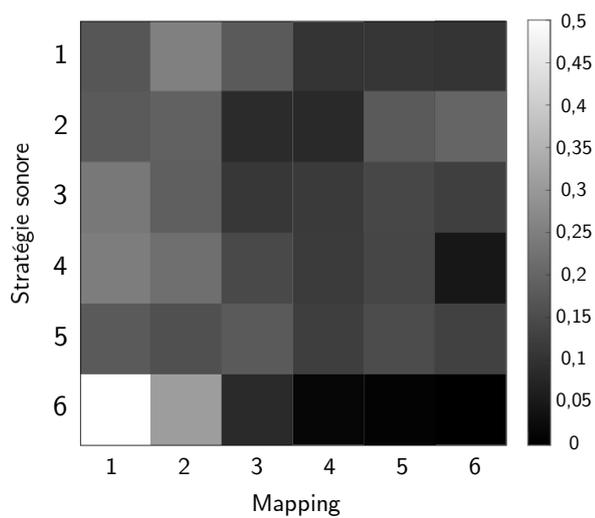


FIGURE A.1 – Représentation graphique des scores obtenus dans l'expérience 1. Les scores sont à lire uniquement en ligne, et les niveaux de gris varient entre 0 et 0,5 comme indiqué sur l'échelle de droite. Pour chaque stratégie, plus le score est élevé et plus le mapping semble naturel aux sujets.

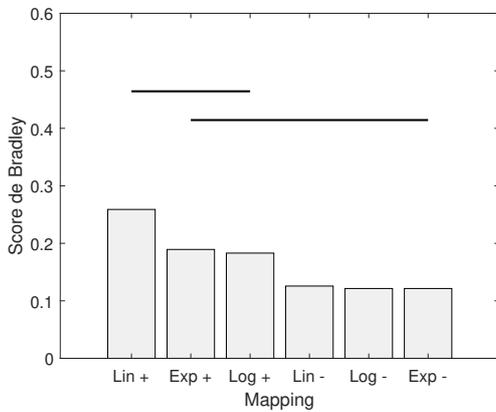


FIGURE A.2 – Scores obtenus pour la stratégie sonore 1, liant la distance à la hauteur. Les barres indiquent les mappings n'étant pas significativement différents.

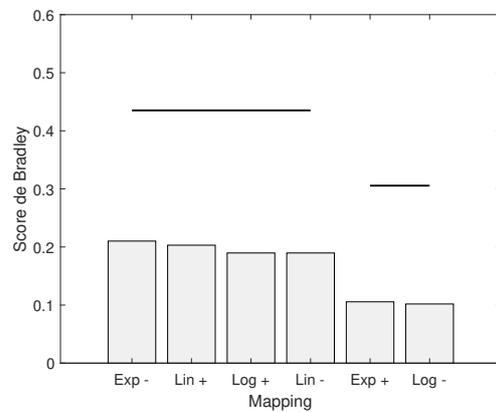


FIGURE A.3 – Scores obtenus pour la stratégie sonore 2, liant la distance à la modulation d'amplitude. Les barres indiquent les mappings n'étant pas significativement différents.

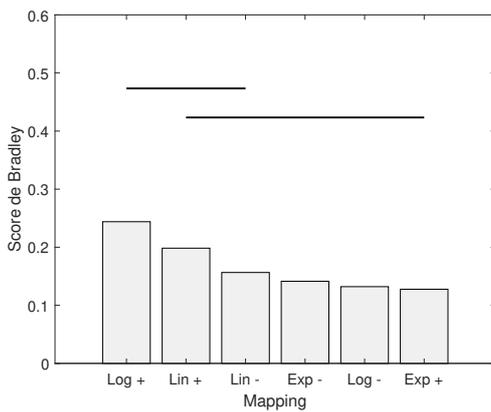


FIGURE A.4 – Scores obtenus pour la stratégie sonore 3, liant la distance à la brillance. Les barres indiquent les mappings n'étant pas significativement différents.

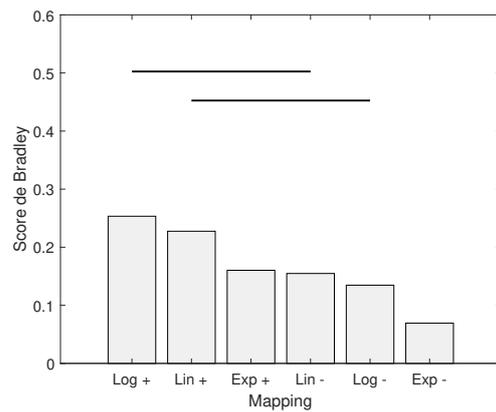


FIGURE A.5 – Scores obtenus pour la stratégie sonore 4, liant la vitesse à la hauteur. Les barres indiquent les mappings n'étant pas significativement différents.

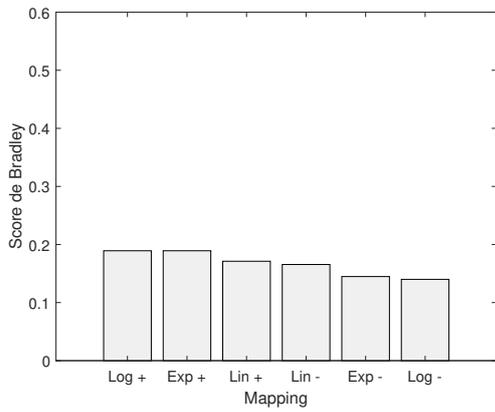


FIGURE A.6 – Scores obtenus pour la stratégie sonore 5, liant la vitesse à la modulation d’amplitude. Les barres indiquent les mappings n’étant pas significativement différents.

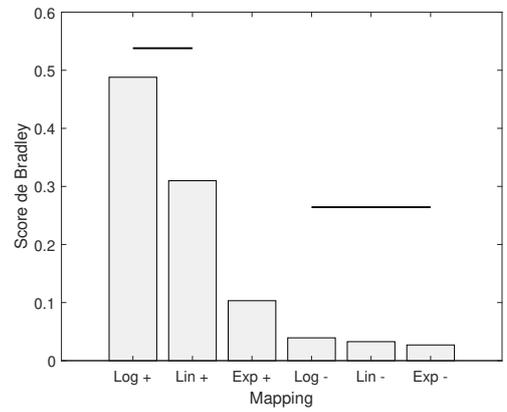


FIGURE A.7 – Scores obtenus pour la stratégie sonore 6, liant la vitesse à la brillance. Les barres indiquent les mappings n’étant pas significativement différents.

Figures relatives à l'analyse des sons du chapitre 5

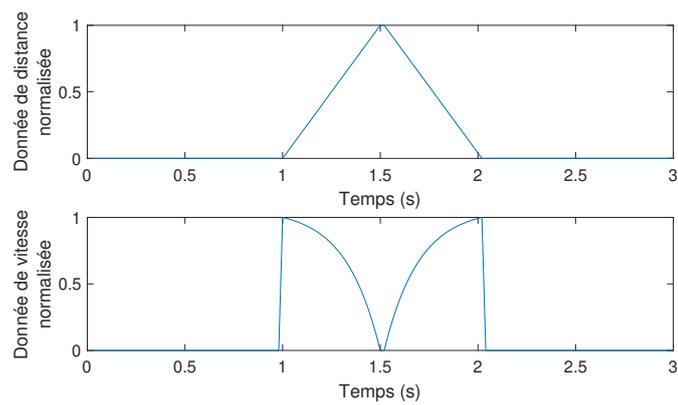


FIGURE B.1 – Données d'entrée pour la comparaison des sons, pour une stratégie sonore liée à la distance ou la vitesse.

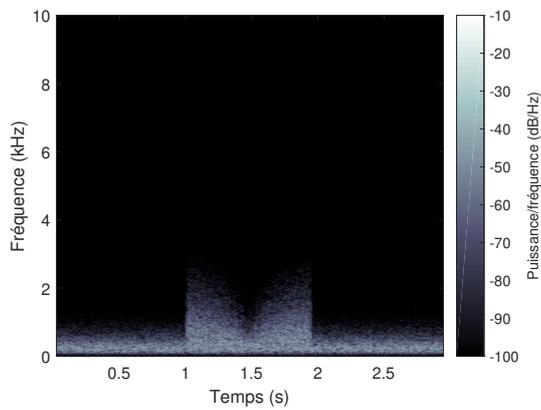


FIGURE B.2 – Spectrogram de la stratégie sonore alliant vitesse de l'objet virtuel au paramètre de brillance utilisée en expérience 2.

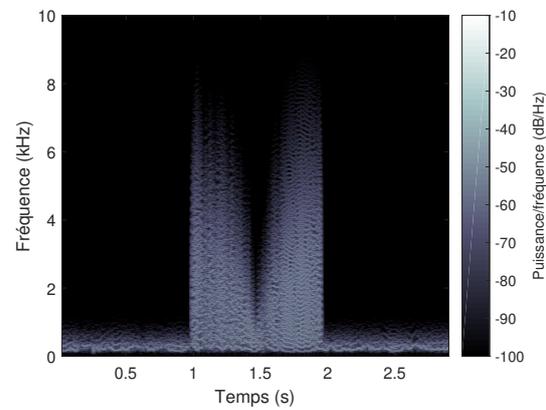


FIGURE B.3 – Spectrogram du son de roulement pur.

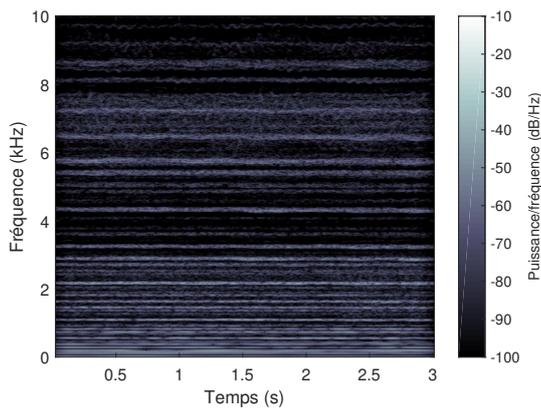


FIGURE B.4 – Spectrogram de la texture sonore choisie pour la stratégie sonore 6.

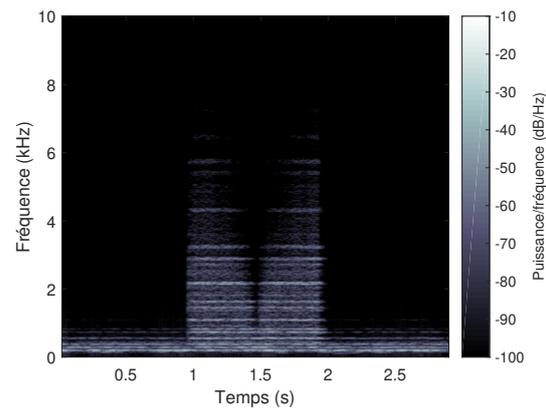


FIGURE B.5 – Spectrogram du son de roulement convolué avec la texture sonore.

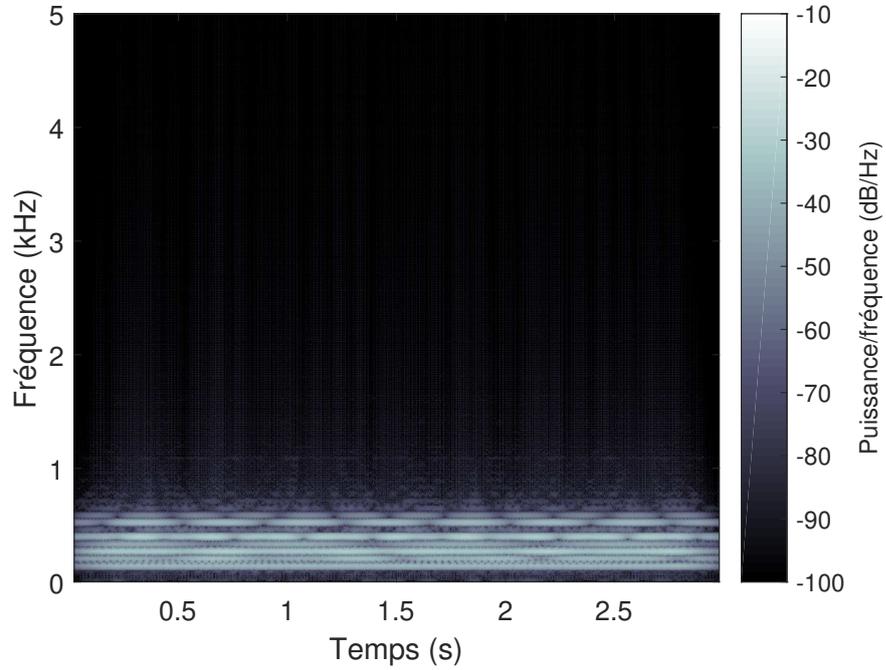


FIGURE B.6 – Spectrogramme de la texture sonore choisie pour la stratégie sonore 1.

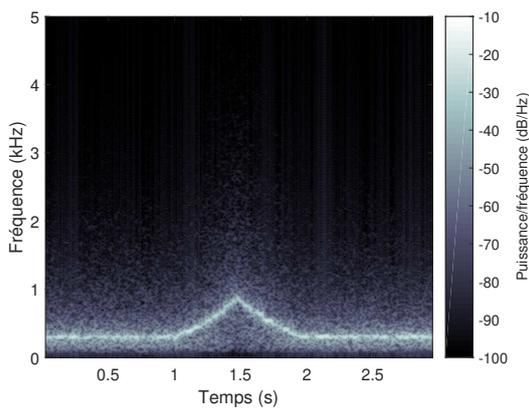


FIGURE B.7 – Spectrogramme de la stratégie sonore alliant distance main-OV au paramètre de hauteur utilisée en expérience 2.

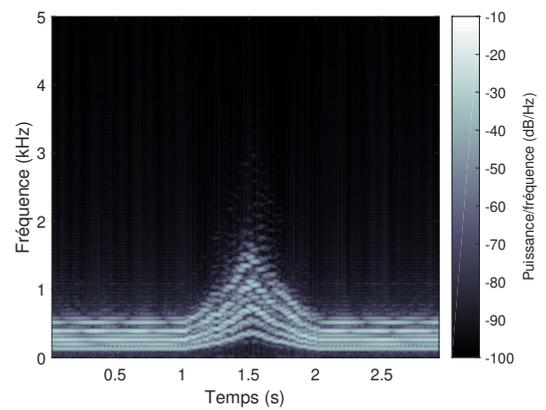


FIGURE B.8 – Spectrogramme de la texture sonore proposée avec un décalage fréquentiel.

Détail des étapes de la découverte de l'interface gestuelle de l'expérience 5

Étape 1 : Les sujets débutent la séance par une découverte du capteur de gestes et de la reconstruction de leur main. À cet instant, ils ne peuvent voir que la reconstruction de leur main comme le montre la figure C.1, et aucun retour sonore ne leur est donné. Ils sont invités à bouger leur main pour observer les conséquences sur la main reconstruite, et constater les limites du capteur.

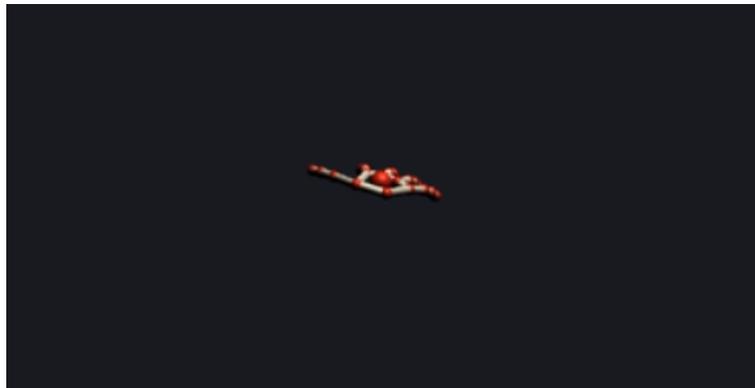


FIGURE C.1 – Vue de l'interface gestuelle lors de l'étape 1.

Étape 2 : L'objet virtuel est ajouté à la scène, et est accompagné de la stratégie sonore que nous avons mis au point. Il est alors possible de le manipuler librement, mais il n'est à ce stade pas possible de le saisir ni de changer sa taille. Les sujets peuvent alors constater la présence d'un ressort virtuel, qui attire l'objet virtuel vers la main.

Étape 3 : Le sujet est invité à faire des balayages de la main sur la gauche et sur la droite, et l'expérimentateur lui demande de faire tout particulièrement attention aux variations des sons lors des mouvements de l'objet virtuel.

Étape 4 : Il est maintenant possible de saisir l'objet virtuel en fermant la main avec la paume orientée vers le bas, voir figure C.2. Les sujets peuvent alors voir l'OV rester dans leur main jusqu'à ce qu'ils ouvrent de nouveau la main, et ils sont invités à enchaîner cette manipulation de l'OV avec des balayages de la main, pour constater les bonnes conditions d'enchaînement des manipulations. L'expérimentateur sensibilise les sujets au respect scrupuleux des gestes pour un bon fonctionnement de l'interface.

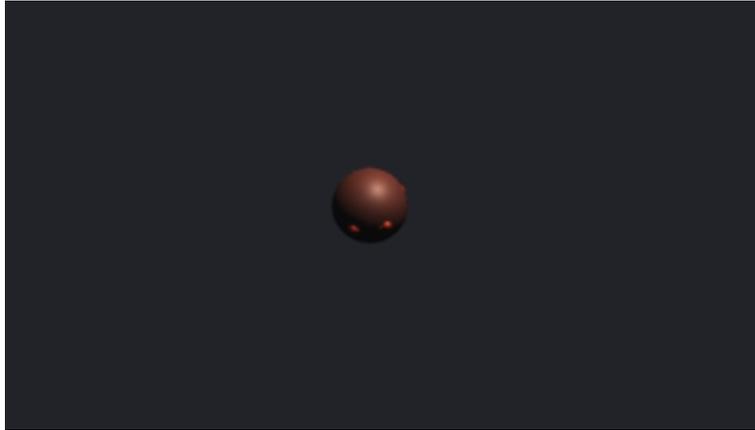


FIGURE C.2 – Vue de l'interface gestuelle lors de l'étape 4.

Étape 5 : L'interaction avec la taille de l'OV est ajoutée, les sujets peuvent donc moduler le rayon de l'objet virtuel en orientant la paume de leur main vers le haut, comme on peut le voir en figure C.3. L'objet vient alors se positionner dans leur main, et varie selon l'ouverture de la main jusqu'à ce que la main soit retournée. Les sujets sont également invités à enchaîner cette manipulation de l'OV avec des balayages de la main ou avec la saisie de l'objet virtuel.

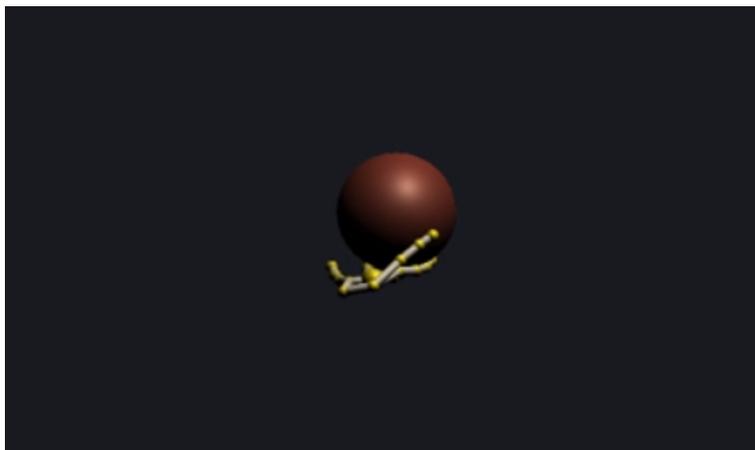


FIGURE C.3 – Vue de l'interface gestuelle lors de l'étape 5.

Étape 6 : À partir de cette étape, nous avons choisi de ne plus montrer la main reconstruite ni l'objet virtuel. Les sujets peuvent alors voir une première partie de l'interface avec l'affichage de la température, et ils sont invités à effectuer des balayages de la main pour en observer les changements. Des icônes sonores sont joués lorsqu'un changement de température est accompli, et l'expérimentateur souligne la différence de hauteur lors des envois de l'OV en zone 1 ou 2.

Étape 7 : Les sujets sont invités à saisir l'objet virtuel et à le déplacer selon l'axe vertical, pour constater les différences de réglages concernant l'orientation de la source. Des icônes sonores sont également perceptibles lorsque les limites du paramètre de positionnement de source sont atteintes en haut ou en bas. Les sujets sont également invités à faire des balayages de la main, pour enchaîner les réglages.

Étape 8 : Cette étape permet aux sujets de manipuler la taille de l'OV et d'observer les conséquences en terme d'intensité de ventilation, que ce soit par l'interface visuelle avec la jauge circulaire, ou avec le son de ventilation dont l'intensité est modulée par ce réglage. De la même manière que précédemment, l'expérimentateur invite les sujets à enchaîner les différentes manipulations pour enchaîner les gestes, mais aussi d'entendre les différents sons de l'interface.

Étape 9 : Les sujets sont sensibilisés au système de verrouillage automatique de l'interface. Ils sont invités à explorer la zone où les interactions sont prises en compte, et à observer la position de leur main lorsque l'interface se verrouille. Le bouton de déverrouillage placé sous le volant leur permet de déverrouiller l'interface.

Vue des 3 colorations possibles de la vision tête haute lors de la phase de conduite de l'expérience 5



FIGURE D.1 – Vue du simulateur pour une distance avec le lièvre inférieure à 40 mètres.



FIGURE D.2 – Vue du simulateur pour une distance avec le lièvre supérieure à 40 mètres mais inférieure à 60 mètres.



FIGURE D.3 – Vue du simulateur pour une distance avec le lièvre supérieure à 60 mètres.

ANNEXE E

Échelle RSME utilisée pour mesurer la concentration des conducteurs

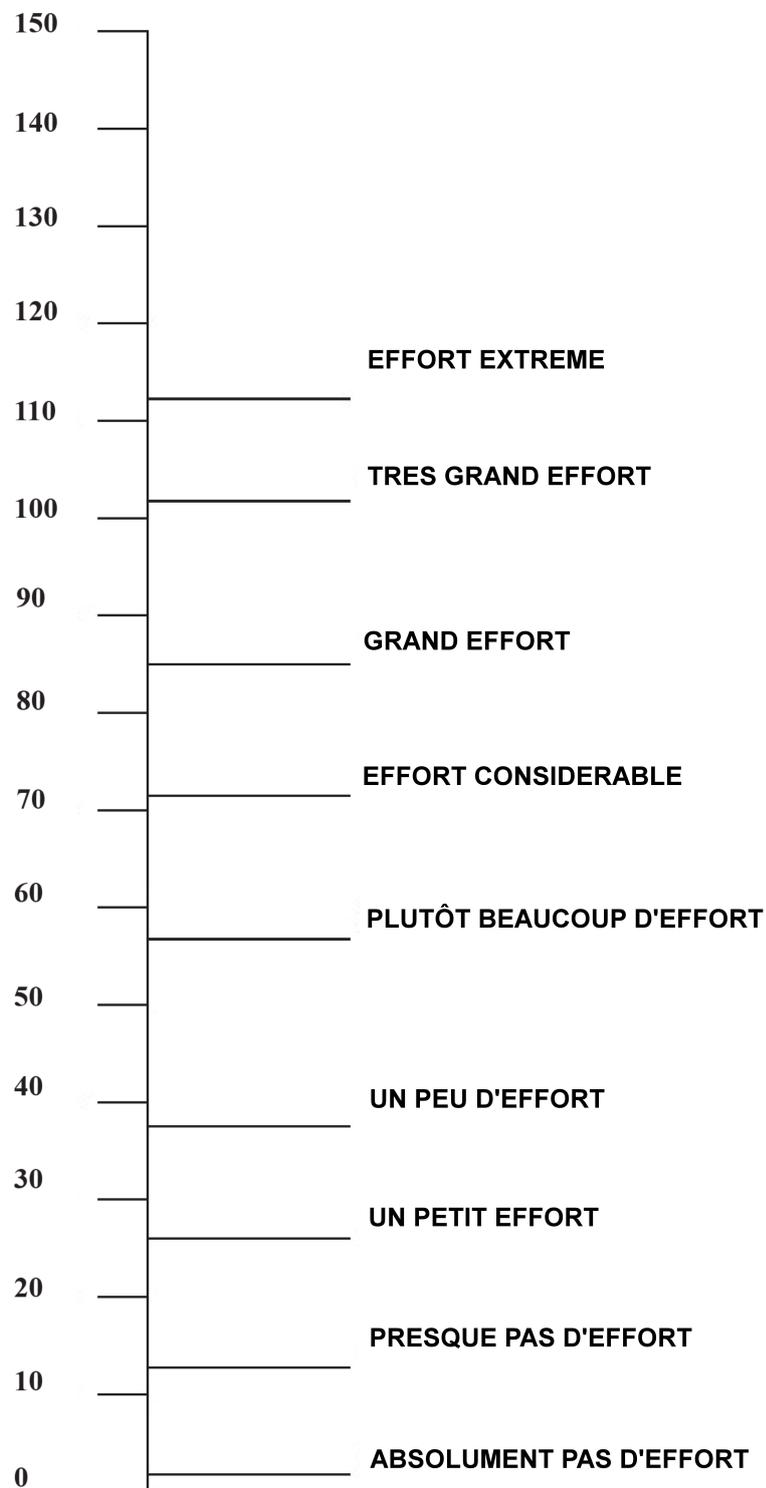
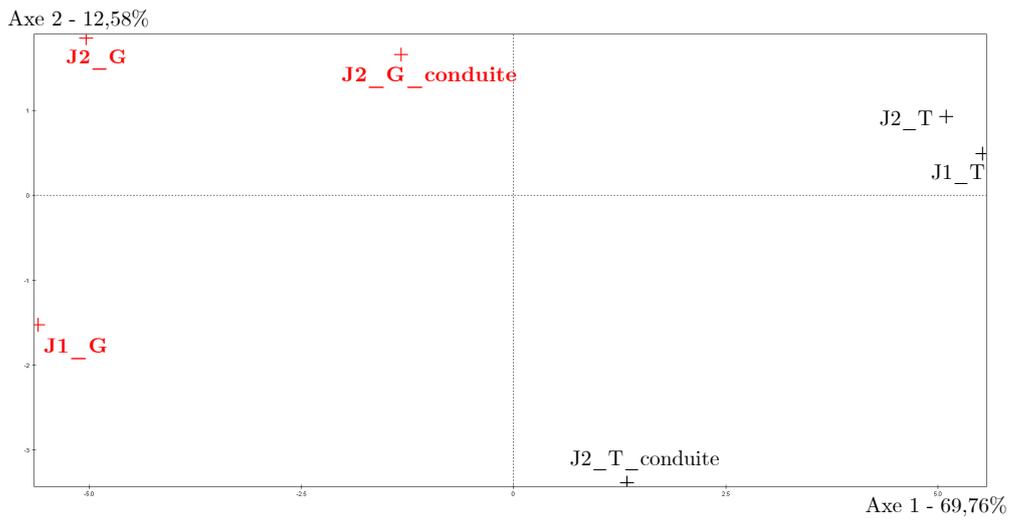


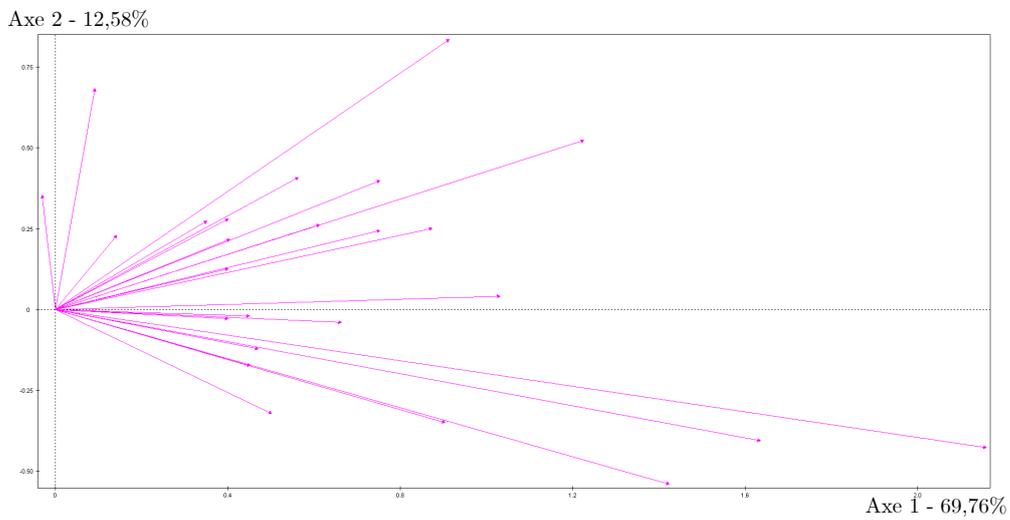
FIGURE E.1 – Échelle RSME traduite en français, utilisée pour l'expérience 5.

ANNEXE F

Détail des réponses au questionnaire de l'expérience 5 comparant les interfaces



(a) Évolution du score de confiance selon les conditions.



(b) Évolution du score de confiance par sujet.

FIGURE F.1 – ACP réalisée sur les scores de confiance.

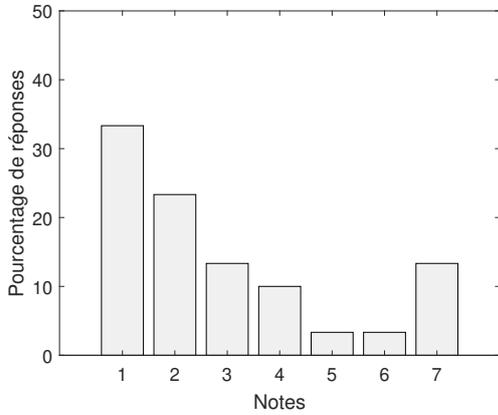


FIGURE F.2 – Répartition des réponses des sujets à :
Désignez l'interface paraissant la plus facile à utiliser à l'arrêt.
 1 : interface tactile ; 7 : interface gestuelle

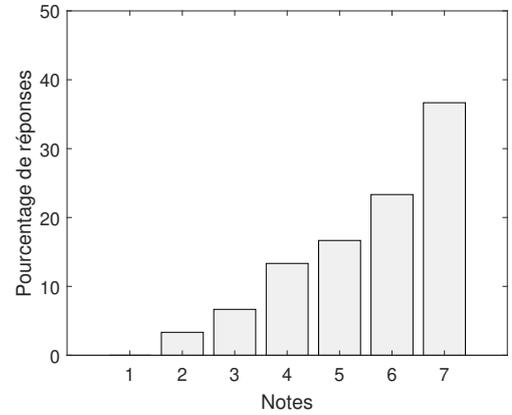


FIGURE F.3 – Répartition des réponses des sujets à :
Désignez l'interface paraissant la plus facile à utiliser en conduite.
 1 : interface tactile ; 7 : interface gestuelle

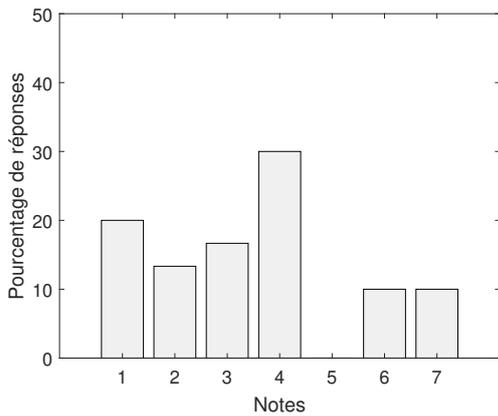


FIGURE F.4 – Répartition des réponses des sujets à :
Désignez l'interface paraissant la plus agréable à utiliser à l'arrêt.
 1 : interface tactile ; 7 : interface gestuelle

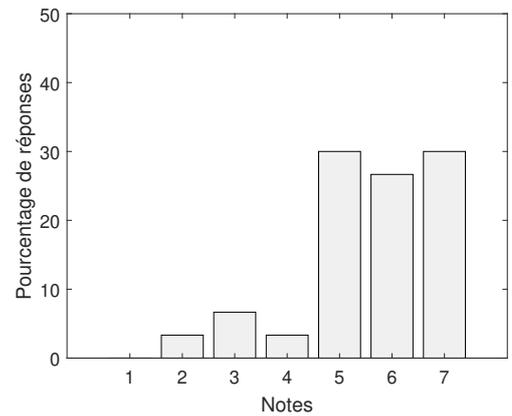


FIGURE F.5 – Répartition des réponses des sujets à :
Désignez l'interface paraissant la plus agréable à utiliser en conduite.
 1 : interface tactile ; 7 : interface gestuelle

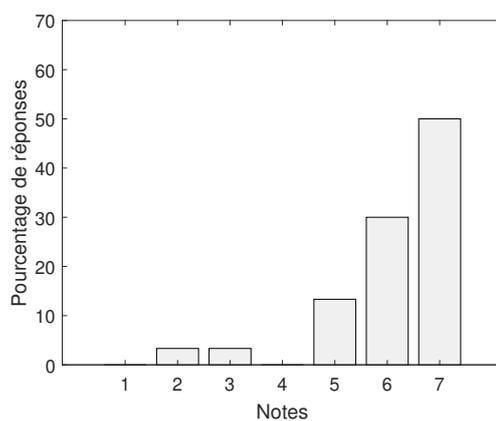


FIGURE F.6 – Répartition des réponses des sujets à : *Désignez l'interface paraissant la plus sécuritaire à utiliser en conduite. 1 : interface tactile ; 7 : interface gestuelle*

Détail des réponses au questionnaire de l'expérience 5 concernant l'interface gestuelle de gestion de la musique

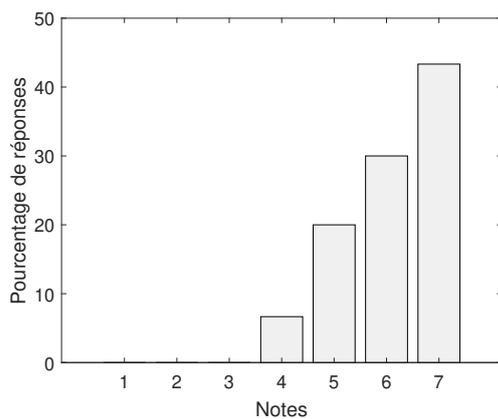


FIGURE G.1 – Répartition des réponses des sujets à la question :

Avez-vous trouvé pertinente l'utilisation des gestes sur la manipulation du changement de musique ?

1 : non pas du tout ; 7 : oui tout à fait

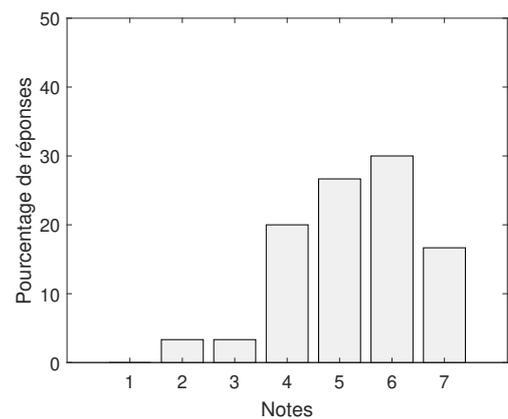


FIGURE G.2 – Répartition des réponses des sujets à la question :

Avez-vous trouvé pertinente l'utilisation des gestes sur l'avance ou le recul dans un morceau de musique ?

1 : non pas du tout ; 7 : oui tout à fait

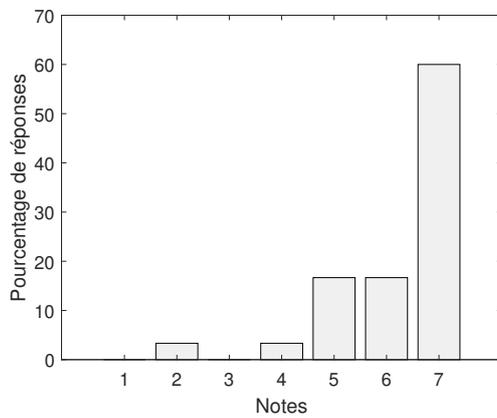


FIGURE G.3 – Répartition des réponses des sujets à la question :
Avez-vous trouvé pertinente l'utilisation des gestes sur la manipulation du positionnement de la source de musique ?
1 : non pas du tout ; 7 : oui tout à fait

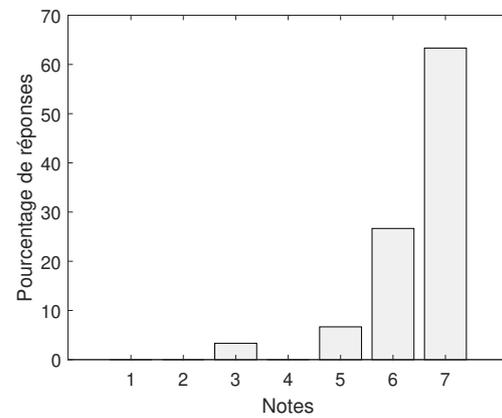


FIGURE G.4 – Répartition des réponses des sujets à la question :
Avez-vous trouvé pertinente l'utilisation des gestes sur la manipulation du volume sonore ?
1 : non pas du tout ; 7 : oui tout à fait

Détail des réponses au questionnaire général de l'expérience 5 concernant
l'interface gestuelle

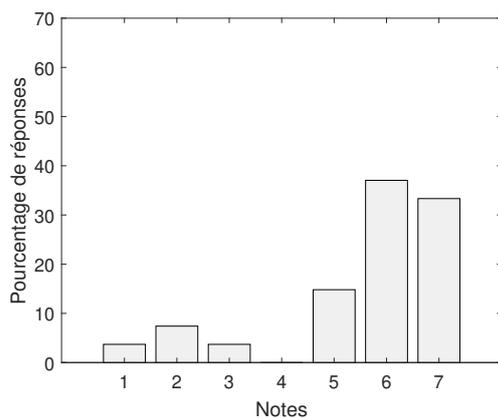


FIGURE H.1 – Répartition des réponses des sujets à la question :
Avez-vous réussi à imaginer le comportement de l'objet virtuel ?
1 : non pas du tout ; 7 : oui tout à fait

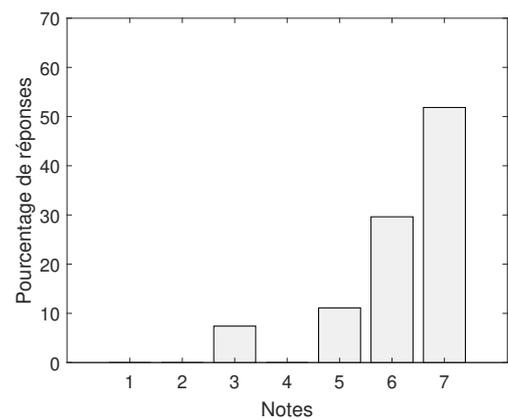


FIGURE H.2 – Répartition des réponses des sujets à la question :
Avez-vous trouvé que le nombre de gestes à retenir était convenable ?
1 : non pas du tout ; 7 : oui tout à fait

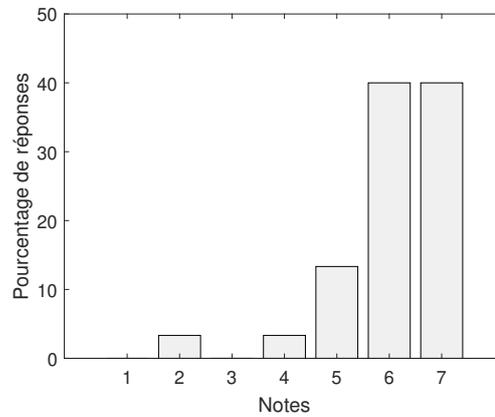


FIGURE H.3 – Répartition des réponses des sujets à la question : *Est-ce que la combinaison du tactile et des gestes vous paraît intéressante ? 1 : non pas du tout ; 7 : oui tout à fait*

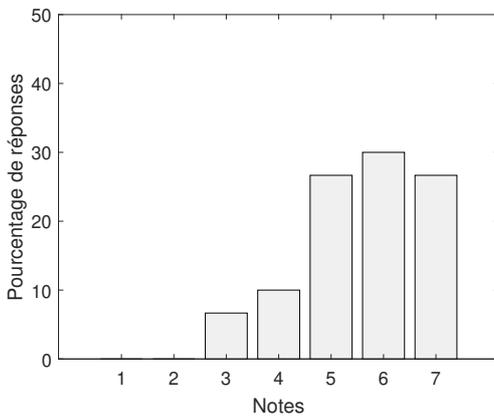


FIGURE H.4 – Répartition des réponses des sujets à la question :
Seriez-vous intéressé-e pour avoir ce genre de technologie dans votre véhicule pour la gestion de la ventilation ?
1 : non pas du tout ; 7 : oui tout à fait

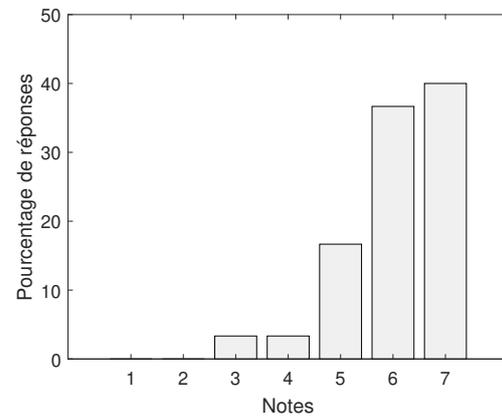


FIGURE H.5 – Répartition des réponses des sujets à la question :
Seriez-vous intéressé-e pour avoir ce genre de technologie dans votre véhicule pour la gestion des médias ?
1 : non pas du tout ; 7 : oui tout à fait

Titre : Manipulations gestuelles d'un objet virtuel sonifié pour le contrôle d'une interface en situation de conduite.

Résumé : Les constructeurs automobiles proposent un vaste éventail de fonctions secondaires à la conduite, en relation avec le GPS, la musique, ou encore la ventilation. L'ensemble de ces fonctions est centralisé sur un écran accessible au centre de l'habitacle, équipé d'une dalle tactile dans nombre de véhicules récents. Cependant, leur manipulation en conduite peut se révéler périlleuse : la sollicitation de la modalité visuelle pour interagir avec l'interface peut entraîner une baisse de vigilance vis-à-vis de la tâche de conduite, ce qui peut mener à des situations dangereuses. Au cours de cette thèse, fruit d'une collaboration entre le Groupe PSA et le laboratoire PRISM, nous nous sommes attachés à proposer une association de gestes et de sons comme alternative à une sollicitation visuelle. Le but est de proposer des interactions réalisables en aveugle, pour permettre au conducteur de focaliser ses yeux sur la route. Pour réaliser conjointement une manipulation de l'interface et une tâche de conduite, une sollicitation multisensorielle peut en effet permettre de faire baisser la charge mentale du conducteur, en comparaison à une situation unimodale visuelle. Pour que le lien entre les gestes et les sons paraisse naturel, un objet virtuel a été introduit, manipulable grâce à des gestes. Cet objet est le support des stratégies de sonification, par analogie aux sons de notre environnement, qui sont la conséquence d'une action sur un objet. L'objet virtuel permet également de structurer différents gestes autour d'une même métaphore, ou encore de redéfinir le menu d'une interface. La première partie de cette thèse est consacrée à la mise au point de stratégies de sonification pour transmettre des informations pertinentes sur la dynamique de l'objet virtuel. Deux expériences perceptives ont été mises en place, qui ont conduit à la discrimination de deux stratégies de sonification intéressantes. Dans une deuxième partie, nous avons œuvré à la mise en situation automobile par un travail sur les stimuli sonores, sur l'interface, et par l'étude de l'intégration multisensorielle. Un design des stratégies de sonification a été proposé pour permettre de se projeter dans une utilisation en véhicule. Par la suite, les évocations provoquées par le couplage des gestes et des sons ont été au centre d'une troisième expérience perceptive. Cette étude a été effectuée en aveugle, où le concept d'objet virtuel était inconnu et découvert progressivement par les sujets. Ces images mentales véhiculées par les stratégies de sonification peuvent en effet être utiles lors de la familiarisation des utilisateurs avec l'interface. Une quatrième expérience perceptive s'est concentrée sur la prise en main de l'objet virtuel, où l'intégration des stimuli visuels et auditifs a été étudiée, dans le contexte du maniement d'une interface. Les sujets ont été placés dans des conditions similaires à la découverte de l'interface en véhicule à l'arrêt, avec des stimuli audio-visuels ; puis à son utilisation en aveugle grâce aux stratégies de sonification. Les enseignements de ces expériences ont permis de bâtir une interface gestuelle, qui a été comparée à une interface tactile dans une dernière expérience perceptive réalisée en simulateur de conduite. Bien que les résultats montrent une utilisation plus performante de l'interface tactile, l'association des gestes et des sons semble intéressante du point de vue de la charge cognitive des conducteurs. L'interface gestuelle peut donc offrir une alternative prometteuse ou un complément aux interfaces tactiles pour une utilisation simultanée à la conduite en toute sécurité.

Mots-clefs : Relation gestes/sons, perception sonore, perception multisensorielle, sonification, charge cognitive, interface homme-machine.

Title: Gestural manipulations of a sonified virtual object for interface control in driving situations.

Abstract: Car manufacturers offer a wide range of secondary driving controls, such as GPS, music, or ventilation, often localized on a central touch-sensitive screen. However, operating them while driving proves to be unsafe: engaging the sense of sight for interface interaction can lead to vigilance reduction towards the driving task, which can lead to high-risk situations. In this PhD thesis, which is a part of a collaborative research project involving both the PSA Group and the PRISM laboratory, we aim to provide a gesture and sound association as an alternative to the visual solicitation. The goal is to enable blind interface interactions, allowing the driver to focus their eyes on the road. When jointly performing interface manipulations and the driving task, a multisensory solicitation can lower the driver's cognitive load, in comparison with a visual unimodal situation. For the gesture-sound association to feel more natural, a virtual object that can be handled with gestures is introduced. This object is the support for sonification strategies, constructed by analogy with sounds from our environment, which are the consequence of an action on an object. The virtual object also allows to structure different gestures around the same metaphor, or to redefine the interface's menu. The first part of this thesis deals with the development of sonification strategies, with the aim to inform users about the virtual object dynamic. Two perceptual experiments were set up, which led to the discrimination of two valuable sonification strategies. In a second part, the automotive application was addressed by designing new sound stimuli, the interface, and by studying the multisensory integration. Sounds were proposed for each of the two sonification strategies, to progress towards an in-vehicle integration. The evocations brought by the gestures and sounds association were the subject of a third perceptive blinded experiment. The concepts around the virtual object were unknown and gradually discovered by the subjects. These mental images conveyed by the sonification strategies can help users familiarize with the interface. A fourth perceptual experiment focused on the virtual object handling for first-time users, where the integration of audio-visual stimuli was studied, in the context of an interface manipulation. The experiment conditions were similar to the driver first discovering of the interface in a parked vehicle thanks to audio-visual stimuli, and then operating it through sonification strategies only. The results of this experiment lead to the design of a gestural interface, which was compared with the results obtained with a touchscreen interface in a final perceptual experiment, carried out in a driving simulator. Although the results show better performances for the tactile interface, the combination of gestures and sounds proved to be effective from the cognitive load point of view. The gesture interface can therefore offer a promising alternative or complement to tactile interfaces for a safe simultaneous use in driving condition.

Keywords: Gestures/sound relation, sound perception, multisensory perception, sonification, cognitive load, human-machine interface.

