



**HAL**  
open science

# De l'étude du confort sonore de l'habitacle du véhicule électrique à son amélioration par l'acoustique augmentée

Matthieu Duroyon

## ► To cite this version:

Matthieu Duroyon. De l'étude du confort sonore de l'habitacle du véhicule électrique à son amélioration par l'acoustique augmentée. Acoustique [physics.class-ph]. Université de Technologie de Compiègne, 2025. Français. ⟨NNT : 2025COMP2868⟩. ⟨tel-05091019⟩

**HAL Id: tel-05091019**

**<https://theses.hal.science/tel-05091019v1>**

Submitted on 31 May 2025

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.

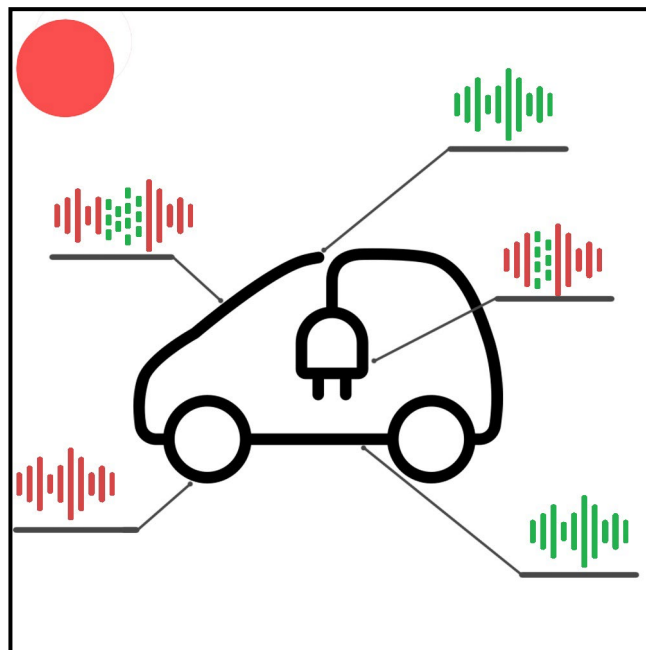


HAL Authorization

Par **Matthieu DUROYON**

*De l'étude du confort sonore de l'habitacle  
du véhicule électrique à son amélioration  
par l'acoustique augmentée*

Thèse présentée  
pour l'obtention du grade  
de Docteur de l'UTC



Soutenue le 14 mars 2025

**Spécialité** : Acoustique et Vibrations : Unité de recherche en  
Mécanique - Laboratoire Roberval

D2868

---

De l'étude du confort sonore de  
l'habitacle du véhicule électrique à son  
amélioration par l'acoustique  
augmentée

---

**Matthieu Duroyon**

Thèse de Doctorat en Acoustique et Vibrations  
Université de Technologie de Compiègne, France  
14 mars 2025

# De l'étude du confort sonore de l'habitacle du véhicule électrique à son amélioration par l'acoustique augmentée

Thèse soumise pour l'obtention du grade de Docteur  
de l'Université de Technologie de Compiègne

**Matthieu Duroyon**

**Spécialité : Acoustique et Vibrations**

*Laboratoire Roberval  
Université de Technologie de Compiègne*

Soutenue le 14 mars 2025 devant le jury composé de :

## **Directeurs de Thèse**

Nicolas DAUCHEZ	Professeur des universités	Roberval, UTC, Compiègne
Nicolas MISDARIIS	Directeur de recherche	STMS Lab, IRCAM, Paris

## **Encadrants**

Patrick SUSINI	Directeur de recherche	STMS Lab, IRCAM, Paris
Louis-Ferdinand PARDO	Expert	Renault Group, Guyancourt

## **Rapporteurs**

Jean-François PETIOT	Professeur des universités	LS2N, Ecole Centrale de Nantes, Nantes
Clara SUIED	Chargée de recherche	IRBA , Brétigny-sur-Orge

## **Examineurs**

Roland CAHEN	Enseignant-chercheur	ENSCI Les Ateliers, Paris
Arthur GIVOIS	Maître de conférences	Roberval, UTC, Compiègne
Sabine MEUNIER	Chargée de recherche	LMA, CNRS, Marseille

*À Cécile Guillot-Pattyn.*

## Remerciements

L'accomplissement d'un travail tel qu'une thèse de doctorat ne saurait être entièrement attribué à une seule personne, quand bien même elle en serait l'auteur. Je profite donc de ces quelques lignes pour remercier les personnes ayant permis l'aboutissement de ce travail.

En premier lieu, j'aimerais remercier les encadrants à l'origine de cette thèse : Louis-Ferdinand Pardo, Nicolas Dauchez, Nicolas Misdariis et Patrick Susini, pour m'avoir accordé leur confiance tout au long de ce travail. La liberté dont j'ai bénéficié dans ma manière de travailler, ainsi qu'un suivi régulier, m'ont permis d'acquérir une maturité et une connaissance du sujet qu'il aurait été impossible d'obtenir autrement.

J'aimerais ensuite remercier Jean-François Petiot et Clara Suied, qui ont accepté d'être rapporteurs de ce manuscrit, ainsi que Roland Cahen, Arthur Givois et Sabine Meunier, membres du jury de ma soutenance. La finesse de la relecture et la précision des questions le jour de la soutenance ont permis de mettre en lumière ce travail.

Je tiens également à exprimer ma gratitude envers l'ensemble de mes collègues que j'ai pu côtoyer au cours de ma thèse. D'abord à l'UTC, où de précieuses pauses café et événements sociaux ont rythmé ces années, grâce aux équipes de doctorant en acoustique et ceux en mécanique des matériaux, composés de Grégoire, Simon, Alexandre, Camille, Christophe, Apolline, Ouardia, Marcià, Guilhem et bien d'autres, qui ont su rendre ces moments agréables, comiques ou mémorables.

Ensuite, à l'IRCAM, où Michèle, Léo, Armand et Clara ont su créer une ambiance unique, au travers de nouvelles décorations à chaque saison, d'un calendrier de l'Avent sonore, de mots croisés collaboratifs et bien d'autres initiatives. Je remercie particulièrement Éléonore Vialatte, qui a passé six mois en stage à l'IRCAM sur une partie de mon sujet de thèse et qui a apporté une contribution précieuse, scientifique, technique et surtout humaine.

Je remercie chaleureusement les acousticiens de chez Renault ayant accepté de contribuer à l'avancement de cette thèse, en particulier Christophe Thévenard et Thomas Antoine, avec qui j'ai pu échanger longuement sur des sujets aussi variés que la musique minimaliste, l'importance du son dans l'automobile, la synthèse sonore de phénomènes physiques, l'histoire de l'acoustique chez Renault et, bien sûr, le confort sonore dans le véhicule électrique.

Je tiens aussi à remercier l'ensemble de mes amis qui m'ont accompagné durant ces trois années de thèse, mais aussi dans ma vie personnelle parfois mouvementée : Justin, Jordan, Marc, Bakari, Manuel, Pierre-Sylvain, Damien et tant d'autres que je ne mentionne pas mais n'oublie pas non plus, sans qui ces années auraient été bien monotones.

Enfin, je remercie ma famille, qui a été présente tout au long de cette thèse. Mon père qui s'est toujours intéressé au sujet de mes études, ma grande sœur qui avait déjà ouvert la voie du doctorat dans la famille, ma petite sœur pour son aide précieuse dans l'organisation d'événements festifs dignes de ce nom, et mon frère pour les vacances et week-ends passés ensemble – pas toujours de tout repos, mais assurément dépayants.

Un immense merci à ma mère, qui parmi tant d'autres choses, a accepté de relire en détail une grande partie de ce manuscrit, et à ma grand-mère, avec qui j'ai partagé de nombreux repas et qui n'aura jamais autant visité le musée Georges Pompidou que pendant ces trois années.

## Abstract

Recent advances in cabin soundproofing in automobiles, along with the generalization of electric motors, silent, have profoundly transformed the user's auditory experience and challenged established notions of comfort, particularly sonic comfort. This thesis aims to explore the concept of sonic comfort in this context. To achieve this, a review of the scientific literature is conducted to clarify and define this notion of sonic comfort. This also enables to identify the tools suited to analyze and detect the emergence of sound sources that were previously masked or nonexistent, while considering the unique role of sound in the automotive environment.

This thesis is structured into three distinct and complementary parts. The first part addresses the description of the sound environment through a series of interviews with sound automotive specialists, focusing on sound descriptors. This work results in the development of a lexicon based on a paradigm designed within the PDS (Perception et Design Sonores) team, known as the SpeaK paradigm [1, 2]. The second part of the thesis focuses on improving the sonic comfort of a specific situation characterized by a sound described as unpleasant or bothersome. The proposed solution involves adding sounds designed according to different masking typologies, referred to as Energetic, Habillage, and Attentional masking. Among these typologies, masks are created using a parametric synthesizer developed specifically for this purpose. A perceptual study in a controlled environment, employing a Best-Worst Scaling methodology, is then conducted. This study, carried out with 50 participants, quantifies the differences between these methodologies across two evaluated dimensions : defect detectability and scene pleasantness [3]. Following this study, a perspective for in situ experiments to validate the results, as well as insights into the limitations and possible improvements of the experimental method, are proposed.

## Résumé

Les progrès récents de l'insonorisation des habitacles dans l'automobile ainsi que l'avènement de la motorisation électrique, silencieuse, ont durablement transformé l'expérience sonore de l'utilisateur et remettent en question les acquis du confort, et notamment du confort sonore. Cette thèse a pour objectif d'explorer le confort sonore dans cet environnement. Pour cela un travail de littérature scientifique permet de clarifier et définir cette notion de confort sonore, et de mettre en évidence les outils adaptés pour traiter et identifier l'émergence de sources sonores, auparavant masquées ou inexistantes tout en prenant en compte la place particulière du son dans un milieu comme l'automobile.

Cette thèse se structure en trois parties distinctes et complémentaires. La première aborde le point de la description de l'environnement sonore à travers un travail d'interviews avec des spécialistes automobiles du son, sur les descripteurs du son. Ce travail aboutit à la construction d'un lexique en suivant un paradigme élaboré au sein de l'équipe PDS (Perception et Design Sonores), le paradigme SpeaK [1, 2]. La deuxième partie de la thèse s'attelle à améliorer le confort sonore d'une situation sonore précise, comportant un son décrit comme désagréable ou gênant. La solution explorée consiste à ajouter des sons conçus suivant différentes typologies de masquage, appelées masquage énergétique, masquage par habillage et masquage attentionnel. Parmi ces différentes typologies, des masques sont alors créés grâce à un synthétiseur paramétrique développé spécifiquement pour cette tâche. Une étude perceptive en atmosphère contrôlée, utilisant une méthodologie *Best-Worst Scaling* est alors menée. Cette étude, menée avec 50 participants, permet de quantifier les différences entre ces typologies sur deux dimensions évaluées : la détectabilité du défaut et l'agrément de l'environnement sonore [3]. À l'issue de cette étude, une perspective d'expérience in-situ pour confronter les résultats, ainsi que la mise en lumière de limites et d'améliorations possibles de la méthode expérimentale sont proposées.

## Abréviations

<b>ACP</b>	Analyse en Composantes Principales
<b>ANFA</b>	Association Nationale pour la Formation Automobile
<b>ASD</b>	<i>Active Sound Design</i>
<b>AVAS</b>	<i>Acoustic Vehicle Alerting System</i>
<b>AVI</b>	Acoustique et vibrations pour l'ingénieur
<b>BF</b>	Basse-Fréquences
<b>BWS</b>	<i>Best-Worst Scaling</i>
<b>CD</b>	<i>Compact Disc</i>
<b>CIFRE</b>	Convention Industrielle de Formation par la Recherche
<b>DDL</b>	degrés de libertés
<b>DL</b>	<i>Difference Limen</i>
<b>DS</b>	Différentiel Sémantique
<b>ESAD</b>	Ecole Supérieure d'Art et de Design
<b>ENSCI</b>	Ecole Nationale Supérieure de Création Industrielle
<b>ESE</b>	<i>Engine Sound Enhancement</i>
<b>HF</b>	Haute-Fréquences
<b>IHM</b>	Interface Homme-Machine
<b>INSEAD</b>	Institut européen d'administration des affaires
<b>IRBA</b>	Institut de Recherche Biomédicale des Armées
<b>IRNS</b>	<i>Iterated Rippled Noise Sample</i>
<b>IRCAM</b>	Institut de Recherche et Coordination Acoustique/Musique
<b>JATOS</b>	<i>Just Another Tool for Online Studies</i>
<b>jnd</b>	<i>just-noticeable difference</i>
<b>LMA</b>	Laboratoire de Mécanique de d'Acoustique
<b>LMS</b>	<i>Leuven Measurement System</i>
<b>LS2N</b>	Laboratoire des Sciences du Numérique de Nantes
<b>MDS</b>	<i>Multi-Dimensional Scaling</i>
<b>MF</b>	Moyenne-Fréquences
<b>MFI</b>	<i>Multidimensional Fatigue Inventory</i>
<b>NLP</b>	<i>Natural Language Processing</i>
<b>PAT</b>	<i>Psychoacoustic Tonality</i>
<b>PDS</b>	Perception et Design Sonores
<b>POMS</b>	<i>Profil of Mood States</i>
<b>PR</b>	<i>Prominence Ratio</i>
<b>RMS</b>	<i>Root Mean Square</i>
<b>RNC</b>	<i>Road Noise Cancellation</i>
<b>STMS</b>	Sciences et Technologies de la Musique et du Son
<b>TNR</b>	<i>Tone to Noise Ratio</i>
<b>UTC</b>	Université de technologie de Compiègne

# Table des matières

<b>Abstract</b>	<b>iii</b>
<b>Résumé</b>	<b>iv</b>
<b>Abréviations</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>Préface</b>	<b>1</b>
<b>1 Etat de l'art</b>	<b>3</b>
1.1 Introduction de chapitre . . . . .	3
1.2 Éléments historique . . . . .	3
1.2.1 La place du son dans l'automobile et chez Renault . . . . .	3
1.2.2 Histoire du véhicule électrique . . . . .	6
1.2.3 Le problème du silence . . . . .	7
1.2.4 Résumé de la section . . . . .	9
1.3 Du confort au design sonore . . . . .	9
1.3.1 Définition du confort . . . . .	9
1.3.1.1 Ce que n'est pas le confort . . . . .	9
1.3.1.2 Les modèles de confort . . . . .	10
1.3.1.3 Ce qu'est le confort . . . . .	11
1.3.2 Du confort acoustique au confort sonore. . . . .	11
1.3.2.1 Le confort acoustique . . . . .	11
1.3.2.2 Le confort sonore . . . . .	12
1.3.2.3 Le confort sonore dans la voiture . . . . .	12
1.3.3 Ajouter du son . . . . .	13
1.3.3.1 Sons intentionnels . . . . .	13
1.3.3.2 Impact de la musique . . . . .	14
1.3.3.3 La fatigue . . . . .	15
1.3.3.4 Acoustique augmentée . . . . .	15
1.3.4 Design sonore . . . . .	16
1.3.5 Méthode scientifique appliquée et design sonore . . . . .	17
1.3.6 Résumé de la section . . . . .	18
1.4 Des paysages sonores à la description verbale . . . . .	19
1.4.1 Paysages sonore : définition . . . . .	19
1.4.2 L'habitacle du véhicule comme paysage sonore . . . . .	19
1.4.3 Caractériser un paysage sonore . . . . .	19
1.4.4 Historique de la description du son . . . . .	20
1.4.5 Différents types d'écoute . . . . .	21
1.4.6 Type de discours . . . . .	22
1.4.7 Construction de lexiques et paradigme SpeaK . . . . .	23
1.4.8 Résumé de la section . . . . .	26
1.5 Objectif . . . . .	27

<b>2</b>	<b>Description d'un nouvel environnement sonore</b>	<b>29</b>
2.1	Introduction de chapitre . . . . .	29
2.2	Collecte des termes . . . . .	30
2.2.1	Rédaction du questionnaire semi-directif . . . . .	30
2.2.2	Modalité orale : interviews . . . . .	31
2.2.3	Réalisation des interviews . . . . .	32
2.2.4	Résumé de la section . . . . .	33
2.3	Construction du lexique . . . . .	33
2.3.1	Transcription et étiquetage . . . . .	33
2.3.1.1	Transcription . . . . .	33
2.3.1.2	La classe « Nom de son » . . . . .	33
2.3.1.3	Étiquetage . . . . .	35
2.3.2	Sélection d'une liste de termes . . . . .	35
2.3.2.1	Méthode statistique . . . . .	36
2.3.2.2	Méthode sémantique . . . . .	37
2.3.2.3	Discussion des résultats obtenus . . . . .	39
2.3.2.4	Le descripteur « <i>Harmonique</i> » . . . . .	41
2.3.2.5	Liste de termes . . . . .	42
2.3.3	Construction des définitions . . . . .	42
2.3.3.1	Objectif des définitions . . . . .	42
2.3.3.2	Groupe de travail . . . . .	42
2.3.3.3	Liste définitive des termes . . . . .	44
2.3.3.4	Construction des définitions . . . . .	44
2.3.3.5	Exemple avec le hululement . . . . .	45
2.3.3.6	Résultats . . . . .	46
2.3.4	Illustration sonore des termes . . . . .	47
2.3.4.1	Extraits contextualisés . . . . .	47
2.3.4.2	Extraits décontextualisés . . . . .	47
2.3.4.3	Extraits synthétisés . . . . .	48
2.3.4.4	Extraits semblables . . . . .	48
2.3.5	Résumé de la section . . . . .	50
2.4	Validation et finalisation . . . . .	50
2.4.1	Validation . . . . .	50
2.4.1.1	Validation quantitative des experts interviewés sur le lexique . . . . .	50
2.4.1.2	Retour qualitatif des experts interviewés sur le lexique . . . . .	52
2.4.2	Lexique final . . . . .	53
2.4.2.1	Roulement . . . . .	53
2.4.2.2	Sirènement . . . . .	53
2.4.2.3	Sifflement aéroacoustique . . . . .	53
2.4.2.4	Sifflement électrique . . . . .	54
2.4.2.5	Bourdonnement . . . . .	54
2.4.2.6	Ventilation . . . . .	54
2.4.2.7	Chargement . . . . .	54
2.4.2.8	Hululement . . . . .	54
2.4.2.9	Fuite . . . . .	55
2.4.2.10	Hachage . . . . .	55
2.4.2.11	Tambourinement . . . . .	55
2.4.3	Résumé de la section . . . . .	55

2.5	Conclusion de chapitre . . . . .	56
<b>3</b>	<b>Création de sons ajoutés pour l’acoustique augmentée</b>	<b>59</b>
3.1	Introduction de chapitre . . . . .	59
3.2	Choisir un environnement sonore . . . . .	60
3.2.1	Interview : partie sur le confort . . . . .	60
3.2.1.1	Les limites du lexique construit . . . . .	60
3.2.1.2	Le lexique comme outil pour s’approprier le contexte . . . . .	60
3.2.2	Synthèse des éléments contextuels . . . . .	61
3.2.3	Identification de situations sonores propices à l’acoustique augmentée . . . . .	62
3.2.3.1	Situation sonore 1 . . . . .	62
3.2.3.2	Situation sonore 2 . . . . .	64
3.2.3.3	Situation sonore 3 . . . . .	64
3.2.4	Résumé de la section . . . . .	64
3.3	Typologies de masquage . . . . .	65
3.3.1	Fonction du masque . . . . .	65
3.3.1.1	Masque Énergétique . . . . .	65
3.3.1.2	Masque Habillage . . . . .	65
3.3.1.3	Masque Attentionnel . . . . .	65
3.3.2	Aspect du masque . . . . .	66
3.3.2.1	Masque Énergétique . . . . .	66
3.3.2.2	Masque Habillage . . . . .	67
3.3.2.3	Masque Attentionnel . . . . .	68
3.3.3	Résumé de la section . . . . .	68
3.4	Construction d’un synthétiseur . . . . .	69
3.4.1	Situation sonore modulable . . . . .	69
3.4.2	Matière première . . . . .	70
3.4.2.1	Synthèse granulaire . . . . .	72
3.4.2.2	Lecture en boucle . . . . .	73
3.4.3	Octaver . . . . .	75
3.4.4	Filtre . . . . .	76
3.4.5	Aspect dynamique . . . . .	77
3.4.6	Résumé de la section . . . . .	79
3.5	Composition des masques . . . . .	79
3.5.1	Masque Énergétique 1 (E1) . . . . .	79
3.5.2	Masque Énergétique 2 (E2) . . . . .	80
3.5.3	Masque Habillage 1 (H1) . . . . .	81
3.5.4	Masque Habillage 2 (H2) . . . . .	82
3.5.5	Masque Attentionnel 1 (A1) . . . . .	83
3.5.6	Masque Attentionnel 2 (A2) . . . . .	83
3.5.7	Résumé de la section . . . . .	84
3.6	Conclusion de chapitre . . . . .	85
<b>4</b>	<b>Évaluation des solutions conçues avec la méthodologie <i>Best-Worst Scaling</i></b>	<b>87</b>
4.1	Introduction de chapitre . . . . .	87
4.2	Méthodes d’évaluation perceptive . . . . .	88
4.2.1	Les évaluations indirectes . . . . .	88
4.2.2	Les évaluations directes . . . . .	88

4.2.3	La méthode de comparaisons par paires . . . . .	89
4.3	Le choix de la méthodologie <i>Best-Worst Scaling</i> . . . . .	89
4.3.1	Le principe . . . . .	89
4.3.2	Les algorithmes de scoring . . . . .	89
4.3.3	La conception d'une expérience BWS . . . . .	91
4.3.4	La question du nombre de participants . . . . .	91
4.3.5	Résumé de la section . . . . .	93
4.4	Méthode . . . . .	93
4.4.1	Schéma expérimental . . . . .	93
4.4.2	Construction des stimuli . . . . .	94
4.4.2.1	Facteur Masque . . . . .	94
4.4.2.2	Facteur Défaut . . . . .	94
4.4.2.3	Facteur Volume . . . . .	94
4.4.2.4	Facteur Brillance . . . . .	96
4.4.3	Participants . . . . .	98
4.4.4	Réalisation technique . . . . .	98
4.4.5	Déroulement de l'expérience . . . . .	99
4.4.6	Résumé de la section . . . . .	100
4.5	Résultats . . . . .	101
4.5.1	Sons fictifs . . . . .	101
4.5.2	Comparaison inter-expérience . . . . .	101
4.5.3	Cohérence et conformité des participants . . . . .	102
4.5.3.1	Cohérence inter-participant . . . . .	102
4.5.3.2	Cohérence intra-participant . . . . .	103
4.5.3.3	Une comparaison intra-participant avec le score individuel ? . . . . .	104
4.5.4	Résultats sur l'expérience de détectabilité : pertinence des typologies . . . . .	105
4.5.5	Résultats sur l'expérience de détectabilité : analyse des facteurs . . . . .	106
4.5.6	Résultats sur l'expérience d'agrément : pertinence des typologies . . . . .	107
4.5.7	Résultats sur l'expérience d'agrément : analyse des facteurs . . . . .	108
4.5.8	Résumé de la section . . . . .	109
4.6	Discussion . . . . .	110
4.6.1	Le choix des sons fictifs, <i>dummy players</i> . . . . .	110
4.6.2	Lien entre détectabilité et agrément . . . . .	110
4.6.3	Résumé de la section . . . . .	112
4.7	Perspectives : expérience in-situ . . . . .	112
4.7.1	Sélection des masques . . . . .	112
4.7.2	Réalisation technique . . . . .	113
4.7.3	Protocole . . . . .	114
4.7.3.1	Première étape : sélection du masque préféré . . . . .	114
4.7.3.2	Deuxième étape : paramétrage du masque préféré . . . . .	114
4.7.3.3	Troisième étape : évaluation de l'acoustique augmentée . . . . .	114
4.7.4	Résumé de la section . . . . .	115
4.8	Conclusion . . . . .	115

<b>5</b>	<b>Proposition d'une analyse de la variabilité inter-participants issue des données de la <i>Best-Worst Scaling</i></b> . . . . .	<b>117</b>
5.1	Introduction de chapitre . . . . .	117
5.2	Inclure les variations inter-participants . . . . .	118

5.2.1	Formalisme statistique . . . . .	118
5.2.2	Résumé de la section . . . . .	120
5.3	Application à notre expérience . . . . .	121
5.3.1	Tests de Student . . . . .	121
5.3.1.1	Confirmation des résultats précédents . . . . .	121
5.3.1.2	Comparaisons statistique des stimuli . . . . .	121
5.3.2	Table ANOVA . . . . .	122
5.3.3	Analyse en composantes principales . . . . .	123
5.3.3.1	Par participant . . . . .	123
5.3.3.2	Par sons . . . . .	123
5.3.4	Résumé de la section . . . . .	125
5.4	Discussion . . . . .	125
<b>6</b>	<b>Conclusion et perspective</b>	<b>127</b>
		<b>129</b>
<b>A</b>	<b>Scores détaillés des participants</b>	<b>131</b>
A.1	Scores détaillés . . . . .	131
<b>B</b>	<b>Traduction du lexique</b>	<b>133</b>
B.1	Traduction du lexique . . . . .	133
<b>C</b>	<b>Limites de l'indicateur psychoacoustique de tonalité</b>	<b>137</b>
C.1	Introduction . . . . .	137
C.2	Echantillons SpeaK . . . . .	138
C.2.1	Présentation . . . . .	138
C.2.2	Comparaison . . . . .	138
C.3	Comparaison ECMA-74 . . . . .	140
C.3.1	Présentation . . . . .	140
C.3.2	Comparaison . . . . .	141
C.4	IRNS . . . . .	144
C.4.1	Présentation . . . . .	144
C.4.2	Comparaison . . . . .	144
C.5	Discussion . . . . .	144
	<b>References</b>	<b>147</b>
	<b>Index</b>	<b>157</b>

# Préface

## Présentation du sujet

L'environnement sonore de l'habitacle du véhicule a été profondément bouleversé par la réduction progressive des sources de bruits existantes. Ceci a rendu audible des sources inattendues, jusqu'à présent inexistantes ou masquées, ou a provoqué une perte d'informations pour le conducteur en faisant disparaître les repères sonores auxquels il était habitué. Ce bouleversement est accentué par la popularisation des véhicules silencieux, en particulier grâce à la motorisation électrique et la disparition du moteur thermique, qui était à l'origine de la signature sonore des véhicules. Ainsi, les habitudes d'écoutes et l'utilisation des sons évoluent, soulevant de nombreuses questions tant pour les constructeurs automobiles que pour les utilisateurs. Quels sont ces nouveaux sons ? Comment sont-ils perçus ? Ont-ils un impact positif ou négatif sur la conduite ? Le silence reste-t-il un objectif souhaitable ? La question du confort sonore se pose-t-elle ? Ces interrogations, parmi bien d'autres, apparaissent spontanément et constituent le point de départ ayant amené à ce sujet de thèse.

À l'origine de ce travail se trouve aussi la volonté du constructeur automobile Renault d'améliorer la satisfaction des utilisateurs de ses véhicules en mobilisant des technologies modernes. Cette démarche s'inscrit dans un changement d'image de marque lié à la révolution électrique initié par le groupe, baptisé « Renaultion ». Parmi ces initiatives, on note l'intégration de technologies inspirées de la réalité augmentée, telles que le développement d'aides à la conduite ou l'adoption de systèmes de visualisation tête haute,<sup>1</sup> permettant de projeter directement sur le pare-brise des informations utiles au conducteur. Cette démarche, visant à utiliser des technologies modernes, s'étend également dans le domaine acoustique. La réduction de bruit, qui est historiquement recherchée, est traitée avec des solutions passives, auxquelles s'ajoutent désormais des solutions actives [4]. Pour augmenter l'environnement acoustique, l'idée émerge de transcrire les technologies de réalités augmentées au domaine du son et de l'acoustique, en complément des solutions de réduction de bruit déjà existantes.

L'évolution de l'environnement sonore dans les habitacles des véhicules électriques, combinée à la volonté de proposer une expérience renouvelée, a ainsi ouvert la voie à une question centrale : est-il possible de concevoir une « réalité augmentée sonore » pour répondre aux nouveaux enjeux du confort sonore ?

Cette thèse s'inscrit dans le cadre d'une convention Convention Industrielle de Formation par la Recherche (CIFRE) entre l'équipe Prestation acoustique de Renault Group, l'équipe Acoustique et vibrations pour l'ingénieur (AVI) du laboratoire Roberval de l'Université de technologie de Compiègne (UTC), et l'équipe Perception et Design Sonores (PDS) du laboratoire Sciences et Technologies de la Musique et du Son (STMS) de l'Institut de Recherche et Coordination Acoustique/Musique (IRCAM).

---

<sup>1</sup> <https://fr.e-guide.renault.com/fra/Austral/la-vision-tete-haute>

Dans ce manuscrit, nous présenterons tout d'abord un état de l'art permettant de mieux cerner le sujet ainsi que le contexte scientifique autour de cette thèse. Cette introduction sera suivie par l'exposé de la problématique centrale ainsi que du plan détaillé du manuscrit. Ce plan est structuré autour de trois thématiques principales qui articuleront les chapitres : la description de l'environnement sonore, la création du son, et l'évaluation du son.

## Note au lecteur

La rédaction de ce manuscrit a été pensée pour une lecture sur un support numérique, permettant d'interagir avec des liens cliquables. L'ensemble des abréviations, références et citations est interactif : il est possible de cliquer pour accéder aux définitions ou aux sources, puis de revenir au point où l'élément a été mentionné pour la première fois. Par exemple, l'abréviation *Compact Disc* (CD) renvoie à la page des abréviations où elle est présentée, et depuis cette page, il est possible de revenir ici en cliquant à nouveau sur CD. Par ailleurs, différents éléments réalisés au cours de la thèse sont librement accessibles en ligne, dont les liens sont insérés dans le corps du texte. Il est tout à fait possible de lire intégralement cette thèse sous un format imprimé, avec toutes les informations et références nécessaires, mais l'utilisation des liens et ressources en ligne est considérablement simplifiée dans la version numérique. Il est aussi conseillé au lecteur de se munir d'un casque audio ou d'un système de restitution audio pour écouter les différents sons ou extraits sonores proposés au fil de ce manuscrit, notamment au moyen de ces liens hypertextes.

Le lecteur trouvera à la fin de chaque section un encart contenant un résumé en quelques lignes des éléments principaux de la section. Ces résumés permettent d'une part, une lecture rapide de chaque étape du manuscrit, d'autre part, une explication simplifiée mais précise des travaux effectués à destination du lecteur non expert.

# 1. Etat de l'art

## 1.1 | Introduction de chapitre

Dans ce chapitre, nous détaillerons d'abord le contexte industriel dans lequel se place cette thèse, l'historique de l'acoustique et du véhicule électrique, et les premières problématiques liées aux véhicules silencieux. Nous examinerons ensuite comment est habituellement définie et étudiée la notion de confort et les différences entre le confort acoustique et le confort sonore. Cette notion de confort sonore est liée à l'interaction de l'utilisateur avec les différentes sources sonores. Les études existantes sur l'impact de la musique au volant, la question de la fatigue induite par les sons ajoutés seront alors présentées, avant d'introduire le concept d'acoustique augmentée, inspiré par la réalité augmentée. Le design sonore est ensuite introduit et se présente comme une solution pour rendre possible cette acoustique augmentée, qui sera une notion centrale dans ce manuscrit. Pour guider ce design et mieux comprendre le confort sonore, les outils issus du champ de recherche des paysages sonores sont détaillés, et en particulier l'approche reposant sur une description par verbalisation. Ils peuvent être utiles à notre cas, car ce champ de recherche se révèle proche de notre définition du confort sonore. Une synthèse d'études menées sur la verbalisation du son est alors présentée. Enfin nous présenterons les objectifs de cette thèse en trois grandes parties autour desquelles s'articuleront les chapitres, qui contiendront eux mêmes des éléments bibliographiques spécifiques à leurs problématiques : la description de l'environnement sonore, la création de son et l'évaluation du son.

## 1.2 | Éléments historique

### 1.2.1 | La place du son dans l'automobile et chez Renault

Historiquement, la problématique du confort acoustique dans les automobiles a découlé des problèmes vibratoires et mécaniques. Les premières automobiles se caractérisaient surtout par de fortes vibrations se propageant à l'ensemble du véhicule, et notamment aux passagers, ainsi qu'un habitacle complètement ouvert,<sup>1</sup> ce qui reléguait les problèmes de confort acoustique et de niveau sonore au deuxième plan. La question de l'amélioration de l'acoustique apparaît avec l'ajout d'un réel habitacle pour isoler l'utilisateur de l'extérieur, ainsi que par le développement de motorisations à l'architecture plus complexe, se traduisant par un régime d'utilisation de moteur ainsi qu'une vitesse d'utilisation du véhicule plus élevée. Un constat technique rapidement effectué est que plus le nombre de cylindres est élevé, moins le moteur vibre ; les motorisations à 6, 8, 10, 12 cylindres se développent, jusqu'à 16 cylindres pour la très luxueuse et aujourd'hui légendaire Cadillac V-16<sup>2</sup> produite en 1930. Les véhicules luxueux commencent alors à avoir une signature sonore distincte et reconnaissable, synonyme de luxe ou de confort. L'amélioration des matériaux constituant l'habitacle, qui isole l'utilisateur de l'extérieur le confine alors avec ces bruits de fonctionnement, qui deviennent audibles et potentiellement source de

<sup>1</sup> <https://www.youtube.com/watch?v=DWhwCCRLSm0>

<sup>2</sup> <https://www.youtube.com/watch?v=v3syL4SF-3g>



**Figure 1.1.** Carnet d'utilisateur d'une Renault 16, datant de 1967, vantant une insonorisation « particulièrement poussée » quelques années après la création d'un service entièrement dédié à l'acoustique.


gêne pour l'utilisateur. La problématique de l'acoustique automobile voit alors le jour.


Un centre technique dédié à l'acoustique se forme chez Renault pour la première fois en 1965, à Rueil avec une modeste équipe de quelques personnes. Il faut dire qu'à l'époque les moyens de mesure et le domaine restaient nouveaux dans le monde automobile. On peut cependant trouver dès 1967 les premiers arguments marketing autour des capacités acoustiques d'un modèle (fig. 1.1). La question de l'acoustique prend une place de plus en plus importante, et les moyens techniques et humains alloués augmentent progressivement. Les premières difficultés techniques apparaissent : un bruit sourd et fatigant est identifié sur les premières Renault R5, qui sera appelé « bourdonnement » et deviendra une préoccupation majeure des acousticiens pour des années, ou encore le « sirènement de triplette » sur la Renault 14. Une culture du son se construit ainsi progressivement chez le constructeur autour de l'acoustique des véhicules, accompagnée d'un vocabulaire spécifique, en constante évolution.

Le département acoustique gagne en importance en même temps que les moyens techniques associés. Les enregistreurs magnétiques Nagra, adaptés à la mesure dans les véhicules grâce à une voie tachymétrique, sont largement utilisés au sein du service dans les années 70 et 80. En 1979, Maurice Mergeay fonde *Leuven Measurement System* (LMS) ; c'est le début d'un outil spécialisé dans l'analyse numérique de mesures acoustiques et vibratoires, encore largement utilisé aujourd'hui. En 1995 le service acoustique de Renault est déplacé au centre technique d'Aubevoye avec de nombreux moyens expérimentaux allant de chambres sourdes jusqu'à un auditorium, en passant par un atelier de préparation des véhicules et des pistes d'essais. Une véritable expertise acoustique se construit au sein du constructeur ainsi qu'une forte culture du son. Un vocabulaire très riche se développe progressivement autour des habitudes d'écoute et des ressentis de chacun, et en 1995 Jean-Christophe Kraemer propose un Audioscope largement partagé à l'intérieur du groupe qui consiste en un CD contenant une centaine d'extraits sonores accompagnés d'un manuel détaillant les sons (Fig. 1.2). Un réel glossaire des sons automobiles couramment

## BRUITS DE CARROSSERIE

On cherche à supprimer les grincements de planches de bord, sifflement de portières et autres bruits parasites qui, par leur présence, voire leur omniprésence agacent le client et nuisent à l'image de finition et de qualité des véhicules. Le claquement des portières, par exemple, mat ou ouaté est synonyme de qualité à l'oreille du client, alors qu'un bruit métallique peu amorti ne l'est pas.

17 • **Le crissement de lécheur de vitre** provient du frottement de la partie caoutchouc du lécheur sur la vitre. 


18 • **Le grincement à la descente de la vitre** a la même origine que le bruit précédent. On parle alors d'un phénomène de " Stip and Slik " comparable aux sons que l'on déclenche en frottant un doigt mouillé sur le rebord d'un verre. 

19 • **Le bruit lié au lâcher de poignée de virage** correspond au choc de cette poignée sur son support. IL est possible d'interposer un élément souple pour amortir ce choc. Cela correspond à

**l'enregistrement 19.** 

20 • Le bruit se produisant pendant la **manoeuvre de la rehausse de siège** correspond au bruit de moteur électrique, d'engrenage, et frottement. 

21 • **Le bruit de glissière de siège** à commande électrique est très proche du bruit précédent.

22 • **Le couinement d'arrêt de porte** avant le claquement à la fermeture correspond à un bruit de frottement aigu. 

23 • **Le claquement de porte** revêt une certaine importance. C'est le premier contact auditif du client avec son véhicule. Il est possible de modifier le bruit de claquement de porte à la fermeture en amortissant les panneaux de tôle, en modifiant les gâches et les serrures, etc..

**Figure 1.2.** Manuel accompagnant l'audioscope, partagé en interne dans le groupe Renault. Ce manuel détaille le nom et l'origine techniques des sons habituellement rencontrés dans l'automobile. Aujourd'hui la plupart des sons identifiés appartiennent aux voitures du passé et leurs descriptions sont devenues obsolètes, mais témoignent des habitudes d'écoute et de l'histoire des sons par laquelle est passée l'automobile.

rencontrés est alors construit et diffusé. Une expertise acoustique se développe, construite autour d'habitudes d'écoutes, de vocabulaire et de travail développées à l'aide des décennies d'expériences accumulées.

Avec le développement de l'utilisation de média dans la voiture comme la musique d'abord écoutée par la radio, puis choisie par l'utilisateur grâce à un autoradio, un lecteur CD ou une connexion Bluetooth. L'acoustique de la voiture ne se restreint plus seulement aux sons du véhicule, mais s'étend aussi au système audio, permettant à l'utilisateur de jouer les sons de son choix. La qualité d'un système de restitution devient importante et de nombreux partenariats apparaissent entre les constructeurs automobiles et des équipementiers audio réputés.<sup>3,4</sup> Les gammes premium incluent volontiers un nombre de haut-parleurs croissant, pouvant facilement dépasser la dizaine de haut-parleurs dans un habitacle, et il est même possible de changer entièrement le système audio de sa voiture pour choisir un système promettant une meilleure qualité de restitution.<sup>5</sup> En parallèle du développement de ces technologies de restitution au sein du constructeur, une identité sonore est conçue par la société Sixième son.<sup>6</sup> D'un point de vue marketing, elle permet de garder une sonorité cohérente pour les différents éléments publicitaires créés, mais aussi de démontrer les capacités des systèmes audio des véhicules lorsque le jingle sonore est joué au démarrage du véhicule. Cette identité traduit aussi la volonté de véhiculer une image de marque à travers un son créé.

<sup>3</sup> <https://automotive.bose.fr/vehicules/renault>

<sup>4</sup> <https://www.arkamys.com/alpine-arkamys/>

<sup>5</sup> <https://www.focal-inside.com/renault/fr/>

<sup>6</sup> <https://www.sixiemeson.com/fr/nos-creations/sonic-branding-fr/renault-identite-sonore/>



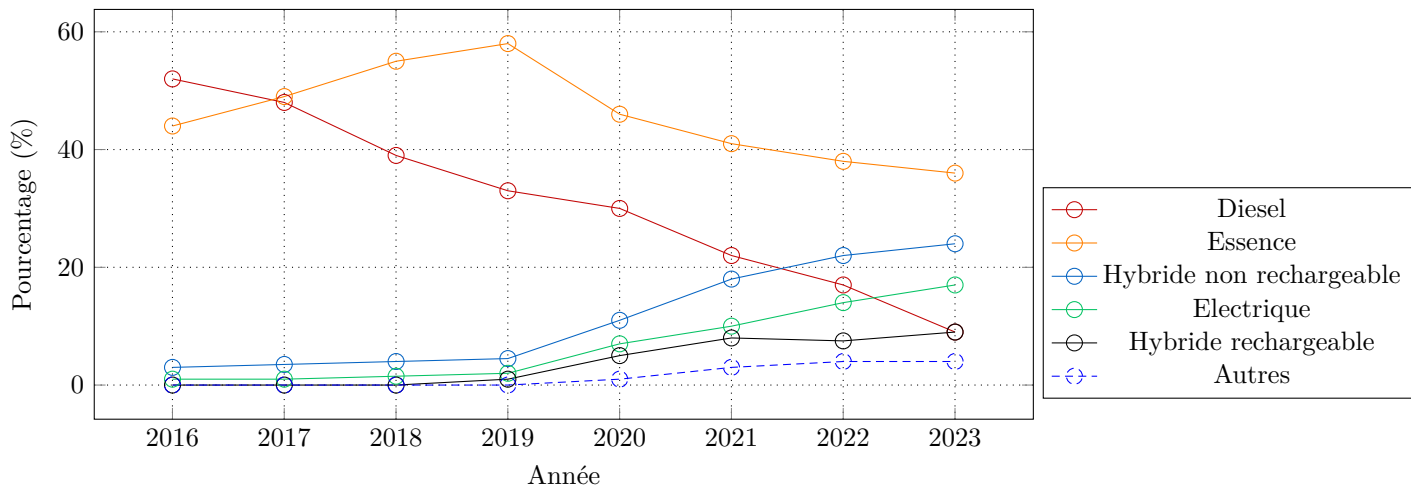
**Figure 1.3.** A gauche, *La Jamais Contente*, première voiture à dépasser les 100 km/h, exposée au Musée national de la voiture et du tourisme de Compiègne. A droite la Venturi Fétish, première voiture de sport électrique commercialisée, au Mondial de l'automobile à Paris en 2006. Ces deux voitures, bien que possédant des performances comparables ou supérieures à leurs concurrentes thermiques, ne suffiront pas à elles seules à populariser la technologie électrique.

### 1.2.2 | Histoire du véhicule électrique

Les premiers véhicules électriques font leurs apparition très tôt dans l'histoire de l'automobile. Leurs performances sont alors significativement meilleures, comme démontré par *La Jamais Contente* (Fig. 1.3) qui établit un record de vitesse en franchissant pour la première fois la barrière des 100 km/h en 1899. On assiste quelques années plus tard à l'apparition de la première voiture électrique de série dès 1907 avec la *Detroit Electric*. L'automobile électrique constitue alors une alternative sérieuse à la motorisation thermique. L'histoire montrera que les meilleures performances techniques n'ont pas suffi à imposer ce type de véhicule, les technologies des batteries et la nécessité d'avoir un réseau d'électrification efficace ne pourront concurrencer le plus faible coût de la motorisation thermique et la facilité de stockage du carburant. Le véhicule thermique s'impose progressivement et connaît un réel essor avec le boom économique de l'après-guerre, celui-ci s'installe comme un bien de masse et symbole de liberté individuelle.

La technologie de la voiture électrique connaît cependant un regain d'intérêt à la suite du premier choc pétrolier (1973) à partir duquel les industriels se réintéressent à la possibilité d'un véhicule électrique, indépendant des hydrocarbures. On peut alors citer le projet avant-gardiste de Renault avec la 4L 100 % électrique. Restée à l'état de projet, elle illustre la volonté de se réappropriation la technologie électrique dans l'automobile. En 1996, General Motors propose à son tour un véhicule électrique produit en série destiné au grand public, l'EV1, près de 90 ans après la première voiture électrique en série. Ce modèle, bien que pionnier, ne connaîtra qu'un succès modeste et sera rapidement abandonné par la marque. D'autres véhicules font alors progressivement leur apparition et misent plutôt sur l'efficacité et la puissance des motorisations électriques, tels que la Venturi Fetish en 2004 (Fig. 1.3) ou la Tesla Roadster I qui sera produite en série à partir de 2008. Mais le coût très élevé de ces véhicules et leur faible autonomie restreignent ces véhicules au marché de niche du véhicule sportif ou de collection.

Il faut réellement attendre la rupture technologique sur le stockage de l'électricité, qui permet la baisse progressive du coût des batteries lithium-ion au cours des années 2000-2010 ainsi qu'une volonté étatique de subventionner le développement de l'automobile électrique avec des dispositifs tels que le bonus écologique en France ou *Umweltbonus* en Allemagne, pour observer une réelle révolution électrique sur le marché automobile. Renault se place alors en pionnier du véhicule électrique, avec le développement et la vente de son premier véhicule électrique de série destiné au grand public : la



**Figure 1.4.** Évolution des parts de véhicule immatriculés par année, en fonction du type de motorisation. Données issues du baromètre du véhicule électrique 2024, publié par l'ANFA

Renault ZOE, devenue véhicule électrique le plus vendu en France en 2013 et symbole de la direction technologique prise par la marque. Le véhicule électrique n'est plus un concept, il devient alors possible, et l'électrification du monde automobile commence. Bien que l'électrification du parc automobile actuel soit aux alentours des 2 %, c'est bien les véhicules conçus et vendus aujourd'hui qui occuperont une grande place dans le paysage automobile de demain puisque l'observatoire de l'Association Nationale pour la Formation Automobile (ANFA) estime à 17 % la part de véhicules neufs entièrement électriques pour l'année 2023 (fig. 1.4), une part qui monte à 50 % si on y inclut les véhicules hybrides. De plus, la part de véhicules électriques est de facto amenée à augmenter, avec l'adoption de la loi européenne entrée en vigueur en avril 2023, sur l'interdiction à la vente de véhicule particulier à motorisation thermique d'ici 2035.<sup>7</sup> Outre les problématiques techniques liées au stockage de l'énergie, le temps de charge ou l'autonomie du véhicule, de nombreuses et nouvelles questions apparaissent avec ce véhicule.

### 1.2.3 | Le problème du silence

La démocratisation du véhicule électrique, et plus généralement du véhicule silencieux, amène une remise en question complète de l'utilisation du son dans la voiture, que ça soit la signature marketing, la réaction émotionnelle du conducteur, ou même comme marqueur social [5, 6]. Le cas d'étude le plus emblématique autour de cette révolution sonore réside probablement dans ce qui a donné naissance à l'*Acoustic Vehicle Alerting System* (AVAS). En effet, l'adoption d'un véhicule à motorisation silencieuse posait la question de sa faible détectabilité dans les plages de vitesses où les niveaux sonores du roulement et de l'aéroacoustique de la voiture sont faibles (<30 km/h) ; le piéton perd les repères sonores auxquels il était habitué [7, 8]. Une solution alors explorée consiste à ajouter un son artificiel synthétisé [9, 10, 11], ou plus récemment mécanique [12], permettant d'avertir le piéton de la présence d'un véhicule en fonctionnement. Ce son artificiel est conçu avec l'objectif de rendre un objet roulant plus détectable, mais aussi avec plusieurs contraintes liées au design : la contrainte d'inclure ce son dans l'environnement urbain déjà bruyant et chargé, ou la contrainte de ressembler à ce qui est déjà attendu pour ce type de son dans l'imaginaire collectif ou dans des œuvres de fictions. Cette signalétique est aujourd'hui devenu une signature sonore assez iconique de la Renault ZOE, créé par le compositeur Andrea Cera en collaboration avec l'équipe PDS de l'IRCAM. Cette signalétique sonore est aujourd'hui répandue dans les véhicules du constructeur, et aisément reconnaissable à l'écoute.<sup>8</sup> Cette solution

<sup>7</sup> <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32023R0851>

<sup>8</sup> <https://www.youtube.com/watch?v=1XUzX2gXq8c>

s'est progressivement propagée au sein des constructeurs, offrant une solution technique intéressante pour répondre à la législation européenne entrée en vigueur en 2019 pour tous les véhicules électriques neufs, qui impose un niveau sonore minimum en dessous d'une certaine vitesse,<sup>9</sup> puis dans le monde.<sup>10</sup>

La question du silence se pose également pour l'utilisateur du véhicule. De nombreuses études, notamment dans le sport, montrent l'importance du retour sonore pour l'apprentissage ou le perfectionnement d'un geste [13], la disparition de ce retour sonore entraînant l'augmentation du taux d'erreur [14], ou une perte de précision, même pour des sportifs de haut niveau [15]. La question se pose aussi pour le conducteur du véhicule électrique. Celui-ci ne peut plus s'aider du son du moteur pour compléter les informations qu'il a de la dynamique du véhicule [16, 17, 18], ou pour deviner la sportivité d'une voiture au vrombissement de son moteur [19]. Une partie de l'imaginaire et de la culture du son autour du véhicule se retrouve ainsi définitivement modifiée. La disparition de ces objets sonores, bien qu'ils contribuent à une baisse significative du niveau sonore dans l'habitacle, ne supprime pas la question de l'acoustique automobile [20]. Au contraire, l'effacement de ces objets au fort pouvoir masquant, tel que le moteur thermique, laisse apparaître de nouvelles sources sonores, plus faibles en niveau, mais bien perceptibles, comme le moteur électrique, encore inhabituel pour le conducteur [21, 22] ou même le son de l'AVAS nouvellement créé. La question de la faible empreinte sonore, qui se posait spontanément pour le piéton se pose aussi pour la conduite dans le silence. Les études sur simulateurs prouvent que ce manque de retour sonore provoque une sous-estimation de la vitesse du véhicule par son utilisateur [18, 23]. Cet effet reste limité en pratique grâce à l'usage du compteur de vitesse : les véhicules silencieux ne provoquent pas de conduite significativement plus rapide ou dangereuse. Hellier et al. [17] montrent en plus de cette sous-estimation de la vitesse du véhicule, un inconfort accru en condition entièrement silencieuse sur simulateur. Un constat se dessine à l'issue de ces premières études : ce silence permet l'émergence de sources sonores inhabituelles pour le conducteur, change ses habitudes d'écoutes lors de la conduite et impacte donc le confort perçu. La nécessité se dessine alors, de mieux comprendre cette notion de confort et d'inconfort dans ce nouvel environnement, et en particulier d'explorer des solutions inspirées de technologies modernes telles que la réalité augmentée. Une approche autre que le traitement passif des sources acoustiques est donc utilisée pour tirer profit de la place importante qu'a prise l'audio dans le véhicule en vu d'améliorer le confort sonore.

<sup>9</sup> <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019R0839>

<sup>10</sup> [https://eur-lex.europa.eu/eli/reg/2017/71\(1\)/oj](https://eur-lex.europa.eu/eli/reg/2017/71(1)/oj)

### 1.2.4 | Résumé de la section

#### Histoire

Cette section permet de se familiariser avec l'histoire de l'acoustique au sein du constructeur Renault, ainsi qu'avec celle du véhicule électrique. Ainsi, on constate que l'acoustique automobile a une histoire riche et fortement influencée par le moteur thermique autour duquel un vocabulaire et des habitudes se sont construits. Plus récemment, un réel enjeu autour de la qualité des systèmes de restitution s'est développé, et il n'est pas rare de voir un véhicule de série proposé avec un système audio complexe de neuf haut-parleurs ou plus. La qualité de l'audio joué est jugée au même titre que la qualité du son de fonctionnement. L'avènement du véhicule électrique bouleverse alors une grande partie des habitudes des utilisateurs. Bien que cette technologie électrique ne soit pas entièrement nouvelle et ait fait fréquemment l'objet de prototypes ou de concept-cars, c'est bien au cours de ces dernières années que celle-ci s'est démocratisée pour le grand public. Cette production en masse de véhicules électriques a alors révélé de nouveaux défis liés au son : un avertisseur piéton a été conçu puis standardisé (l'AVAS), de nouvelles sources sonores émergent dans le véhicule, produites par la motorisation électrique, les habitudes d'écoute des utilisateurs lors de la conduite sont bouleversées.

La problématique guidant cette thèse évolue alors : puisque le système de restitution audio a gagné en importance et en qualité au cours des dernières années ; face à ce changement d'environnement sonore impactant les habitudes d'écoutes et laissant apparaître de nouvelles sources inattendues, est-il possible d'améliorer le confort sonore perçu à l'aide du système audio de la voiture ?

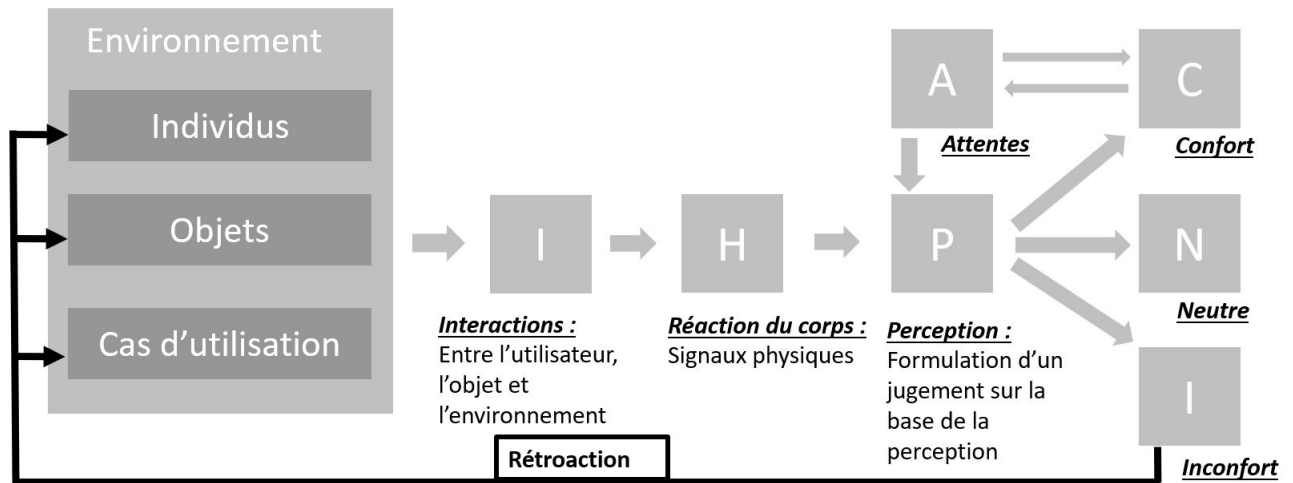
## 1.3 | Du confort au design sonore

### 1.3.1 | Définition du confort

#### 1.3.1.1 | Ce que n'est pas le confort

Le confort est une notion centrale dans ce travail et probablement la notion abordée dans cette thèse la plus universellement partagée. Selon Le Goff [24], cette notion de confort émerge lorsque la société atteint un certain statut qui permet l'apparition d'un « droit au confort ». Ce droit s'est démocratisé avec le développement de la société de consommation. La production de la plupart des objets a suivi un procédé qui s'est complexifié pour permettre leur production industrielle. Cette complexification a donné au travail de design et de conception une grande importance, on s'attend alors à ce que chaque objet manufacturé ait suivi ce processus. Leur confort devient donc une attente spontanée de la part des utilisateurs, ainsi qu'un objectif naturel pour l'ingénieur en amont de leur production. En revanche, bien qu'universellement partagée et attendue, cette notion reste délicate à définir, et pour s'atteler à cette tâche, il est sûrement plus facile de commencer par ce que n'est pas le confort.

Tout d'abord, le confort n'est pas le luxe [25]. Cette distinction est particulièrement importante dans l'industrie automobile où la notion de luxe est omniprésente. Le luxe, est par définition ostentatoire et superflu, il se présente comme inaccessible. Il ne s'agit alors pas du concept de confort poussé à l'extrême, mais bien d'un concept différent. On peut qualifier un objet de luxueux sans qu'il soit confortable dans son utilisation, ou qualifier un objet de confortable sans qu'il soit luxueux. Le confort n'est pas non plus la qualité. Le concept de qualité dans le milieu industriel est normé, on parle ainsi de normes de qualité garantissant un niveau de standard sur les propriétés d'un objet que l'on est capable



**Figure 1.5.** Modèle de confort mettant en évidence les mécanismes aboutissant à la sensation de confort et d'inconfort, traduit de celui présenté par Vink et Hallbeck [29], comme par exemple la boucle rétroactive provoquée par une sensation d'inconfort, sur les éléments de l'environnement.

de quantifier, mesurer et comparer entre les produits. Cette qualité devient alors, par conception, indépendante du contexte d'utilisation d'un objet. Il s'agit d'une propriété intrinsèque d'un objet.

On peut alors être amené à définir le confort par une définition négative, « *Le confort c'est l'absence d'inconfort* ». Cette première définition utilisée par Hertzberg [26] offre une base claire et a marqué les premières recherches sur le confort. L'étude lexicologique menée par Delepaut [25] retrace l'évolution du terme dans la langue française, initialement utilisé au Moyen Âge et associé au domaine de l'intimité. Après une disparition progressive de son usage, le terme réapparaît après un emprunt par la langue anglaise pour décrire une notion matérielle, participant au bien-être et à la commodité de vie, comme défini actuellement par le Larousse : « Ensemble des commodités, des agréments qui produit le bien-être matériel ». Cette observation lexicologique prouve l'existence du concept de confort avant celui d'inconfort. Ce dernier est alors défini comme une insuffisance de confort et non pas comme une absence de confort. L'inconfort n'est pas le strict opposé du confort, ce qui remet en question la définition de celui-ci par la négative, proposée plus tôt. En complément de cette étude lexicologique, Zhang quantifie ce constat [27], en identifiant parmi une liste de descripteurs qui peuvent se rapporter au confort ou l'inconfort de sièges, l'asymétrie qui existe entre les notions liées au confort, et celles à l'inconfort. Ainsi l'inconfort est plutôt lié à des notions physiques, de douleur ou de raideur, alors que le confort est plutôt relié à des concepts plus haut-niveau, tels que l'appréciation, le bien-être ou le calme. L'axe de recherche diffère selon que la problématique porte sur l'inconfort ou, comme dans notre cas, sur l'amélioration du confort.

### 1.3.1.2 | Les modèles de confort

Les travaux de Zhang et la mise en évidence de facteurs indépendants affectant le confort et l'inconfort ouvrent alors une perspective de modélisation du confort [28]. Initialement développés par les études sur les objets tels que des sièges, les modèles de confort construits sont généralisables. En particulier, le modèle construit par Vink et Hallbeck (voir Fig.1.5) [29], qui fait nettement apparaître la différence de facteurs agissant sur le confort et l'inconfort. Ce modèle ajoute un état intermédiaire, neutre, caractérisé par à la fois une absence de confort et d'inconfort.

Ce que montre aussi cette figure c'est l'impact important du contexte d'utilisation, de l'environnement

des individus, sur le confort. De plus, la double flèche entre les attentes et le confort matérialise l'effet d'attente sur le confort : l'utilisateur a une idée préconçue du confort, et l'expérience d'une situation confortable peut changer cette idée pré-existante. Enfin la boucle de rétroaction produite par l'inconfort montre la tendance d'un utilisateur à agir sur l'environnement pour modifier la situation perçue comme inconfortable. En revanche, cette boucle n'existe pas dans le cas d'une situation neutre ou confortable, l'utilisateur ne cherchera pas spontanément à maximiser son confort. Améliorer le confort, et par extension le confort sonore, nécessite en premier lieu d'identifier les points responsables d'un inconfort pour au moins atteindre un état neutre. Dans un second temps, cette amélioration suppose de pouvoir évaluer le confort perçu, afin de distinguer une situation neutre d'une situation confortable. Selon ces modèles, ces solutions apportées pour atteindre cet état ne viennent pas de l'utilisateur lui-même, qui cherche seulement à ne pas avoir de situation inconfortable. Les solutions viennent alors plutôt d'un acteur extérieur et doivent donc être des propositions originales, que l'évaluation validera ou non. Ces propositions originales sont influencées par des attentes ou idées pré-conçues relatives au contexte où à l'utilisation d'un objet. C'est donc pour cela que ce travail d'amélioration du confort doit faire l'objet d'un design [30].

### 1.3.1.3 | Ce qu'est le confort

Les précédentes études montrent que la complexité de la notion de confort ne peut se satisfaire d'une définition négative. Une fois fixée sur ce que n'est pas le confort, on cible maintenant ce qu'est le confort. Le manque de consensus sur la définition n'empêche pas d'identifier plusieurs points précis de la notion de confort faisant eux consensus, identifiés par [28].

- Le confort est une construction subjective par nature .
- Le confort est affecté par des facteurs variés de natures différentes (physique, physiologique, psychologique).
- Le confort est une réaction à l'environnement.

De nombreux éléments de la définition de Slater [31] sont retrouvés dans ces trois points, souvent prise comme référence encore aujourd'hui et traduite ici : « *Un bien-être physiologique, psychologique et une harmonie physique entre un individu et son environnement* ». Un reproche possible à cette définition, identifié par Le Goff [24], est de ne pas inclure la notion d'objet et son utilisation. L'importance de la dimension contextuelle relative à cette utilisation [32] qui est visible sur le modèle de confort présenté plus tôt est aussi absente de cette définition. Ces éléments définissant le confort nous serviront alors à construire une définition du confort dans le cadre de cette thèse, et plus particulièrement du confort sonore.

## 1.3.2 | Du confort acoustique au confort sonore.

### 1.3.2.1 | Le confort acoustique

Ce confort, utilisé pour décrire des objets matériels, tel qu'un siège ou un bureau, s'applique tout aussi bien pour des objets immatériels tels qu'un logiciel, une situation, ou pourquoi pas des sons. Les notions de confort acoustique et de confort sonore sont présentées pour la première fois par Chelkoff dans le cadre d'études sur l'architecture et l'habitation [33, 34]. Le confort acoustique y est présenté comme contribuant à l'amélioration de l'intimité. Cette caractéristique rappelle le sens initial du mot confort dans la langue française. Cette notion de confort acoustique, permise par des solutions techniques, telles que des matériaux plus isolant, vise à réduire des sources physiques de « bruits », entendus comme des

sons non souhaités.

Ce confort acoustique vise un état neutre, qui est d'ailleurs défini dans les standards comme une absence de gêne ou de bruit extérieur, rappelant la définition négative du confort mentionnée plus tôt.

### 1.3.2.2 | Le confort sonore

À ce concept de confort acoustique, s'ajoute le concept, complémentaire, de confort sonore. Celui-ci se traduit par non plus seulement une isolation de l'habitat par rapport à l'extérieur, mais aussi de permettre l'existence de sources choisies ou pertinentes pour l'utilisateur, et cohérentes avec ses attentes ou ses habitudes. Ce type de confort n'est plus mesurable physiquement, comme pourrait l'être l'isolation phonique ou le niveau sonore. Pour illustrer ce concept de confort sonore, des éléments techniques, tels que des portes ou des volets qui permettent, selon la volonté de l'utilisateur, de s'isoler ou s'ouvrir aux sources sonores externes peuvent être cités. Un autre élément contribuant au confort sonore est l'utilisation de médias tels que la radio ou la télévision dans des pièces à vivre. On peut aussi penser à des éléments mêlant architecture et acoustique dans la conception d'une pièce telle qu'un salon, où l'intelligibilité peut être clef. Enfin, dans un cadre plus large, des installations sonores plus complexes peuvent être prises en exemple, conçues pour ajouter du son dans le hall d'une résidence comme c'est le cas pour le Piezoplex.<sup>11,12</sup> Une installation qui vise à informer les résidents de l'état énergétique du bâtiment, cohérent avec l'objectif de neutralité énergétique, installé dans le hall de l'immeuble, pour lequel des sons extérieurs peuvent être attendus, en gardant un niveau sonore assez faible pour maximiser l'agrément.

Ce confort sonore ne vise plus seulement un état neutre, mais ajoute des éléments, cohérents avec le contexte, ou l'utilisation qu'il sera fait d'une pièce dans le cadre de l'architecture ici.

### 1.3.2.3 | Le confort sonore dans la voiture

Ces réflexions sur le, ou les confort, ont été faites autour de l'utilisation du logement, en réaction à l'évolution qu'il a eu au cours des années. Cependant l'utilisation actuelle de la voiture se rapproche de plus en plus de l'utilisation d'un lieu à vivre et est de moins en moins considérée uniquement comme un moyen de locomotion. Cette évolution est amenée à se poursuivre dans ce sens, notamment avec l'utilisation du véhicule électrique qui nécessite des temps de recharge pouvant être long, l'habitacle est pensé comme un lieu où rester, même quand le véhicule ne roule pas. Le terme de « salon roulant » est fréquemment utilisé pour décrire une voiture confortable, et des termes tels que « cocon » ou « isolation acoustique » sont volontiers aussi utilisés pour qualifier l'habitacle de la voiture aussi bien que pour l'habitat [33, 35]. À cela s'ajoute l'utilisation commune aux deux milieux, de média, tels que la radio ou la musique. Leur utilisation est en effet de plus en plus facilitée par des interfaces au sein de l'habitacle pour la voiture. La similitude entre ces deux milieux se prolonge aussi dans le fait que les interactions avec le monde extérieur se fait dans les deux cas à travers des portes ou des fenêtres, qui deviennent alors des éléments importants du confort. Cette proximité entre l'habitat et la voiture, d'un point de vu du confort sonore, semble paradoxalement plus grande qu'avec d'autres moyens de transports tels que le train [25, 36, 37, 38] ou l'avion [30] pour lesquels un axe majeur de l'amélioration du confort réside dans la résolution du problème lié à la promiscuité avec autrui.

<sup>11</sup> <https://www.sorbonne.fr/maison-ile-de-france-batiment-pionner-ciup/>

<sup>12</sup> <https://soundcloud.com/menure/sets/piezoplex-la-voix-du-batiment>

À l'issue de cette réflexion, on converge alors vers une définition du confort sonore appropriée à l'habitacle du véhicule. Cette définition prend en compte les différents points soulignés : le côté contextuel de cette évaluation, qui ne peut être ressenti que dans l'utilisation d'un objet, ainsi que la différence avec le confort acoustique et la réduction de niveau acoustique :

« *Le confort sonore est ce qui contribue à une sensation agréable dans l'utilisation d'un objet, en terme de retour sonore, cohérent avec les attentes de l'utilisateur et l'environnement sonore déjà existant.* »

### 1.3.3 | Ajouter du son

#### 1.3.3.1 | Sons intentionnels

Le silence de l'habitacle du véhicule électrique, comparé à son homologue thermique, contribue-t-il à l'amélioration du confort sonore ? En l'état actuel des connaissances, il n'est pas possible de répondre à cette question avec certitude. En revanche, les éléments inclus dans la définition, tels que l'agréabilité perçue, l'importance des retours sonores et la cohérence des éléments sonores avec les attentes de l'utilisateur, suggèrent que le silence complet ne permette pas nécessairement d'atteindre un confort optimal. Ainsi, cette thèse explore la possibilité d'ajouter des éléments sonores contribuant positivement au confort sonore en agissant sur les différents aspects mentionnés plutôt que sur la réduction du niveau sonore seule. Différents sons intentionnels, comme nommés par Langeveld [39] existent et peuvent être, ou sont déjà, ajoutés dans l'habitacle du véhicule. Celui-ci identifie quatre catégories de sons intentionnels, qui peuvent être volontairement créés. Les deux premières réagissent à une action d'un utilisateur est font d'un fonctionnement d'Interface Homme-Machine (IHM), ce sont les "*earcons*"<sup>13</sup> et "*auditory icons*"<sup>14</sup> qui font parties des sons musicaux discrets. A cela s'ajoute les sons continus, "*sonification*" et "*continuous sonic interaction*" qui sont les deux autres catégories identifiées. Dans notre cas nous nous intéresserons uniquement aux sons continus, susceptibles d'exister tout le long d'une tâche de conduite, et pas aux sons tels que les "*earcons*" ou "*auditory icons*", qui apparaissent ponctuellement.

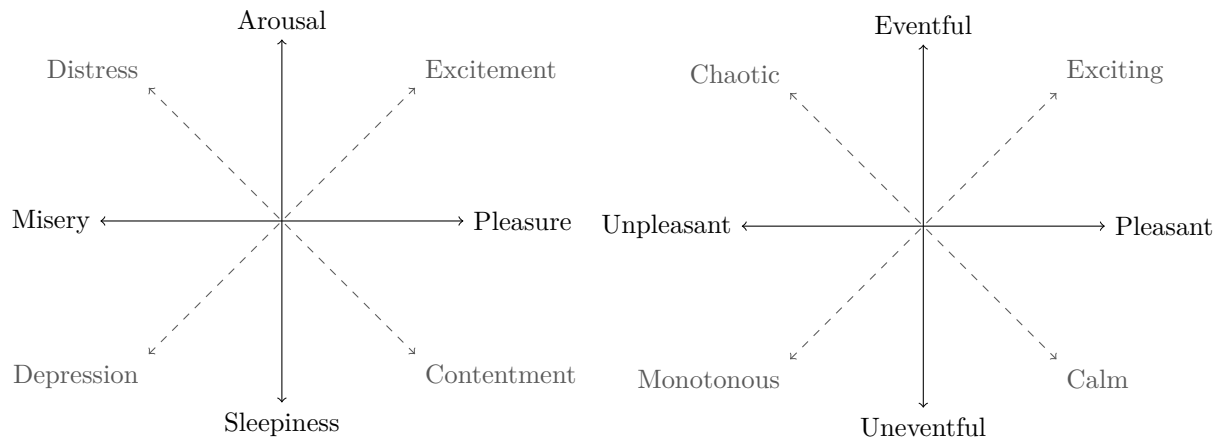
La sonification est définie comme « *l'utilisation de sons non parlés pour véhiculer de l'information [...].* » tel que formulé par Thomas Hermann<sup>15</sup> [40]. Elle peut alors être utilisée pour améliorer les performances dans une tâche, à partir d'informations complémentaires ou en fournissant une redondance. On peut ainsi citer les études dans le champ de la sonification prouvant les meilleures performances obtenues pour un suivi de cible avec un retour sonore direct [41]. Pour se rapprocher du cas de l'habitacle, l'étude de Tardieu et al. [42] met en évidence la diminution du temps alloué à la manipulation d'une IHM sonifiée, au profit de la tâche de conduite. Ces cas d'études, bien qu'ils montrent l'effet positif du son pour améliorer les performances dans une tâche, ne permettent pas de conclure sur un effet sur le confort lors de l'utilisation de la voiture. A cela s'ajoute le fait que l'évaluation du confort est fortement contextuelle, comme le montre l'état de l'art, et moins facilement mesurable qu'un temps de réaction ou une erreur dans un suivi de cible. L'approche de la sonification, qui par définition cherche à véhiculer une information n'est donc pas le type de son ajouté qui sera exploré pour notre problématique.

Les études faites dans les bureaux ouverts sur des sons indésirables montrent la possibilité d'améliorer un ressenti à l'aide de sons ajoutés, et potentiellement l'agréabilité, qui est un des éléments de notre

<sup>13</sup> "The examples given so far mainly consisted of beep-like sounds [...] the abstract mapping of earcons must be learned, as there is no semantic link between the sounds and the data they represent. Differentiation is commonly found in terms of pitch, rhythm, timbre, spatial location, duration, and tempo" [39]

<sup>14</sup> "Contrary to earcons, [auditory icons] are natural, everyday sounds, which are described in terms of their sources (e.g., the air flow sound of a fan to represent the state of a steam vent)." [39]

<sup>15</sup> <https://sonification.de/son/definition/>



**Figure 1.6.** À gauche, le cercle circomplexe des émotions issu du modèle de Russel [50] pour évaluer les émotions perçues ou ressenties, et à droite son adaptation [51] pour évaluer un paysage sonore.

définition de confort. L'étude de Brocolini et al. [43] porte sur l'effet de sons masquants tels que du bruit blanc sur les signaux comme la parole, reconnus comme indésirable dans ce contexte [44]. Bien que l'étude citée conclu sur une absence d'effets positifs de ces masques sur la performance, elle montre cependant leurs effets sur l'intelligibilité des signaux indésirables. L'étude de sons masquants plus riches tels que des sons de cascade sur un environnement sonore industriel désagréable [45] ou de sons naturels sur un paysage sonore avec du bruit de trafic [46, 47] montrent alors qu'il est possible de modifier le jugement d'agrément ou de gêne perçues dans ces scènes. On s'intéresse alors aux sons qu'il est possible d'ajouter dans l'habitacle pour améliorer le confort.

### 1.3.3.2 | Impact de la musique

Lorsque l'on pense à des sons intentionnellement ajoutés, l'écoute de la radio ou de la musique dans la voiture vient alors à l'esprit. Cette utilisation de média extérieurs a déjà largement modifié l'utilisation et les attentes autour de l'habitacle du véhicule au cours de son histoire [48]. Ces sons ajoutés dans le véhicule posent des questions de l'impact sur l'utilisateur, autre que le confort, par exemple l'impact du son sur le comportement, les humeurs ou les émotions au volant. Par exemple, Rochesso et al. étudient la sonification continue d'objets et les interactions avec les utilisateurs, l'étude offre alors des premiers retours sur expériences faits sur ce sujet, à partir d'observations effectuées sur les équipes de designer et les participants interagissant avec les objets sonifiés, notamment un cafetière Moka. Bien que la sonification soit appréciée par les utilisateurs, celle-ci a alors favorisé des manipulations excessives ou inattendues de l'objet, au point de casser un capteur de pression qui y était installé [49]. Cet exemple, assez anecdotique, traduit cependant une préoccupation réelle sur les changements de comportements ou d'utilisations que l'on peut faire d'un objet, une fois augmenté. Dans le cadre de l'automobile, où la sécurité routière est une priorité, la question se pose alors de savoir si l'acoustique augmentée proposée ici risque de provoquer un comportement excessif ou inattendu.

L'effet du son, et plus particulièrement de la musique, sur les humeurs et les émotions a déjà été mis en évidence dans plusieurs études, notamment l'effet du tempo sur la stimulation [52] de participants, ou sur les émotions ressenties ou perçues [53, 54]. Des outils tels que le questionnaire *Profil of Mood States* (POMS) [55, 56] ou l'utilisation de l'espace à deux dimensions proposées par Russell [50, 57] permettent de quantifier l'impact d'un son ou d'une musique sur les émotions ou l'humeur. Cette quantification est alors faite selon deux composantes principales représentées sur la figure 1.6, la stimulation (*Arousal*) et l'agrabilité (*Pleasantness*). Cette méthode d'évaluation a en particulier

trouvé un écho dans l'évaluation que l'on peut faire d'un paysage sonore [51].

Cette modification de l'humeur et des émotions par la musique est aussi à l'oeuvre au volant [58]. Dans le cadre de la conduite, de nombreuses études ont été réalisées pour quantifier l'impact positif ou négatif de la musique sur la conduite. Une méta-analyse [59] de ces études conclut sur les nombreux résultats conflictuels trouvés. En particulier, l'absence d'effets significatifs pour des facteurs comme le tempo de la musique sur le comportement au volant, ou au contraire, que l'écoute de la musique à faible niveau cause un temps de réaction plus grand et donc un risque de collision augmenté [60]. Plus récemment, Navarro et al. [54] identifient dans la littérature trois utilisations de la musique par le conducteur :

- Combattre l'ennui et améliorer l'état émotionnel ou affectif du conducteur.
- Atteindre un niveau suffisant d'activité, en particulier pour les longs trajets.
- Améliorer la sécurité dans le cas de musique décrite comme « facile à écouter » par Navarro.

Ils concluent également sur un effet de la musique, soit directement, soit indirectement par le biais de l'humeur et des émotions, sur la conduite, mais avec une taille d'effet limitée. L'absence de consensus net sur le changement de comportement que la musique crée chez le conducteur ainsi que le manque de méthodologie d'évaluation généralisée explique que cette question ne sera pas traitée dans la suite de ce manuscrit et sort du cadre de l'étude.

### 1.3.3.3 | La fatigue

Une autre question récurrente est celle du risque de fatigue. Il est central et très largement étudié dans les questions de sécurité routière, principalement le risque de somnolence au volant [61, 62]. Par conséquent, l'impact du son a bien été également étudié dans le cadre de l'automobile [63, 64] pour lequel les études concluent à un effet négligeable du niveau sonore sur la somnolence. Ces résultats expliquent le fait qu'aucune mention du son ou du niveau sonore ne soit présente dans les dernières méta-analyses sur le sujet de la somnolence au volant [65, 66].

La fatigue peut cependant être définie différemment et ne pas seulement porter sur la somnolence. La fatigue générale ressentie peut par exemple être mesurée à travers le *Multidimensional Fatigue Inventory* (MFI) [67], mais se prête mal au sujet d'une fatigue induite par un son dans la durée mais plutôt pour une fatigue quotidienne. Ou encore l'évaluation de la fatigue cognitive ressentie par l'ajout d'un son reste alors un sujet ouvert [68].

### 1.3.3.4 | Acoustique augmentée

Parmi les quatre catégories des sons intentionnels présentées plus tôt, la dernière catégorie, interaction sonore continue (*continuous sonic interaction*) a pour objectif de rendre audible une expressivité dans une interaction humain-produit [39]. Ce type de son intentionnel, adapté à son environnement, cohérent avec l'habitacle, pourrait alors être perçu comme agréable par l'utilisateur ou agir sur d'autres aspects de la définition de confort mentionnée plus tôt. Une définition formelle de la notion d'acoustique augmentée, qui est utilisée dans cette thèse est proposée ici comme une interaction sonore continue, conçue autour d'un environnement ou d'un objet sonore précis, ayant pour objectif d'améliorer un jugement perceptif dans l'utilisation de l'objet ou de l'environnement en question.

Il est possible de trouver dans la littérature le terme de *Active Sound Design* (ASD) [69, 70], bien que n'ayant pas de définition aussi formelle que celle proposée ci-dessus, il se rapproche beaucoup de

l'acoustique augmentée telle que définie. Le terme d'acoustique augmentée y est préféré, pour éviter toute confusion avec la notion d'acoustique active qui utilise le contrôle actif tel que le *Road Noise Cancellation* (RNC), et pour se rapprocher des notions de sons intentionnels tels qu'introduit et du design sonore. Ce terme d'acoustique augmentée établit aussi un parallèle avec le concept de réalité augmentée, popularisée par l'industrie du jeu vidéo, mais qui apparaît aussi dans l'industrie automobile [71, 72] à travers les aides à la conduite par exemple. La proximité avec la réalité augmentée réside dans le fait d'ajouter artificiellement des éléments à un environnement réel. Ce chapitre présentera ainsi différentes approches suivies pour guider cet ajout de son.

### 1.3.4 | Design sonore

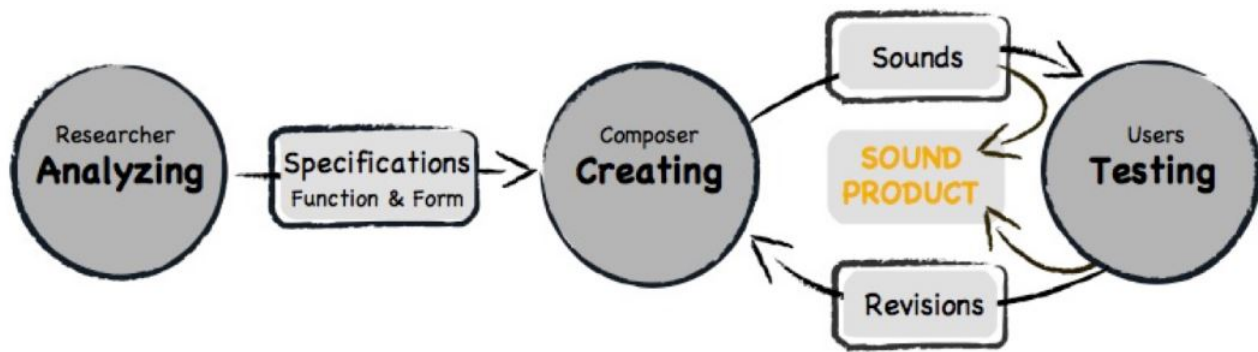
L'acoustique augmentée nécessite donc de créer des sons intentionnels, qui seront joués en continu lors de la conduite. La création de ces sons nécessite donc une tâche de design, plus précisément de design sonore. Une définition du design sonore proposée par Susini et Houix [73] définit le design sonore comme une discipline visant à créer des sons nouveaux, dans l'objectif de faire entendre des intentions dans un contexte d'utilisation. L'intention peut être de différents types : la première intention est d'atteindre efficacement un but grâce au son (fonction), la seconde est de produire une réaction harmonieuse, combinant la satisfaction de l'utilisateur (agrément), la cohérence du son avec le produit (identité) et l'intégration convenable dans son environnement sonore (écologie). Un design sonore efficace doit s'articuler entre ces différentes intentions pour produire des nouvelles interactions grâce au son.

Contrairement à la qualité sonore, le design sonore s'oriente donc vers des intentions sonores. Cette distinction entre la qualité et le design est importante et montre que le design sonore est approprié à notre cas. L'approche explorée par l'acoustique augmentée est originale car elle ne travaille pas sur le son d'un objet précis, mais le son autour. Cela se répercute dans les intentions que le sound design devra traduire. La fonction du son conçu dépendra du cas d'utilisation, son agrément est important, mais c'est l'identité et l'écologie du son conçu qui semblent être les aspects les plus importants du design dans notre cas ; l'acoustique augmentée doit s'intégrer autour de sons déjà existants, ayant leurs propres sens, parfois perçus comme des défauts, parfois comme des informations pour le conducteur.

Historiquement, la paternité du design sonore est attribuée à Luigi Russolo au début du XXe siècle. Certains des instruments qu'il a construits, tels que les *intonarumori*, destinés à synthétiser les « sons-bruits » inspireront le bruitage des premiers films sonores, tels que *La mélodie du monde* (1929) de Ruttman. La proximité entre le design sonore et l'art cinématographique continuera de se développer au fil des années puisque des sons doivent précisément y être créés, avec l'intention d'être entendus, pour illustrer une action spécifique se déroulant à l'écran ; c'est le travail des "*Foley Artist*" [74]. Ce design sonore vise à « augmenter » une œuvre cinématographique en ajoutant des éléments sonores. Lors de la conférence « Concevoir le son pour le développement durable »,<sup>16</sup> Sandra Pauletto présente le travail des *Foley Artist* et y explique que « Cette création est faite plutôt pour être ressentie que réellement écoutée ou entendue. Ces sons sous-entendent un événement, un comportement plutôt que l'exprimer directement. ». Cette philosophie de création guide aussi l'acoustique augmentée, qui vise à créer des sons avec l'intention d'être entendus, mais avec l'idée de ne jamais être entendus seuls, mais dans un contexte. Les sons ajoutés suggèrent une atmosphère ou détournent l'attention d'un défaut.

A la frontière entre le travail d'artiste et celui d'artisan, la dénomination « d'art sonore appliqué » [75] peut être utilisée pour qualifier le design sonore. Puisque celui-ci a pour vocation à travailler sur l'aspect et sur la fonction des sons créés. Le design sonore en d'autres termes, c'est la création

<sup>16</sup> [https://medias.ircam.fr/xa2e51c\\_designing-sound-for-sustainability](https://medias.ircam.fr/xa2e51c_designing-sound-for-sustainability)



**Figure 1.7.** Méthodologie présentée par Susini [76] et utilisé au sein de l'équipe Perception et Design Sonores (PDS) à l'IRCAM

d'un son par un designer pour répondre à une problématique levée par l'ingénieur. La figure 1.7 schématise la méthodologie appliquée à ce processus de design. Les 3 acteurs représentés, le chercheur "Researcher", le compositeur "Composer" et l'utilisateur "User" travaillent alors ensemble. Dans notre cas, la problématique identifiée par le chercheur est celle du confort du conducteur, le son créé aura pour fonction de l'améliorer. Pour cela la fonction et l'aspect du son doivent être discutés, une partie du travail de ce chapitre sera de les spécifier. La boucle de rétroaction traduit l'importance d'une communication efficace entre le compositeur et l'utilisateur du son. Ici c'est surtout l'ingénieur qui est « utilisateur » du design, puisque il identifie la problématique et sera satisfait ou non du résultat, c'est donc pour faciliter cet échange et aider aux spécifications requises que le lexique précédemment construit sera utilisé en y incluant la dimension contextuelle.

### 1.3.5 | Méthode scientifique appliquée et design sonore

Ce travail de recherche incluant le design et la création sonore pose cependant la question de la méthodologie scientifique à adopter. En effet, un domaine aussi riche que la création sonore pourrait sembler trop vague, ou trop arbitraire pour être considéré comme un objet d'étude à part entière. L'approche de la méthodologie scientifique suivie ici s'approche de celle suivie en sociologie et défendue par Bourdieu, puis par Guillaume Lecoindre. « *La profession des scientifiques a pour but collectif de produire des connaissances objectives sur le monde réel* » [77]. Il est donc essentiel dans ce travail sur le design sonore de considérer un objet sonore créé aussi proche du monde réel que possible. Pour cela, le travail de création sonore doit autant que possible être mené comme le ferait un designer sonore, et non comme le ferait un scientifique ou un numéricien, quand bien même il en serait alors plus aisé d'en tirer des conclusions. Sans cela, on se retrouverait non plus à étudier le design sonore, mais la synthèse sonore. De ce fait, dans ce chapitre le processus de synthèse des sons pensé de manière paramétrique sera détaillé, mais en laissant la variable de « créativité » libre. Ainsi nous séparons d'un côté la « quête de créativité » [78] à laquelle est soumis le designer lors de la composition des sons, et la « recherche de la connaissance » du chercheur une fois la matière sonore créée.

### 1.3.6 | Résumé de la section

#### Du confort au design sonore

Cette section permet de préciser la notion de confort, qui ne se perçoit que lors de l'utilisation d'un objet. La distinction est faite avec la notion de luxe, mais aussi avec la notion de qualité. De plus, ce travail de définition du confort appuie sur la différence qu'il existe entre « l'absence d'inconfort » et « le confort ». L'inconfort est principalement causé par des sensations physiques pouvant entraîner de la gêne, voire de la douleur. L'absence d'inconfort résulte en un état neutre, et le confort apparaît alors lorsque l'utilisation de l'objet devient agréable, cohérente avec les attentes de l'utilisateur et son environnement. Un détail de ce modèle, primordial dans cette étude est que l'inconfort pousse l'individu à fixer la cause de celui-ci, tandis qu'une situation neutre ne pousse pas l'individu à améliorer la situation, il se satisfait de l'état neutre. Améliorer le confort nécessite alors de proposer un état supérieur à cet état neutre, que l'utilisateur ne choisirait pas nécessairement spontanément.

Ce travail de définition aboutit à l'adoption d'un modèle de confort issu de la littérature, qui est transcrit à la question du son et du confort sonore. Deux confort sont identifiés : un confort acoustique, centré autour de la réduction des sources sonores gênantes via l'isolation phonique et un confort sonore centré autour de la cohérence des sons avec l'utilisation qui est faite d'un objet. La définition du confort sonore utilisée dans cette thèse est alors proposée : « *Le confort sonore est ce qui contribue à une sensation agréable dans l'utilisation d'un objet, en termes de retour sonore, cohérent avec les attentes de l'utilisateur et l'environnement sonore déjà existant.* »

Améliorer le confort sonore, ce n'est pas seulement réduire les sources existantes, c'est aussi proposer une scène sonore globale, agréable, cohérente avec les attentes et l'utilisation que fait l'utilisateur de l'objet. Une notion très contextuelle dépassant la simple mesure physique. L'impact que cet ajout de sons pourrait avoir sur la fatigue, les humeurs ou les émotions ne sera cependant pas traité dans cette thèse. Cependant, l'amélioration du confort sonore est tout de même explorée à travers l'ajout de son. Cet ajout de son doit être distingué de la sonification, et s'inclure plutôt dans une démarche de design sonore. Cette démarche de design sonore se caractérise par un aspect donné aux sons, et différentes intentions liées à ceux-ci. Ici nous appelons acoustique augmentée l'application de ce design sonore à un environnement sonore précis, pour lequel le son ajouté a pour vocation d' « augmenter » les sources sonores déjà présentes. Dans notre cas, cette augmentation se fait avec l'intention d'améliorer le confort, c'est-à-dire d'améliorer l'agrément perçu de la scène, la cohérence du son avec son environnement et les attentes de l'utilisateur.

## 1.4 | Des paysages sonores à la description verbale

### 1.4.1 | Paysages sonore : définition

La définition de confort sonore utilisée ici s'appuie sur de nombreux éléments contextuels concernant notamment l'environnement, l'individu et sur l'utilisation qu'il fait d'un objet. Le domaine d'étude des paysages sonores offre des outils théoriques et pratiques pour traiter ces éléments. La notion de paysage sonore, ou *soundscape*, apparaît pour la première fois avec les travaux de Murray Schafer [79] en 1977 qui théorise et définit cette notion de paysage sonore comme : « L'environnement des sons. Techniquement, toute partie de cet environnement prise comme champ d'étude. ». Cette définition a depuis évolué, mais traduit déjà l'idée de ne pas considérer le son seul comme objet d'étude, mais aussi les autres sons environnants. Depuis cette étude pionnière présentée par Schafer, le sujet des paysages sonores est devenu un sujet d'étude à part entière et on retrouve aujourd'hui la définition standardisée [80] suivante de paysage sonore : « Environnement sonore tel que perçu, vécu et/ou compris dans son contexte par un ou des individus ». On remarque que, dans la définition tel que partagée aujourd'hui dans les standards, l'individu est remis au centre de la notion de paysage sonore. Ce sujet étant par nature assez ouvert, la définition exacte du paysage sonore reste encore sujette à débat [81].

### 1.4.2 | L'habitacle du véhicule comme paysage sonore

La définition standardisée du paysage sonore depuis maintenant plusieurs années partage de nombreux points communs avec le concept de confort mentionné plus tôt. En effet la littérature scientifique sur le confort s'accorde sur l'aspect très contextuel de cette notion et donc sur la nécessité de traiter les sons dans leur contexte, tels qu'ils sont perçus ou compris par un utilisateur. Cet aspect est appuyé par la définition du confort sonore dans le cadre de cette étude, qui souligne l'importance d'intégrer des sons dans leur environnement. Cette approche implique de ne pas étudier les sons isolément, mais en tenant compte de leur intégration dans leur environnement et leur contexte. Par conséquent, l'habitacle est considéré et traité comme un paysage sonore [22]. Ce qui nous amène ici à considérer les outils issus de ce champ d'étude pour traiter entièrement cette notion de confort avec les contraintes liées au contexte.

En pratique, les études sur les paysages sonores portent majoritairement sur des espaces, tels que des gares [82] ou des villes [83] avec de nombreux événements, plutôt que sur des objets comme la voiture. L'application de la notion de paysage sonore au véhicule électrique reste donc assez limitée et mériterait d'être discutée. De ce fait on gardera dans la suite la dénomination d'environnement sonore plutôt que de paysage sonore. Cependant les outils ou méthodes utilisés pour les études des paysages sonores seront tout de même utilisés dans notre cadre.

### 1.4.3 | Caractériser un paysage sonore

Par définition, la perception qui est faite d'un paysage sonore dépend de l'individu. De plus, les sources ou événements sonores peuvent varier fortement d'un moment à l'autre et ne pas être entendu ou perçu de la même manière par chacun, selon l'utilisation ou l'interprétation qu'ils font d'un lieu ou d'un son. La difficulté avec les paysages sonore est alors d'obtenir un corpus de sons représentatifs du paysages sonore puisqu'ils peuvent être extrêmement variés.

Les études des paysages sonores se tournent alors plutôt vers une approche de description spontanée et libre de la part des individus interrogés [84, 85]. Ces descriptions, orales ou écrites sont alors fondées sur la mémoire que le participant a d'une scène qu'il connaît. Cela permet de traiter, sans hypothèse,

l'ensemble des sources sonores au travers de ce qui a été perçu ou retenu. Pour stimuler ou aider à cette tâche de restitution, les échanges sont alors guidés par des questionnaires [86]. En particulier, l'utilisation fréquente de questionnaires semi-directifs dans ces études est relevée [37, 82, 83, 86, 87] permettant la libre verbalisation autour d'un sujet de la part du participant, en resserrant progressivement sur des détails plus précis de l'environnement.

Mzali [37] propose une approche psycholinguistique et cognitive du sentiment de confort dans sa thèse sur le confort acoustique dans les trains. Deux processus sont identifiés : un ascendant (*bottom/up*) allant des récepteurs sensoriels au cerveau, et un descendant (*top/down*) où une sélection des informations perçues est réalisée en fonction des connaissances de l'individu sur son environnement. La « mémoire préalable » d'un environnement influe sur sa perception en conditionnant justement les informations sélectionnées. Mzali propose donc de s'intéresser à ces processus mentaux et à cette mémoire préalable, sur la notion de confort, en questionnant l'individu. Ces représentations mémorisées par les participants sont alors inférées sur la base d'une analyse lexicale des échanges écrits ou oraux, guidés par des questionnaires. On retrouve ainsi, dans cette étude comme dans les études sur les paysages sonores, un accent important sur l'analyse sémantique des informations recueillies auprès des participants [88].

Ce processus de questionnement, présenté par Mzali, est alors mené en deux étapes. La première étape est un questionnement ouvert exploratoire, sous la forme d'un questionnaire semi-directif, visant à identifier les thèmes principaux liés au confort, sans avoir orienté l'individu. Cette première étape est donc particulièrement délicate puisqu'elle nécessite d'avoir des participants qui répondent de manière la plus exhaustive et détaillée aux questions, tout en restant mobilisés jusqu'aux dernières questions. Comme ce type de questionnaire est soumis à une analyse verbale, il est alors compliqué d'avoir un grand nombre de participants. Il est alors possible d'obtenir les réponses à ces questionnaires de plusieurs manières. Mzali choisit de faire remplir ces questionnaires à l'écrit, lors de trajets en train réalisés par les participants, puisque le sujet se portait sur le confort acoustique dans les trains. Dans le cas où le panel de participants est beaucoup plus large ou dispersé, le questionnaire peut être soumis par mail [83] ou en ligne [89], ce qui facilite sa diffusion mais permet moins de s'assurer que les participants remplissent sérieusement le questionnaire. Une dernière modalité utilisée pour remplir ce type de questionnaire est d'avoir recours à des interviews [82, 87], ce qui permet de s'assurer de l'investissement du participant tout au long de l'échange, mais pose la question du biais potentiellement induit par l'intervieweur, qui doit être réduit au minimum.

Une fois le questionnement exploratoire effectué, la deuxième étape consiste à mener des questionnements fermés à plus grande échelle, basés sur des hypothèses extraites de l'analyse verbale du premier questionnement, que l'on peut ensuite traiter à l'aide d'une analyse statistique classique. Ces deux étapes identifiées par Myriam Mzali, guident alors le plan général de cette thèse. Cette approche sans hypothèse est particulièrement adaptée dans notre cas, puisque la rupture technologique apportée par la popularisation du véhicule électrique évoquée plus tôt change complètement l'environnement sonore sur lequel se focalise ce travail de thèse.

#### 1.4.4 | Historique de la description du son

Les outils d'exploration de paysage sonores nécessitent d'être capable de décrire verbalement une scène sonore. La description verbale du son apparait historiquement plutôt pour ceux d'origines musicales pour lesquels un vocabulaire détaillé a été construit au cours des années dans la musique occidentale. On retrouve ainsi l'utilisation des termes faisant références aux rapport harmoniques, tels que la quinte, la tierce ou l'octave dans les écrits antiques étudiant l'harmonie et la hauteur des sons. Au XIe siècle

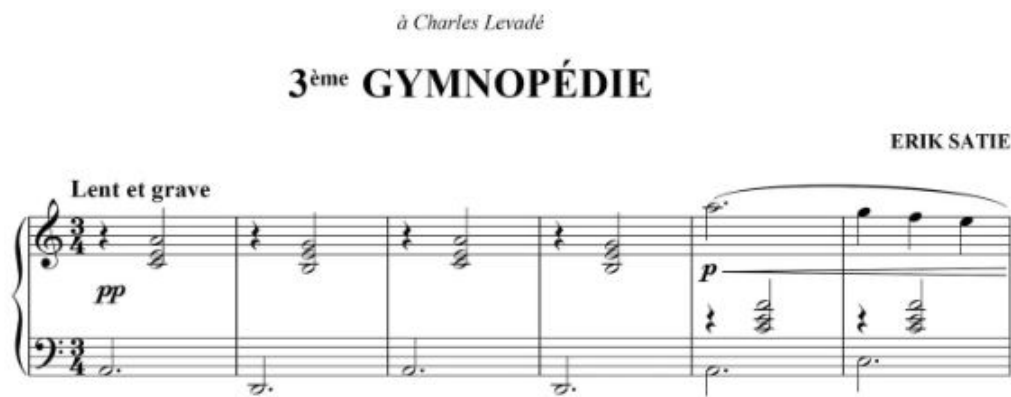


Figure 1.8. Ouverture de la 3<sup>ème</sup> Gymnopédie d'Erik Satie et son annotation « Lent et grave ».

sont proposés les noms pour six notes de différentes hauteurs d'une gamme couramment utilisée, Ut Ré Mi Fa Sol La, proposés par Guido d'Arezzo. Des noms qui évolueront, le Ut devenant un Do, plus facilement chanté, et un Si est ajouté, pour donner le fameux Do Ré Mi Fa So La Si. Plus tard, l'annotation des partitions se répand avec l'utilisation de termes, comme *crescendo*, *piano*, *pizzicato*... précisant le style de jeu à adopter pour donner à un son ou à un instrument la sonorité appropriée. Des annotations plutôt objectives au départ, évoluent au fur et à mesure, et sont utilisées pour apporter à l'œuvre une dimension plus subjective, ne décrivant plus seulement le son mais son interprétation. Un exemple marquant de cette évolution des annotations et des mots utilisés en musique est probablement le compositeur Erik Satie, qui dans sa 3<sup>ème</sup> gymnopédie, annote d'un « lent et grave » (Fig. 1.8) ou d'un « Très luisant » dans sa gnossienne numéro 1.

En parallèle de cette évolution de la musique dite moderne, à la fin du XIX<sup>e</sup> siècle, le paysage sonore change fortement. La révolution industrielle est en marche et les bruits mécaniques et industriels envahissent les villes, comme le décrit Murray Schafer. C'est dans ce contexte que tous ces bruits commencent à être au centre de l'attention. Luigi Russolo (Fig. 1.9) considéré comme le père de la musique bruitiste, dans *l'Art des bruits*, propose une classification de ces sons non musicaux, qu'il appelle « son-bruit », comme instruments possibles pour un orchestre du futur :

- Grondements, éclat, bruit d'eau tombante, bruits de plongeon, mugissements.
- Sifflements, ronflements, renâclements.
- Murmures, marmonnements, bruissements, grommellements, grognements, glouglous.
- Stridences, craquements, bourdonnements, cliquetis, piétinements.
- Bruits de percussions (obtenus en frappant diverses matières : métal, bois, peaux, pierres, etc.).
- Voix d'hommes et d'animaux (cris, gémissements, rires, sanglots etc.).

### 1.4.5 | Différents types d'écoute

Les « son-bruits » commencent alors à être écoutés et décrits, comme l'étaient jusqu'alors les pièces musicales et les instruments. La musique concrète utilise des sons séparés de leurs contextes de leurs sources, notamment des sons du quotidien enregistrés, comme matière première pour la composition musicale. Pierre Schaeffer, probablement influencé par les travaux de Russolo [90] publie son *Traité des objets musicaux* [91]. Il y introduit la notion « d'objet sonore » qui renvoie au son pour lui-même, et non à la source ; associé à la notion d'écoute réduite. Cette conception des sons se prête particulièrement bien à la musique concrète dont il est considéré comme le précurseur. Les termes « bruit » et « son



Figure 1.9. Dynamisme d'une automobile, 1911, Centre Pompidou, Luigi Russolo

musical » sont alors abandonnés, au profit du seul terme « son ». La frontière entre la musique et les sons de tous les jours s'estompe progressivement. Ainsi on privilégiera dans ce manuscrit le terme de son plutôt que bruit pour parler d'un signal acoustique. Associé à ce mode d'écoute réduit, Schaeffer propose une première taxonomie des sons basés sur deux éléments principaux qu'il nomme « facture » et « masse ».

Gaver, dans *The Sounding object* [92] distingue deux modes d'écoute des sons. Le premier "*musical listening*" se rapproche grandement de l'écoute réduite décrite par Schaeffer correspondant à l'écoute que l'on fait d'un son en lui même. Le deuxième mode d'écoute identifié par Gaver est "*everyday listening*" qui consiste à écouter un son comme un événement associé à une source, une action ou un contexte. La différence dans ces modes d'écoute repose sur l'expérience qu'a le sujet d'un son. Un même son peut être écouté de manière réduite, pour reprendre le terme de Schaeffer dans une pièce musicale, ou de manière "*everyday listening*" dans un autre contexte. Michel Chion [93] ajoute à ces deux modes d'écoute un troisième, basé sur l'information sémantique, ou contextuel autour de ces sons. Ce dernier mode d'écoute consiste à écouter un son pour ce qu'il véhicule ou ce qu'il signifie.

#### 1.4.6 | Type de discours

En dépit des travaux de Schaeffer qui proposait une taxonomie pour décrire les sons, et notamment les « objets sonores », qui avait alors surtout une place dans le monde musical, la description du son reste un point délicat. Dans le cas d'interlocuteurs novices sur la question du son, les habitudes poussent plutôt à décrire l'action à l'origine d'un son que le son lui-même. Les premières études de catégorisation libre [94] et de verbalisation libre [92] menées montrent que les gens identifient spontanément le son en termes d'objets et d'événements, et se tournent vers des propriétés acoustiques uniquement lorsque la source n'est pas identifiée. Plus récemment, Lemaitre et al. [95] dans une expérience de catégorisation, montrent que les experts ont plutôt tendance à regrouper les sons selon les ressemblances acoustiques, et les non-experts selon la proximité des causes ou des événements créant le son. Maxime Carron [96], étend ces trois catégories d'écoutes, à des catégories de discours, utilisées pour décrire les sons :

— **Discours causal** : La description causale d'un son se réfère à l'origine de celui-ci, soit par l'action qui a créé le son « *Un claquement* » soit par l'objet à son origine « *une porte qui se ferme* ». Cette

description est associée au mode d'écoute "*everyday listening*" introduit par Gaver [92]. Des sons très similaires entre eux peuvent alors avoir une description causale très différente si l'origine de ces sons est différente : « *un coup de feu* », « *un pétard* », « *l'explosion d'une fusée d'artifice* ». Ce type de description nécessite alors de connaître le phénomène précis pour identifier le son, et serait difficilement utilisable pour décrire des sons abstraits.

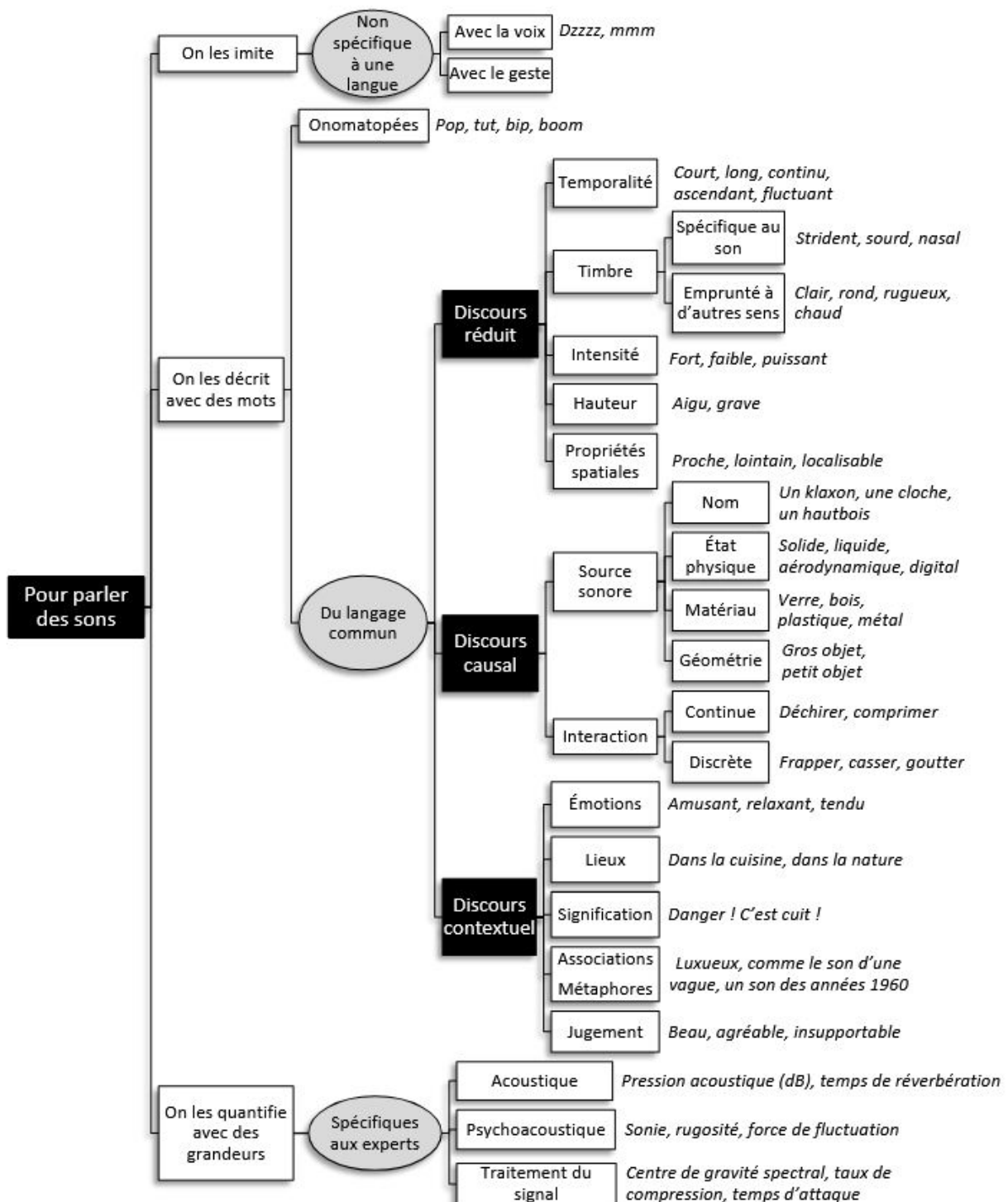
- **Discours réduit** : La description réduite des sons apparaît en premier dans le travail de Schaeffer *Traité des objets musicaux*, qui le premier s'intéresse à ce mode d'écoute et de description et propose sa propre taxonomie des sons musicaux. Cette modalité décrit plutôt les propriétés du son indépendamment de l'objet qui les a créés. À l'exception d'adjectifs faisant parti du langage courant tels que « *fort/faible* » ou « *grave/aigu* », cette modalité est probablement la moins intuitive à utiliser par une population non spécialiste ou habituée. On retrouve en effet un vocabulaire plus développé chez des populations d'experts, travaillant couramment avec le son comme des musiciens [97]...
- **Discours contextuel** : Cette modalité est dédiée à l'intention qu'il y a dans un son [93], et se prête bien aux sons composés pour une œuvre musicale, mais aussi à tous les autres sons d'origine humaine : « agacé par mon refus de lui répondre, il frappa une seconde fois à la porte » / « Inquiet de mon absence de réponse, il frappa à la porte une seconde fois ». Dans ce cas, les adjectifs ou descripteurs ne se réfèrent pas au son, mais traduisent bien une sonorité différente. Dans cette modalité de description, peuvent être identifiés des descripteurs hédoniques, liés aux émotions ou jugements qu'un son peut évoquer tels que « un sifflement agaçant ».

#### 1.4.7 | Construction de lexiques et paradigme SpeaK

Ces travaux mènent au constat qu'il n'existe pas de termes universellement partagés pour la description du son, outre les termes liés aux attributs de niveau sonore, durée et de hauteur [98]. De ce fait, une description du son ne devient réellement utile que dans le cas d'un usage régulier ou d'une formation comme on peut le trouver chez les professionnels. On trouve ainsi assez tôt l'utilisation de la verbalisation chez les sonaristes [99]. Différentes études sur les descripteurs verbaux du timbre attestent cependant de la richesse du vocabulaire existant spontanément pour décrire le timbre du piano [97], de la guitare [87] ou du violon [100, 101]. Dans le cas de sons non musicaux, Anne Faure [102] extrait ainsi une liste de descripteurs utilisés pour décrire le timbre.

Malgré la richesse du vocabulaire à disposition, de nombreux problèmes de communication existent. Langeveld [39] met en évidence la difficulté existante pour la communication entre les designers sonores et les utilisateurs de leurs produits. Pour deux interlocuteurs dans le même domaine d'expertise, tels que des ingénieurs du son, la problématique existe toujours, même si la culture commune permet de faciliter les échanges aux travers d'exemples connus, servant de références pendant les discussions [103]. L'idée apparaît ainsi d'agréger sous forme de lexiques les descripteurs identifiés [102] comme utiles à un domaine pour favoriser leur efficacité. Il n'est cependant pas suffisant de partager le terme pour que le sens de celui-ci soit partagé, Rosi et al.[104] montrent la différence entre plusieurs populations sur l'utilisation d'attributs pour décrire le son.

La méthodologie de construction de lexiques SpeaK [105, 106] permet alors de répondre à ces différents points. Cette construction de lexique est inspirée des lexiques sensoriels déjà utilisés [107, 108] dans d'autres champs sensoriels comme le milieu olfactif ou gustatif. Ces lexiques sont construits avec un panel de spécialistes et nécessitent alors une importante formation [108] pour que chaque terme présenté soit entièrement partagé et compris au sein du panel. Dans le cadre de la méthodologie SpeaK, le lexique n'est pas destiné à un panel d'experts formé à maîtriser chaque terme. Celui-ci est destiné à



**Figure 1.10.** Types de discours existants pour décrire le son, tels que détaillés par Maxime Carron dans sa thèse [96], ainsi que des exemples précis.

faciliter les échanges entre interlocuteurs novices ou ne partageant pas les mêmes habitudes d'écoute. De ce fait ce type de lexique est construit pour être accessible, et il est composé de :

- Une liste de termes.
- Une définition la plus claire possible pour chacun des termes.
- Un ensemble d'exemples sonores pour chaque terme.

Ainsi une base d'échange et description peut-être construite autour du son. Les lexiques SpeaK construits peuvent se baser sur une littérature déjà existante [106]. La contribution originale de Carron, en travaillant à partir de la littérature scientifique et sans corpus de sons, est ainsi de développer une approche "*Words to Sounds*" plutôt que celle "*Sounds to Words*" habituellement utilisés dans les études travaillant avec un corpus de son. Notre travail s'inscrit dans le même paradigme "*Words to Sounds*". La spécificité ici réside en revanche dans le manque de littérature autour du son du véhicule électrique et du manque de culture et d'habitude autour de cet univers.

### 1.4.8 | Résumé de la section

#### Des paysages sonores à la description verbale

À l'issue de la section sur le confort sonore, la dimension contextuelle et subjective du confort est mise en avant, elle favorise l'idée de traiter les sons dans leur contexte et leur environnement, et non comme des sons isolés. Cette approche, dite écologique, au sens étymologique du terme, est notamment à l'origine de la notion de paysage sonore introduite par Schafer. Cette approche consiste, à l'origine, à s'intéresser à l'évolution des sons présents dans un environnement transformé par les révolutions industrielles et booms économiques. La définition actuellement standardisée de paysage sonore : « environnement sonore tel que perçu, vécu et/ou compris dans son contexte par un ou des individus » se prête alors à des environnements aussi vastes que des villes transformées par des décennies d'industrialisation, des parcs, des gares ou pourquoi pas, à l'habitacle de voiture.

Ce champ d'étude a alors mis au point une série d'outils permettant de caractériser ces environnements avec leur dimension contextuelle et individuelle. Ces outils sont principalement construits autour de la description verbale d'un environnement à partir des souvenirs ou des perceptions qu'en ont les individus, permettant d'accéder à la représentation qui en est faite. Un outil comme le questionnaire semi-directif est alors couramment utilisé dans ce domaine et semble également adapté à notre cas. L'intérêt de ce type d'outil est de pouvoir travailler avec peu d'hypothèses sur un domaine et ne pas avoir besoin de se restreindre à un cas d'utilisation précis, ou une série d'enregistrements qui pourrait alors omettre un aspect important de l'environnement sonore.

Dans le cadre du son, la description verbale a été historiquement plutôt utilisée en musique. C'est avec l'apparition du mouvement Bruitiste, puis de la musique concrète, qui traite les « bruits » comme des « sons » et les transforme en des objets musicaux, qu'apparaît une formalisation autour de la description de sons non-musicaux. Trois types d'écoutes différentes sont identifiés, autour desquels sont associés trois types de discours, qui sont repris dans cette thèse.

La description causale, associée au mode d'écoute de tous les jours qui décrit un son par sa cause, l'évènement à son origine, ou l'objet le produisant. La description réduite, associée au mode d'écoute réduit ou « l'écoute musicale », qui consiste à écouter les sons comme des « objets sonores » tels qu'appelés par Pierre Schaeffer, c'est-à-dire un son *per se*, détaché de son origine ou son contexte, où seules les propriétés acoustiques sont importantes. Enfin le dernier type de description, contextuelle, associée au type d'écoute appelée écoute contextuelle ou sémantique, consiste à décrire le sens ou l'interprétation qui est fait d'un son.

Ces trois modes de description associés à trois modes d'écoute, ne sont cependant pas utilisés de la même manière selon l'expertise en son. De plus, même si des habitudes existent avec un vocabulaire riche au sein d'une population d'experts, la communication avec un autre groupe avec des habitudes différentes n'est pas évidente. Le paradigme SpeakK, qui consiste à développer un lexique constitué d'une liste de termes, de définitions et d'exemples sonores, destinés à des populations non expertes est alors adopté dans cette thèse. Les trois modes de descriptions identifiés serviront alors à construire les définitions des termes.

## 1.5 | Objectif

À l'issue de ce travail bibliographique permettant de définir et introduire les différentes notions clés à ce sujet, l'objectif de cette thèse vise à répondre à la problématique suivante : Peut-on améliorer le confort sonore de l'utilisateur du véhicule électrique à l'aide de l'acoustique augmentée ? Trois composantes se dessinent alors autour de cette problématique, autour de laquelle seront construits les chapitres de ce manuscrit.

La première composante de cette thèse portera sur la description des sons dans l'habitacle du véhicule électrique. Elle est abordée du point de vue de la verbalisation qui apparaît comme le moyen le plus adapté pour inclure la dimension contextuelle, indissociable du confort. La description verbale du son est étudiée à l'aide des outils développés dans le champ des paysages sonores comme le questionnaire semi-directif et l'analyse sémantique. Ces étapes ont pour but de converger vers la création d'un lexique descriptif des sons dans cet environnement, selon l'approche déjà proposée avec *SpeaK*. L'originalité de la méthode proposée repose sur le fait de baser cette construction du lexique sur des descriptions recueillies par ce questionnaire semi-directif.

La deuxième composante de cette thèse est orientée vers la création de sons, nécessaire au concept d'acoustique augmentée qui est défini précédemment. Ce chapitre a alors pour but de détailler les solutions concrètes explorées, en ajoutant du son, pour améliorer le confort dans l'habitacle. Ces solutions se présentent sous la forme de trois stratégies qui sont présentées, inspirées de résultats scientifiques ou de propositions techniques déjà existantes. Cette création de son se place dans un contexte précis, qui est fixé à partir des travaux de descriptions présentés dans la première composante de cette thèse.

Enfin, la troisième composante consiste à évaluer perceptivement les sons qui ont été composés pour une situation sonore précise. Les tests construits sont réalisés au casque dans un environnement contrôlé, ce qui permet de tester une grande variété de sons composés, de paramètres et de conditions expérimentales, au prix d'une écologie réduite. Ce test nous permet alors de confirmer l'efficacité d'une méthodologie alternative, *Best-Worst Scaling* (BWS), récemment utilisée dans des études similaires, par rapport à une approche plus classique de type Différentiel Sémantique (DS). Cela nous permet de confirmer la pertinence des stratégies identifiées précédemment. À l'issue de ces résultats une perspective est présentée : la possibilité de faire une expérience in situ pour confronter les résultats obtenus au cours de l'étude à une situation contextuelle, plus proche de la réalité. Enfin, cette composante se termine sur un chapitre présentant un travail en cours qui porte sur l'amélioration de la méthodologie BWS. Une proposition est apportée et étudiée pour combler les limites mises en évidence au cours de cette partie, sur la difficulté d'analyser une expérience paramétrique avec cette méthodologie. À l'issue de ces parties la conclusion générale est présentée. Les annexes permettent ensuite de fournir plus de détails sur les résultats présentés dans le manuscrit.



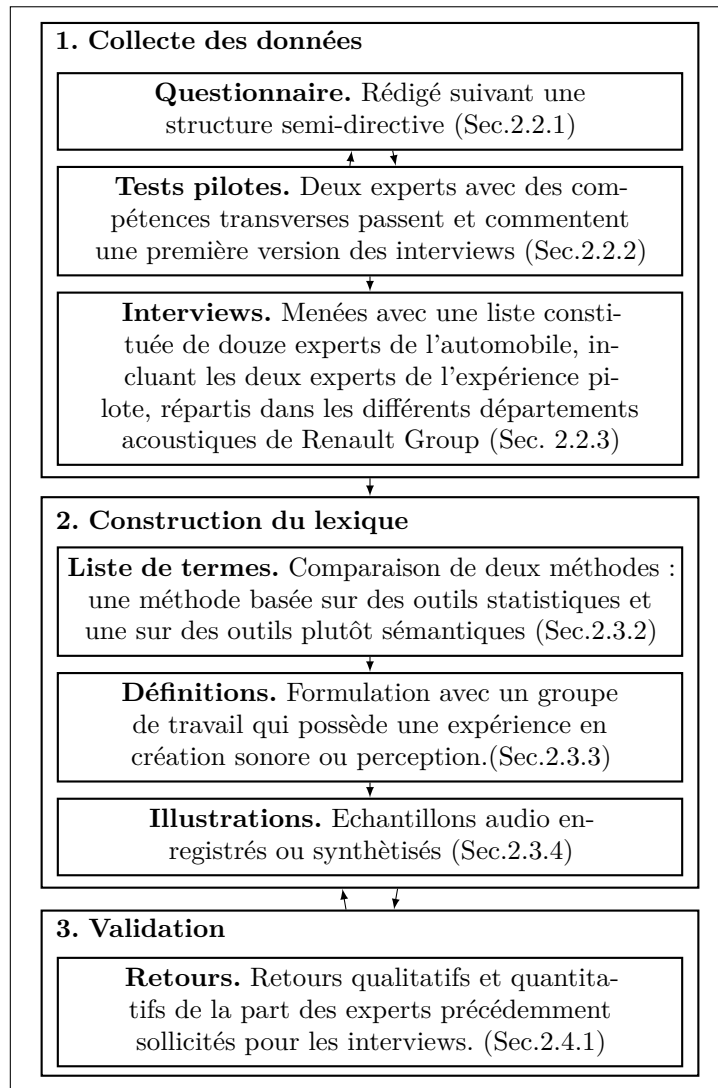
## 2. Description d'un nouvel environnement sonore

*« Un dernier mot [...] je n'en oublie pas pour autant que l'oreille n'est autre qu'un de nos récepteurs de sensations. L'heure est venue de quitter le laboratoire pour la vie qui nous entoure. C'est ce que nous avons fait dans nos études du paysage sonore. » [79]*

### 2.1 | Introduction de chapitre

Ce chapitre traite de la description verbale de l'environnement sonore de l'habitacle du véhicule électrique. Cet objectif est motivé par l'évolution de cet environnement, et en particulier des véhicules silencieux comme l'est le véhicule électrique. En effet les grandes familles de sons existantes historiquement à l'intérieur de la voiture, à savoir les bruits d'origine aérodynamique, le bruit de roulement ainsi que le bruit causé par la motorisation [20], ont fait l'objet de recherches et de développements pour réduire leurs niveaux. Cette tendance atteint son paroxysme avec la rupture provoquée par la généralisation de la motorisation électrique, réputée silencieuse dans l'automobile. La littérature scientifique présentée plus tôt fait alors état d'un réel changement l'environnement sonore de l'habitacle du véhicule. La disparition d'objets sonores au fort pouvoir masquant tel que le moteur à combustion interne laisse apparaître de nouvelles sources sonores telles que le moteur électrique, encore inhabituel pour le conducteur et change alors les attentes et les habitudes d'écoute.

Dans l'objectif d'améliorer le confort sonore, cette évolution rapide de l'environnement sonore amène la problématique qui guidera ce chapitre : comment mieux comprendre les attentes autour du confort sonore de l'habitacle du véhicule électrique ? Pour cela, l'importance de travailler sur la perception des sons ainsi que sur les éléments contextuels qui y sont liés a été montrée. Le cas particulier de l'habitacle du véhicule électrique, que l'on traite comme un paysage sonore conformément à ce qui a été présenté dans le chapitre précédent (Chap. 1), nous encourage alors à utiliser les outils d'exploration de ce domaine qui se basent sur une description verbale du son et de son environnement. Nous verrons dans un premier temps comment sont recueillis les éléments de description de cet environnement sonore au travers d'interviews guidées par un questionnaire semi-directif. Ces interviews nous permettent d'obtenir des descriptions du son, ainsi que des détails autour de l'utilisation de ces termes, et du contexte dans lequel ils sont utilisés. Ces descriptions seront alors utilisées dans un second temps pour construire un lexique autour des trois types de discours présentés dans l'état de l'art : causal, réduit, contextuel. Enfin, la troisième partie de ce chapitre est tournée vers la validation du lexique construit et la mise en ligne de celui-ci. Les différentes parties de ce chapitre sont résumées à la figure 2.1. Ce chapitre détaille donc la méthode originale suivie pour permettre l'élaboration de ce lexique associé à un environnement sonore nouveau, à partir d'interviews avec des experts, ici des acousticiens du secteur automobile.



**Figure 2.1.** Processus chronologique pour permettre la construction du lexique, divisé en trois étapes successives : la collecte des données, la construction du lexique et la validation. Chaque étape est détaillée dans la section référencée.

## 2.2 | Collecte des termes

### 2.2.1 | Rédaction du questionnaire semi-directif

Pour explorer l'environnement sonore de l'habitacle, conformément à ce qui a été présenté (Sec. 1.4) dans l'état de l'art, un questionnaire semi-directif est rédigé. Celui-ci est conçu à destination des experts de l'automobile qui ont déjà une expérience du véhicule électrique et de son acoustique. Le choix de ce type d'outil implique une analyse sémantique, ou au moins statistique, de la verbalisation qui en découlera, favorisée par ce questionnaire autour du sujet du son. En particulier la première partie du questionnaire (Tab. 2.1) est dédiée à la description des différentes sources sonores connues, sur lesquelles la personne interviewée a une expertise. La seconde partie du questionnaire est dédiée à la question du confort sonore et sera abordée dans le chapitre suivant (Chap. 3). Les consignes appliquées pour sa rédaction sont synthétisées ici :

- Le nombre total de questions doit être limité. Il est conseillé de ne pas dépasser un maximum de 10 questions, l'idéal étant 5 ou 6. Dans notre cas, celui-ci est séparé en deux parties différentes de 6 et 4 questions, traitant respectivement de la description du son et du confort.
- Les questions doivent avoir le moins d'ambiguïtés possible. Il est important de s'assurer que les

	Questions
Q1	Selon vous, quelle est la place de l'écoute dans votre métier ?
Q2	Pouvez-vous lister les sons sur lesquels vous avez travaillé, par exemple, au cours de la dernière année ?
Q3	Parmi les sons que vous nous avez cités, pouvez-vous en choisir trois que vous maîtrisez le mieux ?
Q4	Pouvez-vous dans un premier temps décrire l'origine technique de ces sons ?
Q5	Pouvez-vous maintenant décrire le son en lui-même, c'est à dire sans faire référence à la cause du son ?
Q6	Suite à nos échanges, voyez-vous maintenant des termes supplémentaires que vous n'avez pas cités précédemment ?

**Table 2.1.** Première partie du questionnaire constituée de six questions relatives à la description des sons de l'habitacle du véhicule.

termes techniques utilisés dans les questions sont connus et compris par le participant ou que les questions soient claires.

- Les questions doivent être les plus courtes possible, sans pour autant sacrifier la qualité. Il est recommandé d'éviter des questions en plusieurs parties avec « et/ou », par exemple.
- Il est recommandé de donner certaines contraintes dans la verbalisation, par exemple en demandant de citer tous les événements sur une période, avec des formulations du type « Pouvez-vous lister les sons sur lesquels vous avez travaillé, par exemple, au cours de la dernière année ? ». Cette contrainte n'est pas rigide et peut-être outrepassée mais aide à cadrer la question.

### 2.2.2 | Modalité orale : interviews

Pour obtenir les réponses à ces questionnaires, la modalité choisie est celle d'interviews [82, 87, 104, 109, 110] d'une durée fixée à une heure. Ce format possède plusieurs avantages, le premier étant d'ordre pratique car cela permet de manifester l'importance de l'échange auprès des participants parfois très sollicités. Contraindre le temps de l'interview garantit aussi un équilibre dans le poids des réponses de chaque participant. Cet équilibre est conditionné par le travail primordial de l'intervieweur qui doit être un acteur de cette verbalisation, sans biaiser les réponses. Le rôle de l'intervieweur détaillé dans les travaux de Vermersch [109] est primordial au bon déroulement des interviews et à la qualité des résultats. La dimension « relationnelle », qui est présentée par Vermersch, entre l'intervieweur et l'interviewé est construite en plusieurs étapes. D'abord en amont de l'interview, par une prise de contact si possible en personne ou au moins au téléphone avec le participant. Cet échange introduit le sujet de la recherche et le but du questionnaire et permet de commencer à construire la dimension relationnelle. D'autre part cette dimension est aussi construite au début de l'interview avant de commencer le questionnaire présenté plus tôt, en demandant des détails sur la carrière de l'interviewé, sa relation à la musique ou au son dans son travail, ses habitudes lors de la conduite, etc. Cette relation permet alors de mener plus facilement les interviews, en particulier dans notre cas où l'interview est menée avec des experts d'un domaine, ce qui peut créer un déséquilibre entre l'expert et l'intervieweur [104, 111].

Une fois cette dimension relationnelle construite, l'interview peut commencer. Le rôle de l'intervieweur sera alors de maintenir, encourager ou favoriser la verbalisation au maximum à travers l'utilisation de relances. Lors de ces échanges, l'intervieweur peut avoir la position d'un interlocuteur naïf [87], ou la position d'un co-expert [104, 111]. La position de l'interlocuteur naïf est privilégiée ici pour permettre de maintenir la discussion autour d'éléments simples. Les relances les plus classiques utilisées

Domaine	Experts
Roulement	CT JT PB
Aéroacoustique	PS SS
Motorisation	LP ER QM
Bruits parasites	SR
Ventilation	OH
Global	MD TA

**Table 2.2.** Tableau récapitulatif des domaines d'expertise des experts interviewés

ici, consistent alors à demander systématiquement de redéfinir ou détailler des termes techniques ou acronymes utilisés. Une autre stratégie de relance consiste à favoriser la verbalisation autour du sujet de la description du son en demandant des synonymes ou antonymes usuels. Ce type de relances permet de garder une conversation dynamique sans risquer d'orienter le participant dans le choix d'un mot ou d'un autre. Un grand soin est aussi porté à la prise de notes pendant l'interview. Ceci permet de relancer le participant sur un élément qu'il a évoqué plus tôt, sans avoir à l'interrompre. Cela permet aussi de s'assurer de reprendre les mêmes éléments de langage. Un exemple transcrit de l'échange avec un des experts du roulement traduit cette stratégie. La prise de notes permet ici de se référer à un son au travers des éléments techniques utilisés par le participant pour le décrire, comme illustré dans l'exemple reporté ci-après.

**Intervieweur :**

« Okay, et est-ce que, par exemple, si on change un des paramètres que tu m'as cités tout à l'heure, imaginons que voilà par exemple, on passe d'une jante tôle à une jante aluminium. Tu estimerais que quel serait l'impact sur ce son là ? »

**CT :**

« Ça, ça augmente son niveau à partir de 100 Hz dans cette bande-là. Tout ce qui est au-dessus de 100Hz, c'est légèrement augmenté par le changement de la jante. »

### 2.2.3 | Réalisation des interviews

Ce questionnaire a été rédigé à destination d'experts de l'acoustique automobile, il est donc important de constituer un panel de participants représentatifs au sein du groupe Renault tout en gardant un nombre limité pour ne pas rendre l'analyse trop longue. Pour cela une liste de personnes d'intérêts a été dressée et représentée au tableau 2.2, avec l'aide de Louis-Ferdinand Pardo, co-encadrant industriel de cette thèse. Les experts sélectionnés font partie de la branche « métier » du constructeur, ils travaillent donc sur le développement des véhicules et sont plus éloignés du client mais plus proches de la technique. Les experts sélectionnés couvrent plusieurs domaines, certains travaillent sur le roulement, d'autres l'aéroacoustique, la motorisation, les bruits parasites ou la ventilation. A cela s'ajoute deux experts aux compétences transverses, ayant une connaissance générale de ces sujets. Les domaines cités ici le sont à titre indicatifs, les dénominations internes peuvent varier, de plus, la plupart des experts interviewés ont pu changer de département au cours de leur carrière et ne restreignent pas nécessairement la discussion à leur domaine d'activité au moment de l'interview. Parmi les experts sollicités, ceux ayant répondu positivement se répartissent comme indiqué au tableau 2.2.

La première version du questionnaire est éprouvée lors d'interviews pilotes réalisées avec les experts ayant une connaissance globale des problématiques liées à l'acoustique. Il en résulte une ré-écriture de certaines questions pour lever les imprécisions qui existaient. Ceci permet de converger vers le questionnaire présenté au tableau 2.1. Par exemple, les termes de « description réduite », « description

causale » ou « description hédonique » tels qu'introduits précédemment, sont évités dans les questions. L'utilisation de périphrases telles que « Décrire l'origine technique de ces sons. » ou « Décrire le son en lui-même, c'est à dire sans faire référence à la cause du son. » y est alors préférée. Ces phases pilotes ont aussi permis à l'intervieweur de s'habituer à la gestion de la durée de l'interview. Comme ces interviews pilotes n'ont pas engendré de changements majeurs, les réponses des participants ont été conservées au même titre que celles des autres experts par la suite.

### 2.2.4 | Résumé de la section

#### Collecte des termes

L'intérêt de construire un lexique représentatif de l'environnement sonore du véhicule électrique a été discuté plus tôt. Cependant, cet environnement est assez nouveau, et il n'existe pas encore d'habitude et de littérature détaillées sur celui-ci. Pour permettre la construction du lexique et avoir une base de description complète, une interview guidée par un questionnaire semi-directif, avec 6 questions autour de la description du son puis 4 autour du confort, a été réalisée. L'interview, d'une durée comprise entre 45 minutes et 1h15, a été menée avec 12 acousticiens membres de Renault Group, travaillant dans les différents domaines de l'acoustique liés à la motorisation électrique, l'aéroacoustique, la ventilation etc.

Ce travail permet d'avoir la base de données la plus uniforme possible pour la suite de la construction du lexique. Une fois les interviews passées et enregistrées, les échanges sont entièrement transcrits pour pouvoir analyser des échanges écrits.

## 2.3 | Construction du lexique

### 2.3.1 | Transcription et étiquetage

#### 2.3.1.1 | Transcription

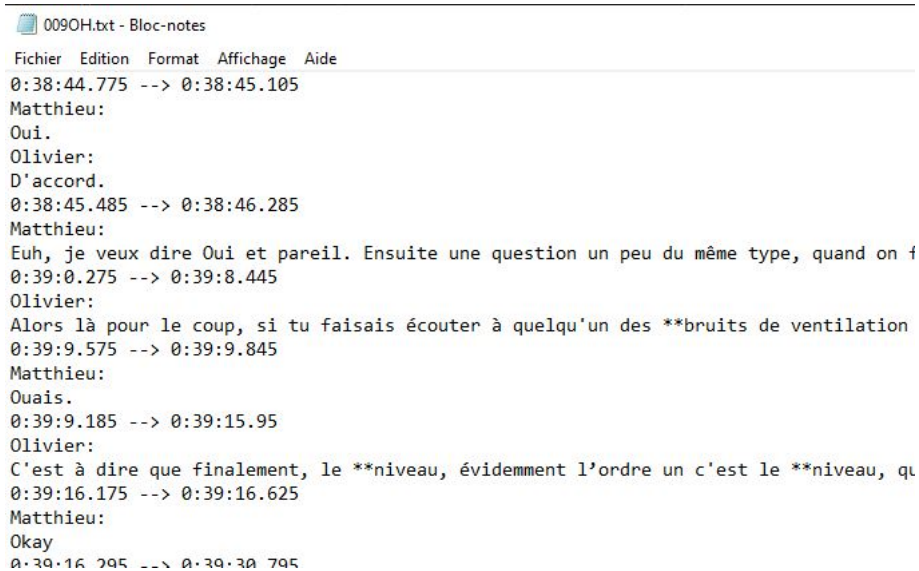
Le travail d'analyse verbale qui suit ces interviews peut commencer seulement après une transcription méthodique des échanges oraux. Chaque interview ayant été soigneusement enregistrée, avec l'accord des experts, la retranscription par écrit est alors possible a posteriori. Cette étape a été réalisée à l'aide d'un logiciel de transcription, Descript<sup>1</sup> suivi d'une importante relecture manuelle pour corriger les transcriptions automatiques.

Chaque interview, une fois transcrite, est mise au même format en vue d'une analyse numérique future, à savoir un fichier .txt comportant l'identifiant de l'interlocuteur, le timecode de l'échange et le texte transcrit (Fig. 2.2). Ce travail a parfois nécessité la réécriture de propos formulés à l'oral, tels que la répétition de termes causée par une formulation hésitante ou l'utilisation d'onomatopées. Ces réécritures ont été réduites au strict minimum pour ne pas trahir les échanges initiaux.

#### 2.3.1.2 | La classe « Nom de son »

Une fois ce travail de transcription fait, avant toute analyse numérique, un étiquetage des termes est effectué. Celui-ci a pour but de relever, dans chaque interview, les différents descripteurs du son, en les catégorisant dans les trois classes de discours précédemment identifiées : Causal, Réduit, Hédonique. Pour faciliter l'identification de ces descripteurs, on s'appuie sur les travaux de Dubois [98] qui identifie

<sup>1</sup> <https://www.descript.com>



```

009OH.txt - Bloc-notes
Fichier Edition Format Affichage Aide
0:38:44.775 --> 0:38:45.105
Matthieu:
Oui.
Olivier:
D'accord.
0:38:45.485 --> 0:38:46.285
Matthieu:
Euh, je veux dire Oui et pareil. Ensuite une question un peu du même type, quand on f
0:39:0.275 --> 0:39:8.445
Olivier:
Alors là pour le coup, si tu faisais écouter à quelqu'un des **bruits de ventilation
0:39:9.575 --> 0:39:9.845
Matthieu:
Ouais.
0:39:9.185 --> 0:39:15.95
Olivier:
C'est à dire que finalement, le **niveau, évidemment l'ordre un c'est le **niveau, q
0:39:16.175 --> 0:39:16.625
Matthieu:
Okay
0:39:16.295 --> 0:39:20.795

```

Figure 2.2. Capture d'écran d'une interview transcrite mise au format prévu pour une analyse numérique.

cinq catégories linguistiques couramment utilisées autour de la description des sons et des bruits de manière spontanée. On s'attend alors à retrouver des types de discours précis en fonction de ces catégories linguistiques :

- **Les noms ou groupes nominaux.** Cette catégorie se réfère directement au son ou au mode de production, parfois avec un nom seul, « *L'alarme* », « *Le klaxon* », mais plus souvent avec un groupe nominal construit autour d'un noyau en « *N de M* ». Le noyau est souvent un terme générique tel que « *Bruit* » ou « *Son* » suivi d'un complément de nom, comme dans la phrase « *Le bruit de ventilation est trop fort* », où le son est identifié par le groupe nominal « *Le bruit de ventilation* ». Ce groupe nominal est souvent réduit à un seul mot issu du complément du noyau, qui permet alors de nommer le son au fil de la discussion plus succinctement : « *La ventilation est trop forte* ». Cette catégorie linguistique favorise le discours causal, en complétant avec des éléments techniques ou contextuels relatifs à l'origine du son.
- **Les noms dérivés de verbes.** Cette catégorie permet de compléter celle présentée au dessus et contient les noms par lesquels sont identifiés des sons, dérivés du verbe à leur origine tels que « *Crisser* », « *Claquer* », « *Rouler* ». Parmi les noms dérivés de verbe, majoritairement un seul type de dérivé avec le suffixe « *-ement* » est repéré tels que « *Le sirènement* », « *Le crissement* », « *Le roulement* », bien que le suffixe « *-age* » soit aussi possible, mais a été beaucoup plus rare. Cette catégorie linguistique peut alors apparaître dans la précédente en faisant partie d'un groupe nominal plus riche tel que « *Le sirènement du moteur électrique* ». Les descripteurs appartenant au type de description causal, et plus particulièrement à la classe nom de son sont alors attendus dans cette catégorie.
- **Les adjectifs.** Dans cette catégorie les adjectifs sont susceptibles de se retrouver dans les types de discours hédoniques ou réduit. Les adjectifs verbaux, qui sont dérivés de verbes, se réfèrent plus souvent au côté hédonique tels que « *Génant* », « *Plaisant* », « *Entrainant* », alors que les adjectifs classiques se réfèrent plus souvent à une description réduite : « *Fort* », « *Harmonique* », « *Rugueux* », « *Brillant* ».
- **Les groupes prépositionnels.** Cette catégorie apporte une fonction de complément circonstanciel, complément du nom, de l'adjectif, du verbe autour d'un son, par exemple : « *Ce son apparaît à 70 km/h sur une piste lisse.* », le groupe prépositionnel « *à 70 km/h sur une piste lisse.* » apporte une description fortement contextuelle et donnera à priori des éléments pour la formulation de la

définition. Ce groupe prépositionnel peut lui-même contenir des noms ou des adjectifs identifiés dans les catégories précédentes.

- **Les propositions relatives.** Cette catégorie apporte des détails sur l'objet créant le son à travers une proposition relative telle que « *C'est le bruit d'un pneu dont la cavité entre en résonance* » et donc offrira par nature des descriptions plutôt contextuelles.

Au sein de la classe des descripteurs causaux, une différence est remarquée dans les usages. Certains descripteurs sont utilisés comme des noms à part entière et servent à reconnaître un objet sonore précis et le nommer, tandis que d'autres descripteurs sont plutôt utilisés pour donner des détails techniques autour d'un son parfois mal identifié ou non nommé. Par exemple la phrase décrivant un son avec un élément causal « *le sifflement du moteur électrique est perceptible* » donnera lieu à l'identification du nom « *sifflement* » dérivé du verbe « *siffler* ». A l'inverse des phrases subordonnées du type : « *Le bruit causé par une fenêtre légèrement ouverte* », donne aussi un discours causal, mais sans permettre l'identification d'un nom de son. De ce fait dans la classification des descripteurs du son, une sous-classe « Noms de son » sera ajoutée à la classe causale, ce qui permet de distinguer les descripteurs nommant la source et ceux donnant des détails sur la cause.

### 2.3.1.3 | Étiquetage

La liste des catégories linguistiques illustre aussi la difficulté de saisir le contexte complet autour de l'utilisation d'un mot, notamment avec des objets complexes tels que les groupes prépositionnels ou les propositions relatives. Ces éléments sont très importants pour mieux comprendre l'utilisation des termes mais sont difficiles à quantifier. La tâche d'étiquetage menée consiste à identifier les descripteurs les plus utilisés parmi les objets linguistiques simples, tels que les noms et adjectifs, et les classer dans les classes de descripteurs identifiées tout en conservant les éléments contextuels contenus dans les objets linguistiques plus complexes. Le tableau 2.3, qui a été construit pour chaque échange enregistré, illustre comment sont classés ces termes simples, en gardant le contexte puisque la transcription complète de l'échange y est associée. A cela s'ajoute un découpage de la séquence sonore, qui permet de réécouter le moment exact de l'interview correspondant à la transcription rapportée. Ce travail d'étiquetage nous permet d'avoir une liste de termes décrivant le son selon les classes de descripteurs et les éléments contextuels associés, ce qui facilitera l'analyse qui s'en suivra. Il apparaît notamment la possibilité qu'un terme soit répertorié dans plusieurs classes de description, comme « *identifiable* » dans l'exemple du tableau, lorsque la transcription ou l'enregistrement ne permettent pas de choisir l'une ou l'autre des classes. La construction de ce tableau a plutôt favorisé des faux positifs, dans le sens où lorsqu'un doute existe, un terme sera étiqueté dans une ou plusieurs catégories, quitte à ce que celui-ci apparaisse comme non pertinent plus tard. Ainsi, des termes tels que « *discrète* » ou « *pic* » qui peuvent ne pas décrire directement le son. Cela se justifie par le fait que l'importance du descripteur sera donnée par les outils d'analyse statistique et sémantique dans la prochaine étape. On peut ainsi se permettre d'avoir des faux positifs à cette étape d'étiquetage qui seront filtrés par la suite.

### 2.3.2 | Sélection d'une liste de termes

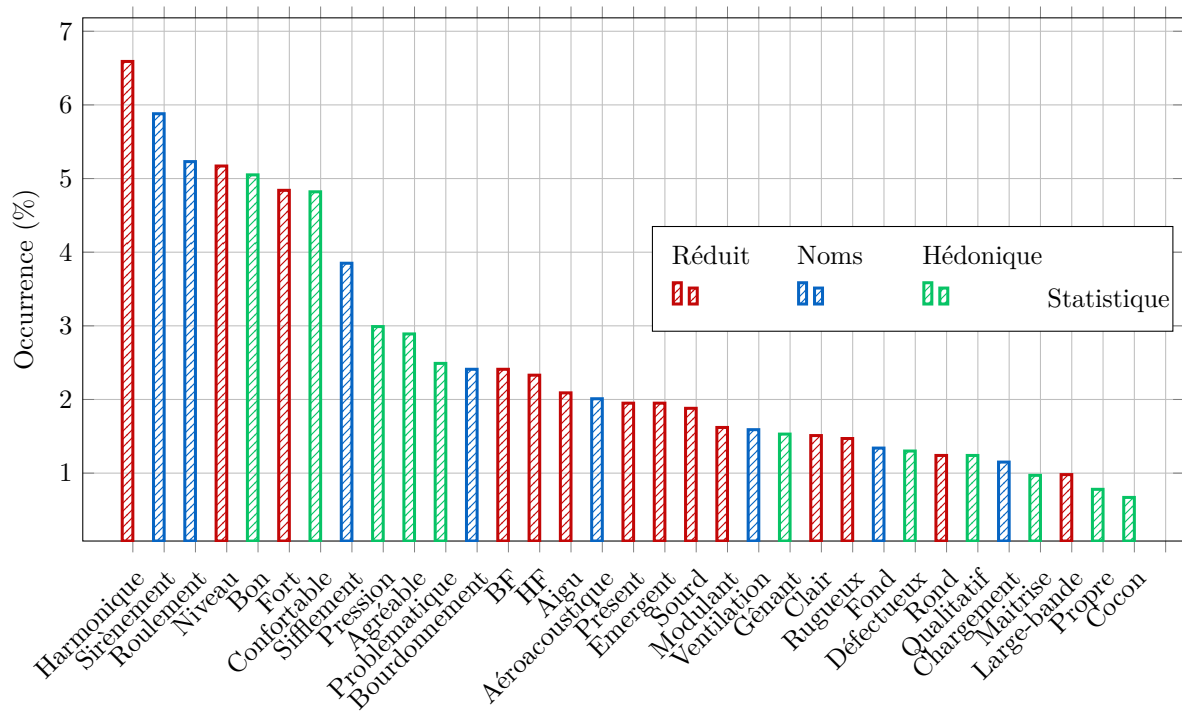
L'objectif final de ce travail est de construire un lexique dédié à la description de l'environnement sonore du véhicule électrique. Le premier élément à construire ici est la liste des termes. Parmi les différentes classes de descripteurs identifiées dans l'étiquetage des interviews, la liste des termes du lexique sera construite à partir de la classe nom de son.

Nom de Son	Causal	Réduit	Hédonique	Transcription
				<b>Hulument1PBCausal.wav</b>
Mode de cavité du pneu	marque, taille, pression  Mode de cavité du pneu	discrète, pic, identifiable	identifiable	[00 :39 :27] <b>Intervieweur</b> : <i>Ok donc bon pareil je vais repasser à travers les paramètres qu'on avait cités tout à l'heure. Si on changeait la <b>marque</b> du pneu, la <b>taille</b> ou la <b>pression</b> ?</i>  [00 :39 :37] <b>PB</b> : <i>Ouais alors là j'ai quelque chose qui peut peut-être un peu plus nous aider en terme de perception, tu peux avoir du <b>mode de cavité de pneu</b>. Et un <b>mode de cavité de pneu</b> c'est une fréquence, pas <b>discrète</b>, faut pas exagérer, mais par contre ça te fait un <b>pic</b> aux alentours des 180-200 Hz qui est très facilement <b>identifiable</b> à la réécoute.</i>

**Table 2.3.** Tableau représentant le processus d'étiquetage sur un extrait d'une interview transcrite. On remarque en particulier que les termes de l'intervieweur sont aussi reportés, puisque son rôle est de reprendre uniquement les termes ayant déjà été mentionnés par l'interviewé, ou présents dans le questionnaire présenté plus tôt.

### 2.3.2.1 | Méthode statistique

Pour obtenir cette liste de termes, dans la construction de leur lexique, Carron et al. [106] utilisent une analyse d'occurrence des termes relevés dans la littérature scientifique couvrant la période de 1877 à 2013, date du début de l'étude. Ainsi, tous les descripteurs réduits du timbre apparaissant plus de cinq fois sont conservés pour former une première liste de descripteurs candidats pour le lexique. Dans notre cas, nous avons remplacé le manque de littérature par les descriptions verbales obtenues lors des interviews. Pour en déduire une liste de termes, nous appliquons en premier lieu le même type d'analyse, que l'on appellera méthode statistique, qui reporte simplement le nombre d'occurrences des termes. Dans le cadre d'une analyse de verbalisation libre, la préparation du corpus à analyser passe souvent par plusieurs étapes, à savoir la tokenization, la suppression des "stop words" puis la lemmatisation [112]. Ce type d'analyse sémantique plus poussée sera utilisé dans un deuxième temps. Dans notre cas, l'étape d'étiquetage pré-sélectionne déjà les termes candidats et les classe en quatre classes de descripteurs mentionnés plus tôt : nom de son, causal, réduit, hédonique. On a ainsi une sélection de l'ensemble des termes relatifs à la description du son. Sur cette sélection, on applique seulement une racinisation des termes avant de calculer leurs occurrences d'apparition dans l'ensemble des interviews. Cela consiste à réduire chaque terme à sa racine - ou son radical - en y retirant tout affixe, ce qui permet de regrouper sous un même morphème les différentes variations d'un terme. Par exemple « *agréable* », « *désagréable* », « *agréablement* », « *agréabilité* » sont des versions dérivées ou fléchies du verbe « *agréer* » de radical « *agré-* ». Tous ces termes seront regroupés après racinisation sous le morphème « *agré-* ». Puis le nombre d'occurrences est compté automatiquement. La figure 2.3 reporte ainsi les résultats obtenus des termes les plus fréquemment utilisés avec la méthode statistique. Les termes sont reportés dans la figure avec leurs formes lexicales la plus cohérente avec leur fonction lorsque cela faisait sens. Les noms de sons sont ainsi des noms communs, et les descripteurs réduits et hédoniques sont mis sous la forme d'adjectif ou d'adverbe, bien qu'ils ne se trouvent pas nécessairement sous cette forme lors de leur utilisation spontanée. On reporte dans le tableau 2.4 les racines utilisées ainsi que les différentes déclinaisons des termes.



**Figure 2.3.** Histogramme représentant les occurrences des descripteurs les plus présents dans les interviews selon la méthode statistique. 77,2% des occurrences cumulées sont ainsi représentées dans la figure

L'occurrence des termes est reportée en pourcentage parmi le nombre total de descripteurs liés au son après racinisation : 5222 termes reportés. Cela permet d'assurer, quand bien même l'étape d'étiquetage comporterait des erreurs d'attribution, que le nombre d'occurrences entre les différentes classes est bien comparable. Cette normalisation met aussi en évidence l'omniprésence de certains descripteurs lors des discussions autour de l'environnement sonore de l'habitable.

### 2.3.2.2 | Méthode sémantique

En complément de la méthode statistique présentée à la section précédente, une méthode dite sémantique, poussant l'analyse plus loin qu'une simple occurrence de termes, est aussi utilisée.

Dans notre cas, nous utiliserons le logiciel Leximancer,<sup>2</sup> qui permet l'analyse de texte sans modèle à priori [113]. Le logiciel revendique aussi une analyse lexicale résiliente à la fréquence d'occurrences des termes, qui se base sur la corrélation entre les apparitions des différents termes : les co-occurrences [114]. Il a notamment été utilisé par Monache et al. [110] pour une analyse d'interviews guidées par un questionnaire semi-directif, similaire à notre étude. Bien que ce logiciel ne soit pas open-source, le fonctionnement général de celui-ci est présenté dans le manuel d'utilisateur.<sup>3</sup> Ce logiciel permet, à partir d'un corpus de texte et sans modèle fourni, d'identifier une liste de concepts, une valeur d'occurrence pour ces concepts, et les liens entre ces derniers. C'est cette possibilité à identifier un concept et son poids dans un corpus complet qui nous intéresse spécifiquement.

Nous expliquons ici plus en détail le fonctionnement du logiciel et la manière dont celui-ci a été utilisé. Smith [114] présente le principe de ce type d'analyse sémantique autour de 6 points :

<sup>2</sup> <https://www.leximancer.com/>

<sup>3</sup> <https://doc.leximancer.com/doc/LeximancerManual.pdf>

Racine	Suffixe	Préfixe
Harmoni- Sirènement Roulement Niveau Bon Fort	-que -eux	In-
Confort- Sifflement	-able -ant	In- Ré-
Press- Agré- Problèm- Bourdonnement Basse-Fréquence Haute-Fréquence Aéroacoustique	-er -ion -able -abilité -ment -e - atique	Sur- Com- -Dés
Prés- Émerg- Sourd- Modul- Ventilation Gèn- Cla- Rug- Bruit-de-fond Déf- Rond- Quali- Chargement Maitr- Large-Bande Propre Cocon	-ent -ence -ent -ence -er -issant -ant -ation -er -s -ant -e -ire -rté -osité -ueux -aut -éctueux -eur -té -tatif -iser -ise	As-

**Table 2.4.** Racines utilisées pour le dénombrement d'occurrences des termes, ainsi que les différentes formes fléchies et dérivées observées dans les interviews. Chaque entrée de la figure 2.3 est détaillée ici. Les occurrences reportées sont donc les occurrences des racines présentées dans le tableau.

- **Préparation du texte.** Cette première partie de l'analyse est commune à toutes les analyses sémantiques en *Natural Language Processing* (NLP) [112, 115] et consiste à supprimer les "stops words", lemmatiser puis tokeniser les éléments du texte.
- **Découverte ontologique supervisée et/ou non-supervisée.** Cette étape consiste à identifier et ordonner, dans le texte complet une fois tokenisé, des « graines ». Ces graines, qui sont donc des mots, pourront former par la suite des concepts en fonction de leurs occurrences et co-occurrences dans le texte. Cette étape peut être non-supervisée dans le cas où aucune attente n'est formulée sur l'analyse du texte. Les concepts apparaîtront alors comme les groupements de graines les plus importants. Une liste de graines peut aussi être fournie en amont au logiciel pour superviser la recherche. Dans notre cas, l'option de la supervision sera choisie, en utilisant la liste déjà constituée dans le chapitre sur l'étiquetage des interviews.
- **Remplir le thesaurus.** Cet aspect est central dans le fonctionnement du logiciel. Une approche "data-driven" permet de constituer ainsi une liste de concept à partir des graines. Cela est possible en appliquant un poids à chacune des graines en fonction du nombre d'occurrences de celles-ci, ainsi que la probabilité de co-occurrences avec d'autres graines et leurs distances respectives dans le texte. Ainsi, deux graines apparaissant souvent ensemble, telles que « *moteur électrique* », ont de fortes chances d'être agrégées dans le thesaurus sous un seul et même concept.
- **Identification de Tag.** Une fois ce thesaurus construit, des "tags" peuvent y être appliqués. Par exemple, si on a gardé trace de l'identité d'un interlocuteur ou de la question à laquelle il répondait, ces informations peuvent être utilisées ici pour rattacher des concepts à un groupe de personnes comme dans l'étude de Monache et al. [110]. Cette option n'est pas explorée ici.
- **Représentation graphique et interaction utilisateur.** Cette étape consiste à représenter dans un plan à deux dimensions la proximité et les liens entre les différents concepts identifiés. Le logiciel identifie en plus à ce moment, des thèmes qui sont des groupements de concepts par proximité. La figure 2.4 affiche la représentation graphique des interviews à partir de la liste de graines fournies. Cette représentation ne sera pas utilisée dans la suite de l'analyse.

La figure suivante (Fig. 2.5) reporte maintenant le classement des différents concepts identifiés par la méthode sémantique et leurs importances relatives dans le corpus complet des interviews. Ils sont présentés dans le logiciel avec une valeur d'occurrence, représentative du nombre de fois que ce concept est évoqué ou mentionné. De la même manière que pour la méthode statistique, on peut alors normaliser nos valeurs d'occurrence des concepts par le nombre total d'occurrences relevées dans l'ensemble des interviews. Malgré la nature différente des méthodes, la comparaison avec la méthode statistique est alors rendue possible, ce qui permet de les afficher sur le même graphique.

### 2.3.2.3 | Discussion des résultats obtenus

La méthode statistique explicite l'importance du terme « *Harmonique* » dans les interviews. Ce descripteur est un cas particulier : son utilisation est détaillée dans le prochain paragraphe (Sec. 2.3.2.4 ). Outre cela, l'analyse d'occurrences révèle que la méthodologie par interviews et relances a effectivement permis de diversifier les modes de description du son. En effet les descripteurs réduits et hédoniques ont un nombre d'occurrences comparable aux descripteurs causaux, quand bien même les experts ont, à priori, un vocabulaire plus développé pour les éléments techniques, ce qui favoriserait une description causale. Outre le descripteur « *Harmonique* », les deux descripteurs réduits les plus fréquents sont « *Niveau* » et « *Fort* », ce qui traduit l'enjeu central de la question du niveau sonore perçu dans l'habitacle, même pour des véhicules électriques réputés silencieux.

Sur la figure 2.5, on observe une bonne corrélation globale des deux méthodes ( $r^2 = 0.71$ ). Cette corrélation, similaire pour la classe seule des noms de sons ( $r^2 = 0.70$ ) que l'on peut voir figure 2.6,

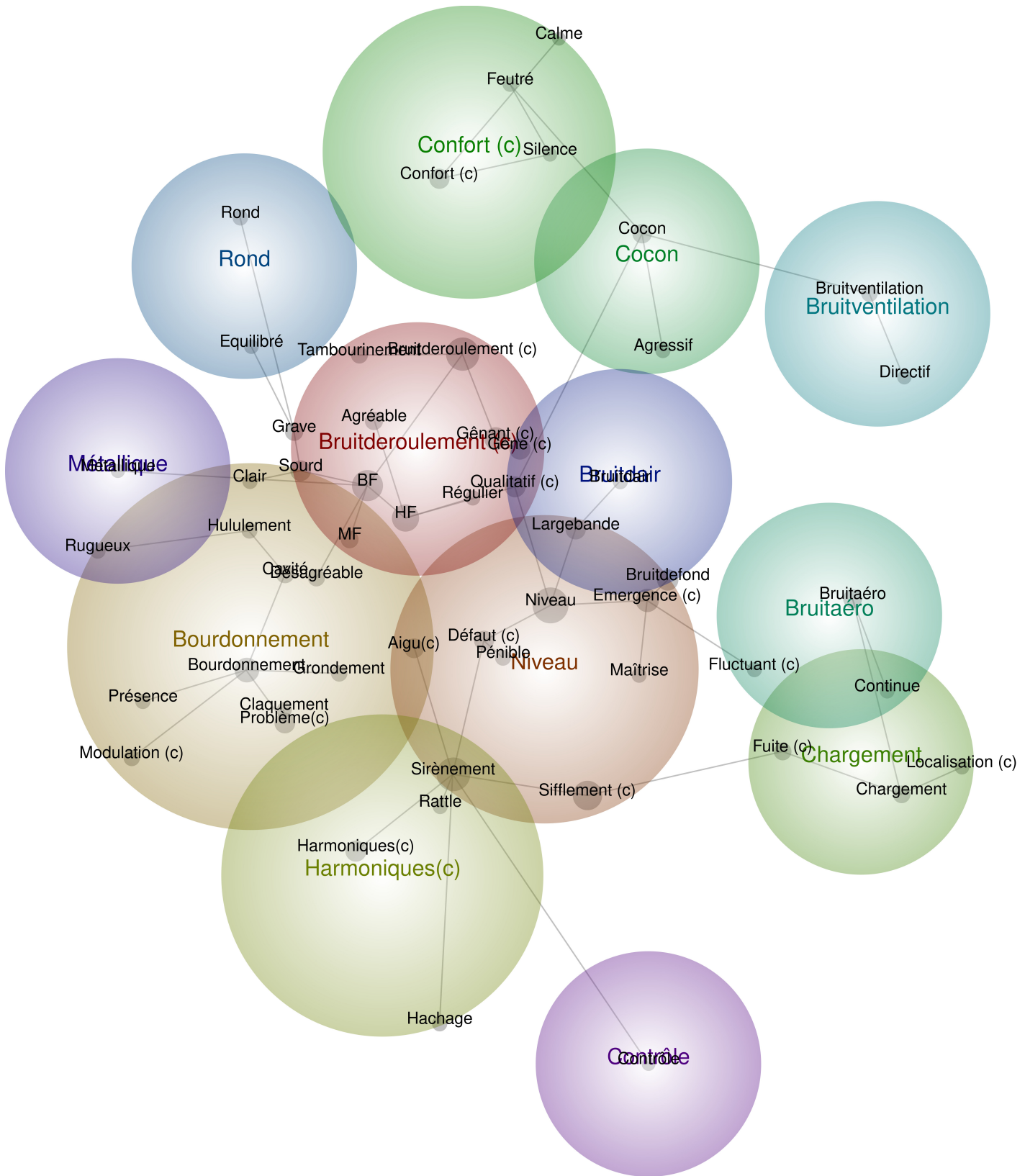
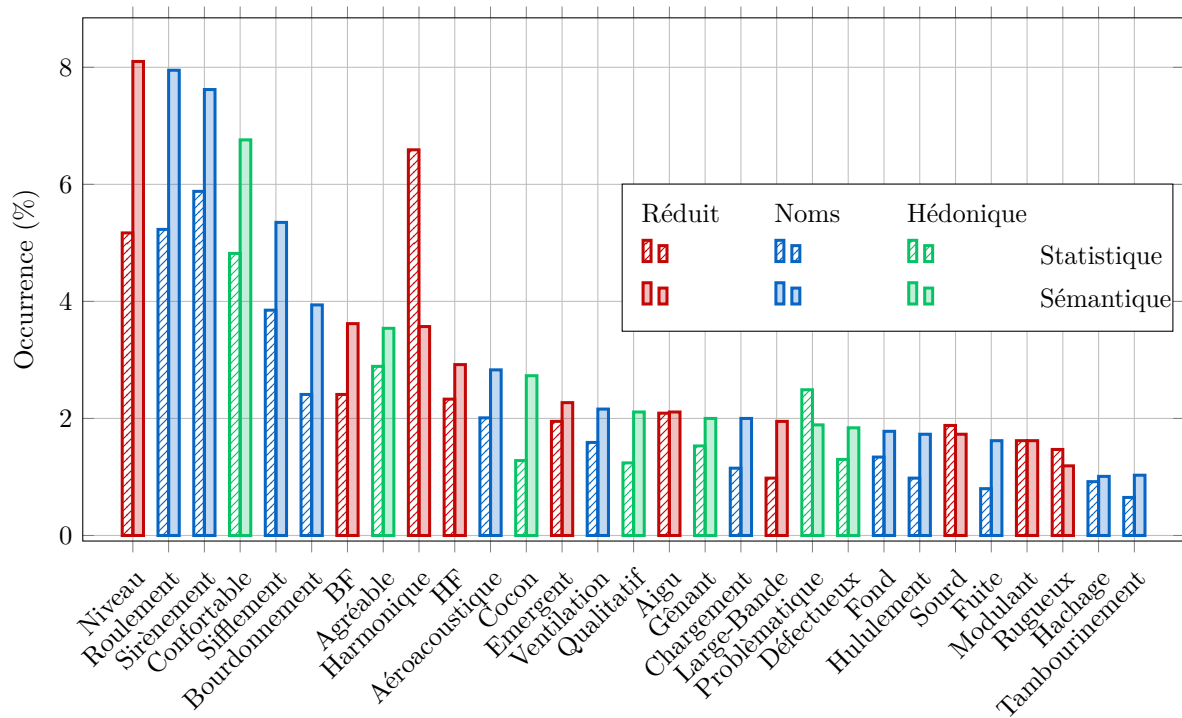


Figure 2.4. Représentation graphique de la carte des concepts identifiés par Leximancer.



**Figure 2.5.** Histogramme représentant les occurrences des descripteurs les plus présents dans les interviews selon la méthode sémantique. Les valeurs d'occurrence statistique sont conservées pour comparaison. Le pourcentage d'occurrence est calculé par rapport au nombre total de termes pris en compte par chacune des méthodes pour permettre une comparaison cohérente des valeurs. Le graphique représente ainsi 87.2% des occurrences cumulées des concepts avec la méthode sémantique et 82.8% des occurrences de termes obtenues par la méthode statistique

conforte l'idée que la liste de termes établie par ces méthodes est pertinente et reproductible. En comparant plus finement les résultats, on remarque en particulier la disparition de certains termes comptés dans la méthode statistique tels que « *Bon* » ou « *Fort* ». La disparition du terme « *Bon* » s'explique par l'homographie avec l'interjection orale courante « *bon* », ce qui crée un faux positif dans la méthode statistique difficile à éviter, et sera filtré dans la méthode sémantique comme un concept ayant peu d'importance. Le terme « *Fort* » quant à lui est assimilé au concept « *Niveau* », le logiciel d'analyse sémantique ayant fusionné ces deux concepts qui apparaissent très fréquemment proches l'un de l'autre.

#### 2.3.2.4 | Le descripteur « *Harmonique* »

Enfin, la forte différence entre les deux méthodes sur l'occurrence du terme « *Harmonique* » visible sur le graphique 2.5 illustre la spécificité de ce terme. Ce descripteur a un rôle très particulier et important dans la description des sons du véhicule électrique. Tout d'abord, il peut être utilisé comme un nom commun pour décrire une composante émergente ou tonale d'un son. Ce nom est particulièrement utilisé pour parler d'une composante du son d'un moteur électrique par exemple. On parle aussi de « *Raie harmonique* » ou « *D'harmonique de rang N* ». Le terme « *Raie harmonique* » sera gardé dans la construction du lexique. Dans ce cas, le descripteur est plutôt dédié à une discussion entre experts, et souvent accompagné de spécifications techniques ou numériques.

D'autre part, « *Harmonique* » peut aussi être utilisé cette fois-ci comme un adjectif. Dans ce cas il se réfère à un son dont la hauteur est proportionnelle à la fréquence de rotation du moteur à l'origine de ce son. Le terme harmonique ici est donc une extension de la définition classique d'un son harmonique en acoustique que l'on peut retrouver dans le Larousse, à savoir : « *Un son susceptible d'être émis par une*

source sonore et dont la fréquence est un multiple entier de celle correspondant au mode fondamental de vibration. On dit aussi son harmonique. ». L'adjectif peut être utilisé en dehors d'un cadre technique puisqu'on observe qu'il est rarement suivi de spécification technique ou numérique comme l'est le nom commun mentionné plus tôt. Dans la construction du lexique, l'adjectif harmonique-moteur sera utilisé pour éviter toute confusion. Enfin, l'utilisation de l'adjectif « *Harmonique* » pour parler de l'harmonie d'un son, au sens musical, et quant à lui très peu utilisé dans le corpus des interviews et ne sera donc pas considéré dans la suite de notre étude.

Un exemple de discussion issu des interviews illustre parfaitement le cas dans lequel le nom et l'adjectif sont alternativement utilisés avec un rôle différent.

**Intervieweur :**

« *D'accord, il y a des sifflements différents en fonction de différents modèles de voiture ou des moteurs ?* »

**ER :**

« *Alors on n'est pas forcément sur les mêmes types d'harmoniques. Parce que les harmoniques... c'est un son qui va être harmonique au moteur. Donc on est sur une fréquence qui est proportionnelle à la vitesse de rotation du moteur. L'harmonique en question elle va dépendre de la topologie de la machine, du nombre de pôles, du nombre d'encoches etc. donc parfois on va avoir des... Typiquement nous chez Renault on a des machines dont les harmoniques c'est souvent l'ordre 24 et 48 qui ressortent. On peut avoir des choix différents qui vont faire ressortir d'autres harmonique, par exemple une harmonique 30 et puis une autre harmonique 60. Finalement ça ne va pas changer beaucoup... dans tous les cas la fréquence va être proportionnelle à la vitesse de rotation du moteur et donc des roues avec un jeu d'émules..* »

### 2.3.2.5 | Liste de termes

De ce travail de classement des termes liés aux noms de sons, on déduit une liste de termes qui constituera la première étape de notre lexique. Pour cette liste, nous nous arrêterons aux 13 premières entrées, ce qui correspond à 94 % des concepts selon la méthode sémantique de noms de sons évoqués en interview (91 % avec la méthode statistique) et 42 % de tous les concepts liés aux sons dans les interviews, toutes classes confondues (33 % avec la méthode statistique). Cette première liste servira donc de base à la construction de notre lexique. On gardera l'ordre d'importance d'occurrences des descripteurs le long de la création du lexique. Ainsi, dans la version finale du lexique, les descripteurs ayant eu le plus d'occurrences seront présentés en premier.

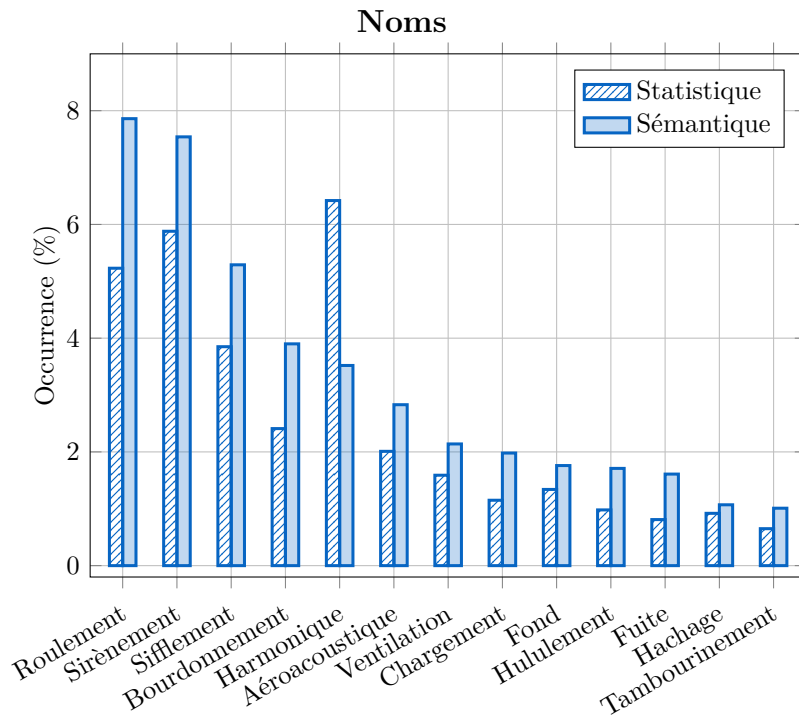
## 2.3.3 | Construction des définitions

### 2.3.3.1 | Objectif des définitions

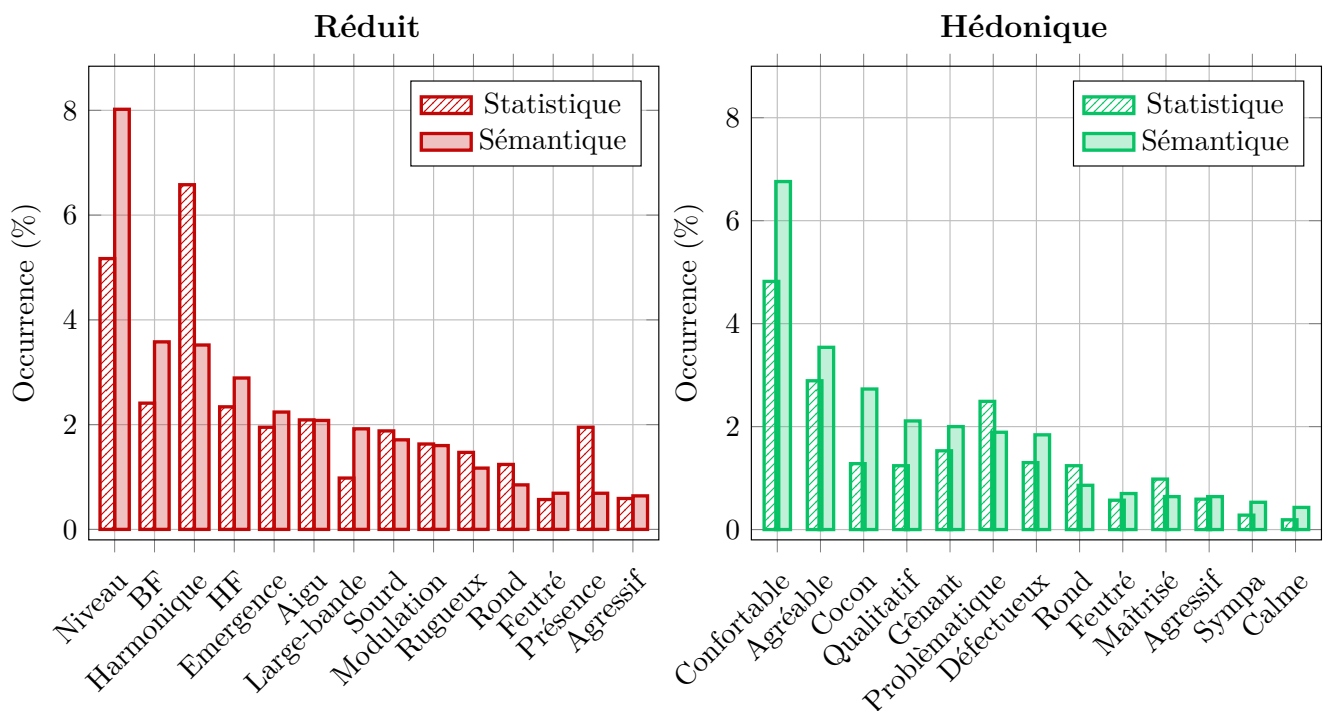
La deuxième étape de notre lexique consiste à construire les définitions liées à chacune des entrées identifiées. Les définitions ont pour objectif de décrire le son nommé en utilisant des éléments de description : nom de son, réduit et hédonique, identifiés lors des interviews, reportés à la figure 2.7. Ces définitions doivent rester accessibles et claires pour des personnes étrangères au monde automobile et donc ne pas être une simple description technique de la cause des sons.

### 2.3.3.2 | Groupe de travail

Un groupe de travail constitué de trois personnes, l'auteur principal, Patrick Susini, directeur de recherche à l'IRCAM dans l'équipe PDS ayant déjà participé à la contribution de deux autres lexique



**Figure 2.6.** Comparaison de l'occurrence des noms de son entre la méthode statistique et sémantique. Le coefficient de détermination entre les deux méthodes est  $r^2 = 0.71$ . L'ordre dans lequel sont présentées les entrées est donné par la méthode sémantique.



**Figure 2.7.** Occurrence d'apparition statistique et sémantique des descripteurs réduits (à gauche) et hédoniques (à droite) parmi tous les descripteurs.

Termes génériques	Termes précis
(Bruit de )Fond	Chargement+Roulement
(Bruit) Aéroacoustique	Chargement+ Fuite + Sifflement aéroacoustique
Sifflement	Sifflement électrique ou Sifflement aéroacoustique

**Table 2.5.** Distinction entre les termes génériques et les termes précis

SpeaK<sup>4,5</sup> ainsi que Eléonore Vialatte, aujourd'hui diplômée de l'Ecole Supérieure d'Art et de Design (ESAD)-TALM le Mans et qui effectuait un stage dans l'équipe à ce moment, a été constitué. Ce groupe est extérieur au monde de l'automobile, à l'exception de l'auteur principal, et possède une expérience et des compétences respectives en création sonore et perception du son. La consigne de ce groupe est d'abord de discuter de la pertinence de chacune des entrées identifiées dans la liste de termes, puis d'établir une définition pour chacun de ces termes. Cette définition finale devra être la plus synthétique possible et permettre à un non-spécialiste de reconnaître le son. Elle comportera une description réduite, causale et hédonique du son. L'intérêt de choisir un groupe de travail extérieur au monde automobile, et avec un domaine de compétence plus proche de la perception, permet de construire ces définitions avec une attention particulière sur la description du son, et une certaine naïveté quant à l'origine et l'importance du phénomène à l'origine de ce son.

### 2.3.3.3 | Liste définitive des termes

La première étape de ce travail est donc de s'arrêter sur la liste des termes qui constituera le lexique. La liste de 13 noms obtenue figure 2.6 n'est pas optimale puisqu'elle reporte seulement les occurrences d'apparition des termes ou des concepts. Le biais principal étant l'incapacité, avec cette méthode, de différencier les termes génériques des termes plus précis. Ces termes génériques sont, par exemple, utilisés dans les interviews pour décrire une famille de sons alors que les termes précis sont utilisés pour décrire les différents sons au sein d'une famille plus large. Le tableau 2.5 détaille parmi la liste des noms de sons représentée figure 2.6, lesquels sont plutôt susceptibles d'être génériques, ou bien précis. Ainsi la liste finale, gardée pour la construction du lexique, sera composée des termes « *Chargement* » et « *Roulement* » plutôt que « *Bruit de fond* ». Nous garderons aussi les termes « *Chargement* », « *Fuite* » et « *Sifflement aéroacoustique* » pour détailler les différents types de sons aérodynamique.

Ainsi, après discussion avec le groupe de travail, le lexique final est composé des 11 termes suivants : *Roulement*, *Sirènement*, *Sifflement Aéro*, *Sifflement Électrique*, *Bourdonnement*, *Ventilation*, *Chargement*, *Hululement*, *Fuite*, *Hachage*, *Tambourinement* .

Avec la même logique de construction, nous aurions aussi pu construire un lexique de granularité moins fines en gardant seulement les termes génériques et ainsi obtenir la liste de termes suivante : *Roulement*, *Bruits Aérodynamique*, *Sirènement*, *Sifflement*, *Bourdonnement*, *Ventilation*, *Hululement*, *Hachage*, *Tambourinement*

### 2.3.3.4 | Construction des définitions

Une fois la liste de terme fixée, la construction de définition s'effectue sur la base des échanges enregistrés pendant les interviews en trois étapes détaillées ici :

- **Ré-écoute des interviews** : Tous les passages des interviews mentionnant le nom de son à définir, identifiés pendant le travail d'étiquetage (Fig. 2.3), sont fournis aux membres du groupe de travail.

<sup>4</sup> <https://speak.ircam.fr/lexique/room-effects/>

<sup>5</sup> <https://speak.ircam.fr/lexique/lexique-ircam/>

En complément de ces extraits, l'ensemble des enregistrements du son en question, qui ont pu être trouvés et qui aident à identifier et caractériser le son, est aussi fourni. Une partie de ces enregistrements a directement été fournie par les experts interviewés.

- **Formulations personnelles :** Après plusieurs jours en possession de ces éléments matériels, chaque membre du groupe de travail propose sa propre définition du terme, qui comporte une description causale, réduite et hédonique du son. A cette étape-là, les définitions ne sont pas encore succinctes, elles doivent surtout comporter tous les éléments qui semblent importants.
- **Formulation globale :** Une fois ces définitions personnelles formulées, une session de travail en commun vise à fixer une formulation finale de la définition, à l'aide de chacune des définitions personnelles et des éléments fournis à la première étape.

### 2.3.3.5 | Exemple avec le hululement

Le son de hululement est identifié dans l'analyse des interviews avec une occurrence de 0.98 % avec la méthode statistique et 1.73 % avec la méthode sémantique parmi tous les descripteurs mentionnés en interviews. Il se place alors à la dixième place en terme de Nom de son revenant le plus fréquemment selon la figure 2.6. De plus, le groupe de travail a considéré ce terme comme pertinent, il figure donc dans la liste des termes à définir. L'étape de ré-écoute des interviews, pour mieux apprécier le contexte autour de ce son, est fait à partir de 26 min 30 d'enregistrement, réparties parmi 6 participants mentionnant le « hululement », sa cause, sa description ou son jugement. A partir de ces ré-écoutes, ainsi que des extraits sonores fournis, chaque membre du groupe de travail formule sa propre définition du son :

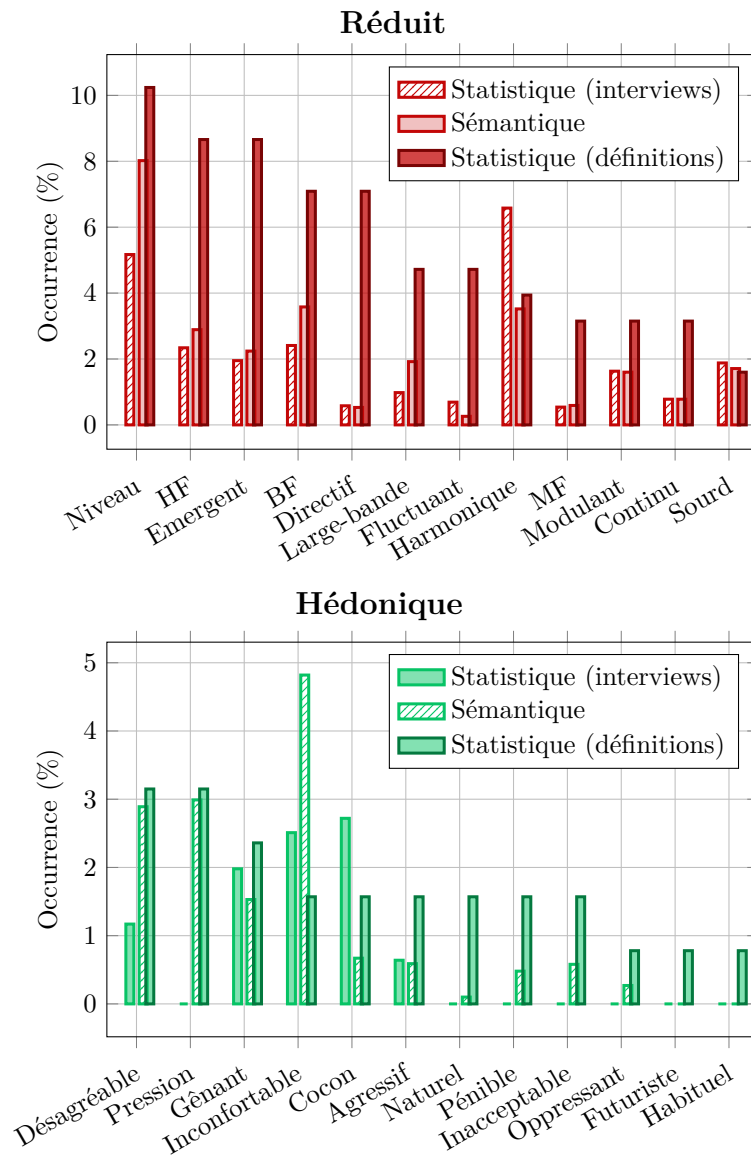
**Formulation personnelle membre 1 :** *Le mode de cavité, aussi appelé hululement, est un son émergent, modulant en amplitude et basse fréquence. Il se caractérise aussi par une constance dans sa hauteur et on peut identifier deux raies harmoniques le composant. Son caractère émergent le rend identifiable dans le bruit de fond ce qui le rend d'autant plus inconfortable et fatigant qu'il est constant. Son caractère inconfortable augmente lorsqu'il devient plus aigu.*

**Formulation personnelle membre 2 :** *Bruit basse fréquence désagréable et continu, composé de deux raies proches qui oscillent rapidement et dont l'amplitude module en même temps que le bruit de roulement. Ce bruit est comparable au chant de la chouette. Ce signal provient de la peau du pneumatique qui va exciter l'air contenu dans le pneu. C'est le mode de cavité du pneu.*

**Formulation personnelle membre 3 :** *C'est un son produit par un mode de cavité pneumatique qui se superpose au bruit large bande de roulement. Cela correspond à deux raies Basse-Fréquences (BF) autour de 200 Hz avec une émergence (6 dB) qui produisent un battement, qu'on peut visualiser sous la forme d'une modulation d'amplitude rapide. C'est inconfortable, voire fatigant car constant, comme un grelot permanent.*

À l'issue de l'élaboration des définitions individuelles, une session de travail en groupe est menée pour s'accorder sur une définition commune. Les éléments qui décrivent les causes sont repérés en **bleu**, ceux qui relèvent du discours réduit en **rouge**, et ceux qui se réfèrent au côté hédonique, en **vert** :

**Formulation commune :** *Le mode de cavité, ou hululement, d'un pneu pour le roulement aux alentours de 80 km/h, ou de l'intersection pare-brise pavillon pour l'aéro, produit un son, aussi appelé hululement. Il correspond à deux raies proches en basses fréquences (autour de 200 Hz), avec une modulation d'amplitude rapide ( $\approx 10$  Hz) très semblable à un battement. Son émergence ( $\approx 6$  dB)*



**Figure 2.8.** Comparaison de la fréquence d'occurrence des descripteurs réduits et hédoniques entre la méthode statistique, sémantique et la fréquence d'occurrence dans les définitions formulées.

*par rapport à un bruit large bande (roulement) le rend identifiable. C'est un son plutôt constant en fréquence, amplitude, et vitesse de modulation qui, par conséquent, devient fatigant.*

### 2.3.3.6 | Résultats

Les définitions construites sont d'abord comparées en occurrence de termes avec les interviews dans leur globalité, on réplique la figure 2.7 en ajoutant les occurrences des termes dans les définitions formulées (Fig. 2.8). Ce résultat permet, d'une part, de valider a posteriori les méthodes d'analyses sémantiques et statistiques utilisées plus tôt. On s'assure aussi, d'autre part que le vocabulaire spontanément utilisé par les experts est bien retranscrit dans les définitions.

### 2.3.4 | Illustration sonore des termes

Le troisième élément constitutif de ce lexique, essentiel pour obtenir un lexique selon la méthodologie SpeaK [105] et qui garantit un usage non-expert, est l'illustration des termes présentés. Elle est faite à partir d'extraits sonores de quelques secondes, permettant à un utilisateur du lexique de se faire une idée du son. Ces extraits doivent donc illustrer, autant que possible, chacune des facettes de la définition. Comme pour la formulation des définitions, ils doivent être suffisamment synthétiques pour pouvoir être écoutés rapidement, et être suffisamment clairs pour que l'identification du descripteur soit aisée.

En s'inspirant du travail réalisé par Maxime Carron et Thomas Rotureau [106], l'illustration des définitions est présentée en plusieurs groupes ayant chacun leurs spécificités. Ces quatre groupes ci-dessous, permettent d'appréhender le son dans différents cas.

#### 2.3.4.1 | Extraits contextualisés

Les extraits contextualisés ont pour but de présenter le son tel qu'il peut être réellement entendu par un utilisateur. Il présente donc un cas d'utilisation où le son a été enregistré, le plus souvent avec un micro installé à la place du conducteur. Cette série d'extraits a été construite avec l'aide des experts interviewés. Une fois les interviews réalisées et après une première ré-écoute, chaque expert a été recontacté pour demander des exemples de sons qu'ils avaient évoqués. Certains d'entre eux ne peuvent pas tous être facilement reproduits : il n'est donc pas possible de mener une campagne d'enregistrement en provoquant chacun de ces sons séparément. On peut par exemple citer le bourdonnement, ou le hululement, qui sont deux sons assez rares et corrigés rapidement si possible. Par conséquent, les différents extraits contextualisés sont difficilement comparables entre eux car ils ne correspondent pas à la même situation de conduite, au même véhicule, ni au même placement de micro.

#### 2.3.4.2 | Extraits décontextualisés

Les extraits décontextualisés permettent de compléter les extraits contextualisés. En effets dans le cas d'un son faible en niveau (hachage), peu émergent (hululement), ou qui apparaîtrait uniquement entourés d'autre sons (sifflement électrique), il peut être difficile de discerner le son ciblé par la définition. Ces extraits sont donc une version décontextualisée de ces enregistrements et présentent le son seul, soit grâce à des moyens expérimentaux, soit grâce à des moyens numériques.

L'utilisation de moyens numériques se prête particulièrement bien pour les sons de caractère émergents, tels que le sirènement, les sifflements ou le hululement. Dans ces cas, les extraits contextualisés sont modifiés avec un filtre passe-bande centré autour de l'émergence à isoler. L'utilisation de filtre convolutif *Denoiser*, qui permet d'enlever un son large-bande d'une piste à partir d'un extrait de ce même son, combiné au filtre passe-bande permet ainsi d'isoler efficacement un sifflement aéro ou hululement du bruit de fond. Le cas de sons harmoniques-moteurs, tels que le sirènement ou sifflement électrique, nécessite d'utiliser des outils plus élaborés permettant un suivi d'harmoniques dont la fréquence fondamentale évolue dans un spectre. Les logiciels conçus pour être utilisés dans l'industrie automobile, où ces sons harmonique-moteurs sont habituels, tel que Ansys VRX Sound,<sup>6</sup> permettent de réaliser efficacement ce suivi d'harmonique, et d'isoler ainsi une émergence de son bruit de fond, malgré la variation en hauteur de celle-ci.

<sup>6</sup> <https://www.ansys.com/resource-center/brochure/vrxperience-sound-datasheet>

Lorsqu'il est possible d'isoler l'objet à l'origine du son, les moyens expérimentaux sont plus adaptés pour isoler un son. Par exemple, le son de hachage causé par le découpage de la tension continue de la batterie en tension alternative, est étudié de manière précise. Il est possible de se rapprocher des experts travaillant sur les bancs moteurs expérimentaux [116] leur permettant d'obtenir des mesures ou des enregistrements du bloc moteur ou du bloc de hachage seul. Un autre exemple de son décontextualisé, issu de manipulations expérimentales, est le cas du son de bourdonnement. Ici la cause exacte du bourdonnement sur un véhicule précis avait été identifiée et corrigée. Il s'agissait d'une mauvaise répartition des masses dans un système de compression d'air, créant un balourd, responsable d'une fréquence de résonance très basse fréquence, résultant en un bourdonnement dans la cabine. L'identification de la cause technique a permis aux acousticiens qui travaillent sur le sujet, de volontairement provoquer un bourdonnement en rajoutant des masses pour créer ce balourd, et ainsi être capable de l'enregistrer précisément.

#### 2.3.4.3 | Extraits synthétisés

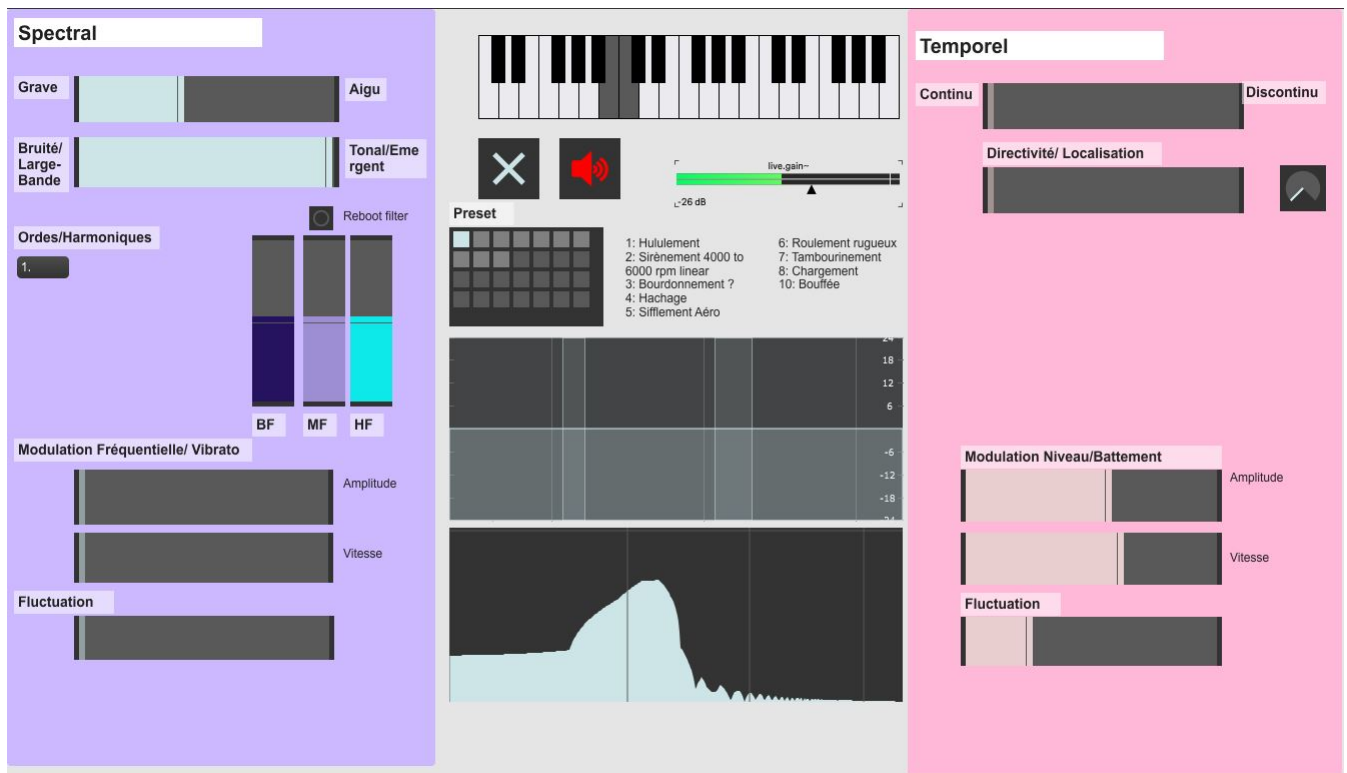
Les extraits décontextualisés sont donc sortis de leur cas habituel d'apparition et on peut en effet s'attendre à ce que le comportement acoustique ne corresponde pas à la réalité de ce que l'on peut entendre. En revanche, ils sont d'une grande aide pour entendre en détail ce son et ces particularités, autrement masquées par d'autres sons. Afin de compléter ces deux modes d'illustration, un troisième mode utilise des sons de synthèse qui reproduisent un comportement particulier du son à représenter. Cette synthèse ressemble au son sans avoir nécessairement de réalité physique derrière. On appellera ce type d'illustration une « Synthèse essentielle », le but est de seulement recréer l'essence même du son et ses éléments principaux.

Un synthétiseur a ainsi été construit pour cette tâche à partir de celui construit (voire Chapitre. 3) pour l'ensemble des sons synthétisés. Grâce à l'interface implémentée représentée figure 2.9, ce dernier permet à des non-spécialistes d'être capable de synthétiser un son sur la base d'une description verbale. Les descripteurs reportés dans ce synthétiseur sont ceux relevés dans l'analyse lexicale. Volontairement, le synthétiseur ne laisse apparaître aucune valeur numérique de fréquence, bande-passante, période de modulation ou autre, pour s'assurer que l'utilisation de celui-ci se fait uniquement sur la base de l'écoute. Par conception, ce type de synthèse n'est pas amenée à recréer des sons réalistes, ni très détaillés. C'est pour cette raison que nous appelons cette synthèse « Synthèse essentielle » et qu'elle a pour but d'être écoutée accompagnée des extraits contextualisés et décontextualisés.

L'exemple en capture figure 2.9 est celui du hululement. On y repère ainsi les éléments principaux de la définition telle que formulée dans le lexique. Tout d'abord, le clavier représente la co-existence des deux raies mentionnées dans la définition puisque deux touches sont enfoncées, créant ainsi deux pistes de hauteurs légèrement différentes. Chacune de ces pistes est ensuite programmée de la même manière à partir de l'interface capturée, seule la hauteur change. On y voit ainsi que le son est à mi-chemin sur le curseur « Grave-Aigu » et avec un caractère Tonal/Emergent assez élevé. On s'assure ainsi que chaque piste mentionnée plus tôt créera bien une raie émergente BF. Enfin, le caractère de modulation temporelle présent dans la définition, est réalisé avec le bloc temporel à gauche, et un curseur à mi distance sur les dimensions « Fluctuation » et « Modulation ».

#### 2.3.4.4 | Extraits semblables

Enfin, le dernier type d'illustration provient des descriptions spontanées obtenues en interview, ou des termes utilisés provenant d'un univers extérieur à l'automobile. On peut encore une fois citer le Hululement, défini par « *Cri des rapaces nocturnes, notamment des hiboux.* » selon le dictionnaire



**Figure 2.9.** Capture de l'interface Max permettant de synthétiser les sons du lexique à partir d'une description verbale. Le bloc à gauche permet de contrôler les éléments liés au spectre du son. Le bloc à droite contrôle les éléments temporels. Le clavier midi représenté en haut permet de choisir le nombre de voix jouées simultanément. Les autres éléments sont des éléments de visualisation.

Larousse. Ce terme a semble-t-il été choisi pour sa proximité entre le son d'origine, i.e. le cri du rapace, et celui décrit dans l'automobile, i.e. le mode de cavité du pneu. Des extraits qui illustrent le cri du rapace, et plus précisément de la chouette sont alors ajoutés à la définition pour maintenir ce lien. Pour d'autres sons, il est parfois mentionné la grande similitude entre le son en question et celui d'une autre nature. Par exemple, dans le cas du sifflement aérodynamique, la forte ressemblance avec le sifflement de bouilloire a été spontanément reportée comme très similaire dans sa sonorité, et donc ajouté à ce groupe d'illustration.

### 2.3.5 | Résumé de la section

#### Construction du lexique

Une fois les interviews réalisées, la construction du lexique peut commencer. La première étape de cette construction a été de procéder à une étape d'étiquetage, qui consiste à identifier dans le texte les différents descripteurs utilisés ainsi que la catégorie, découlant des trois modes de descriptions identifiés précédemment : Nom de son, issu de la catégorie de description causale, réduit et hédonique, issu de la catégorie de description contextuelle. Une fois cette étiquetage effectué, les descripteurs les plus importants sont identifiés en comparant deux méthodes analytiques. Une méthode dite statistique, ordonnant ces descripteurs selon la fréquence d'apparition de la racine du terme dans les interviews, et une méthode sémantique, ordonnant les descripteurs selon l'importance calculée dans le corpus complet du concept associé par l'outil Leximancer. Ces deux méthodes corrèlent fortement et permettent d'identifier une liste de Noms de sons, qui sera donc la liste de termes candidate pour notre lexique.

Pour chacun de ces termes, une définition formulée par un groupe de travail, sur la base des interviews enregistrées et des exemples sonores fournis par les experts, est proposée. Une illustration sonore des termes et définitions est aussi construite en quatre catégories : des extraits contextuels, décontextualisés, synthétisés et rappelant l'origine du terme.

## 2.4 | Validation et finalisation

### 2.4.1 | Validation

Une fois la première version du lexique construite, obtenir un retour sur celui-ci est primordial pour le valider. D'une part, parce que la construction d'un lexique à partir d'interviews semi-directives est originale et nécessite alors une validation. D'autre part, parce que la liste des illustrations sonores est assez hétérogène, parmi laquelle une sélection par des oreilles expertes est aussi nécessaire. De ce fait, le meilleur moyen pour valider ce lexique complet et donc de se tourner à nouveau vers le groupe d'experts sollicités en premier lieu pour les interviews à l'origine de ce lexique.

Les 12 experts qui ont participé aux interviews sont ainsi recontactés. Le retour qui leur est demandé est constitué de deux étapes :

- Un questionnaire par mail permettant d'avoir une évaluation quantitative du lexique, du choix des termes, des définitions et des illustrations sonores.
- Un échange en discussion libre permettant d'avoir plus de détails qualitatifs sur les retours obtenus par écrit.

#### 2.4.1.1 | Validation quantitative des experts interviewés sur le lexique

Le questionnaire est envoyé avec la première version du lexique contenant la liste des termes identifiée, les définitions formulées par le groupe de travail et l'ensemble des illustrations sonores choisies. Avec les mêmes contraintes que lors de la rédaction du questionnaire semi-directif menant aux interviews, ce questionnaire doit ne pas être trop long et aussi doit s'assurer que les réponses de chaque participant ont un poids équivalent. De ce fait le questionnaire est composé presque exclusivement de questions fermées auxquelles le participant pourra répondre via une échelle de Likert en 5 points, pour laquelle on attribue la valeur +2 à la notation « Tout à fait d'accord » et -2 à la valeur « Pas du tout d'accord ».

	A	B	C	D	E	F	G	H
1	<b>Roulement</b>							
2	<a href="https://speak.ircam.fr/lexique/ev-descriptor/roulement-rugueux/">https://speak.ircam.fr/lexique/ev-descriptor/roulement-rugueux/</a>							
3	<b>1. Général</b>							
4	Ce son-là est adapté pour décrire le paysage sonore de l'habitacle du véhicule électrique							
5	D'accord							
6	<b>2. Définition</b>							
7	Décrit bien les propriétés du son							
8	Ni en désaccord ni en accord							
9	Décrit bien le jugement que vous y associez							
10								
11	Décrit bien la cause technique							
12								
13	<b>3. Les extraits</b>							
14	"Audioscope lisse" illustre bien le son et la caractéristique associée?							
15								
	<input type="checkbox"/> Tout à fait d'accord <input checked="" type="checkbox"/> D'accord <input type="checkbox"/> Ni en désaccord ni en accord <input type="checkbox"/> Pas d'accord <input type="checkbox"/> Pas du tout d'accord							
19								
20	"Synthèse Essentielle Rugueux" illustre bien le son et la caractéristique associée?							
21								
22	<b>4. Commentaires</b>							
23	Avez-vous des remarques à faire sur ce son ou sa page Speak ?							
24								
25								
26								
27								

**Figure 2.10.** Capture d'écran du questionnaire envoyé aux experts, rédigé via un tableur pour faciliter la mise en place de la procédure. On peut observer au bas de la figure les différentes feuilles du tableur permettant de répondre aux questions pour les autres entrées du lexique.

Le questionnaire comporte trois sections, ciblées sur les trois éléments constitutifs du lexique, à savoir une section portant sur la pertinence du terme, une seconde sur la définition et une troisième sur les extraits sonores, auxquelles s'ajoute une quatrième section permettant un commentaire libre. La section sur la pertinence du terme appelé « général » (Fig. 2.10) porte uniquement sur l'intérêt du terme pour décrire l'environnement sonore de l'habitacle du véhicule électrique. La section sur la définition évalue différemment les trois types de description du son identifiés. Enfin la section sur les extraits sonore demande une évaluation différente de chacun des extraits présentés. La capture 2.10 montre l'outil utilisé pour partager ce questionnaire et par la même occasion les questions posées dans chacune des sections évoquées.

Le questionnaire est exactement le même pour chaque entrée du lexique, seul changent les noms des extraits sonores illustrant chaque terme. Pour respecter la contrainte du temps, il n'est pas explicitement demandé aux experts de remplir le questionnaire pour chaque son mais à minima sur les trois sons que le participant a décrit pendant l'interview. Comme le nombre d'experts sollicités est faible puisqu'il s'agit de personnes déjà interviewés plus tôt (N=12), il a été possible de personnaliser l'envoi du questionnaire et de rappeler explicitement les sons mentionnés par chacun pour avoir un retour. Ainsi on est assuré d'obtenir au moins un retour pour chaque entrée du lexique. Parmi les 12 experts recontactés, 9 ont ainsi pu donner leurs retours pour évaluer le lexique pour un total de 30 sons évalués. Les retours obtenus sont synthétisés figure 2.11. Maintenant qu'on a une valeur numérique pour chaque question on peut procéder à une analyse statistique descriptive en calculant la moyenne des réponses.

	Général	Réduit	Hédonique	Causale	Extrait
Roulement	2	1,8	1,4	2	1,7
Sirènement	2	0,7	1	0,2	1
Sifflement	1,6	1,4	1,8	1	0
Bourdonnement	0,4	1,2	1,8	0,8	0,6
Ventilation	2	2	2	2	1
Chargement	2	1	2	1,5	1,2
Hululement	0	2,0	2,0	1,0	1,6
Fuite	1	2	2	2	0,8
Hachage	1,7	2	2	2	0
Tambourinement	1,6	1,6	1,8	1,4	0,6

**Figure 2.11.** Score moyen sur les experts qui ont évalué les différentes catégories du lexique sur une échelle de Likert en cinq points allant de -2 à +2. Les lignes correspondent aux différentes entrées du lexique, et les colonnes aux différentes catégories évaluées.

#### 2.4.1.2 | Retour qualitatif des experts interviewés sur le lexique

Cette première phase de retour nous permet déjà de quantifier la qualité des différents éléments de notre lexique et d'aborder les points problématiques plus en détails. Après la collecte des retours auprès des experts, ceux-ci ont donc été sollicités une dernière fois pour un échange de 30 minutes au téléphone ou en présentiel lorsque cela était possible pour avoir plus de détails sur les retours donnés. Ce type de retour est particulièrement utile pour aider à la reformulation des définitions lorsque celles-ci posent problème. Ces discussions sont entièrement guidées par les évaluations fournies par les experts. L'expert et l'intervieweur reviennent systématiquement sur chaque retour formulé pour ajouter des éléments si nécessaire. Par exemple, on peut noter sur les retours quantitatifs reportés figure 2.11, les mauvaises notations autour du terme « Bourdonnement ». La définition initialement proposée à la suite du processus de définition (Sec. 2.3.3) présentée plus tôt est : « *Le bourdonnement est un son très BF à la limite des vibrations. Il se caractérise par un niveau très élevé. Il se rapproche d'un "sirènement" dans sa dépendance au régime moteur (harmonique-moteur) et intermittent bien que beaucoup plus basse fréquence et plus élevé en niveau. Il est très inconfortable et peut provoquer une sensation de douleur ou de malaise physique, notamment en provoquant une sensation de pression au niveau des oreilles. Il est causé par l'excitation de modes de résonance très basse fréquence (comme le mode de caisse) par un harmonique basse fréquence d'une machine tournante.* ».

Les retours qualitatifs qui en sont fait au cours de la discussion précisent que le parallèle avec le sirènement n'est pas pertinent et qu'il manque le côté fatigant très important causé par le bourdonnement. De plus, ce retour, en passant systématiquement sur chaque question, a permis de comprendre que les faibles notations sur les exemples sonores étaient principalement causées par le système d'écoute utilisé lors de l'exploration du lexique. En effet le bourdonnement est un son très basse fréquence et les extraits sonores pour l'illustrer le sont aussi, ce qui rend leur restitution par un système audio délicate. Cette mention a ainsi été ajoutée dans la définition finale : « *Le bourdonnement est un son BF voire très BF ( $\approx 40\text{Hz}$ ) ce qui le rend difficile à restituer sur des systèmes audio classique, avec un niveau assez élevé. Il possède un caractère harmonique-moteur et peut être discontinu en phase d'accélération. Il est oppressant et exerce une pression acoustique sur les tympans, comme la sensation d'être sous l'eau, il entraîne beaucoup de fatigue. Le bourdonnement rappelle les moteurs thermiques et souligne un défaut. C'est un bruit solidien émis par un ou plusieurs panneaux de l'habitacle. Le bourdonnement se rapproche du battement d'ailes de bourdon et du "tambourinement" dans sa sonorité.* ».

A partir de ces retours détaillés sur chaque entrée du lexique, les définitions du lexique initialement formulées sont alors modifiées en conséquence pour arriver à leurs versions définitives. Ces retours permettent aussi de sélectionner les extraits sonores les plus adaptés parmi ceux présentés pour illustrer les définitions. Ils ont aussi permis de trancher sur la question autour du terme « sifflement électrique » et « sirènement » qui sont deux termes se référant à un son très proches, autant par leurs causes que par leurs aspects. Il a ainsi été choisi de garder deux termes séparés.

### 2.4.2 | Lexique final

On reporte ici l'intégralité du lexique : la liste de termes ainsi que les définitions formulées à l'issue du groupe de travail et du retour par les experts lors du travail de validation. Le lexique est gratuitement accessible en ligne à l'adresse suivante : <https://speak.ircam.fr/lexique/ev-descriptor/>. Une traduction du lexique en anglais est aussi proposée sur le site. La procédure de traduction proposée des termes est détaillée dans l'annexe B.

#### 2.4.2.1 | Roulement

Le roulement est un son large-bande plutôt BF avec des zones de fréquences privilégiées qui dépendent du type de pistes, majoritairement BF pour une piste granuleuse/gravillonnée, HF pour une piste mouillée ou équilibré BF/Haute-Fréquences (HF) pour une piste lisse. Il est aussi large-bande mais moins directif que le « chargement » et la « ventilation ». Un bon roulement est faible en niveau, régulier et glissant, rappelant un tapis roulant. Il devient irrégulier lorsqu'apparaissent des fluctuations ou des émergences tel que le « hullement », s'il est trop BF et erratique il se rapproche du « tambourinement ». Le roulement est un son naturel intrinsèque au véhicule, qui informe le conducteur ou la conductrice sur la vitesse du véhicule. La vitesse du véhicule ainsi que le type de véhicule influent surtout sur le niveau. Il est causé par le contact pneumatique/chaussée qui vient exciter différents modes des jantes, des panneaux, etc.

#### 2.4.2.2 | Sirènement

Le sirènement est un son émergent, Moyenne-Fréquences (MF), de type harmonique-moteur et légèrement modulant. Son caractère harmonique-moteur, proche du « sifflement électrique » informe du fonctionnement du moteur et rend le son naturel. Son caractère modulant peut devenir dissonant, ce qui rend le son désagréable ou acide. Son caractère émergent peut être discontinu sur des phases d'accélération, ou continu lorsque la vitesse moteur est stable, pouvant laisser place à du « sifflement électrique » ; il devient alors inconfortable voire pénible. Il est produit par un engrènement de pignons.

#### 2.4.2.3 | Sifflement aéroacoustique

Le sifflement aéroacoustique correspond à un son HF, de caractère émergent, très directif et stable. Il est facilement rattaché à sa source par sa localisation et son émergence, on dit qu'il est signé. Sa signature fréquentielle ainsi que sa hauteur évoluent avec la vitesse. Il est moins fluctuant et aussi émergent que la « fuite ». Il est très désagréable voire inacceptable. Le sifflement aéroacoustique est dû à l'excitation par un flux d'air d'une lame d'air entre deux pièces telles que des barres de toit, un mauvais accostage sur rétroviseur ou une aspérité dans un conduit de ventilation, ce qui produit un effet flûte, rappelant aussi une bouilloire sifflante.

#### 2.4.2.4 | Sifflement électrique

Le sifflement électrique correspond à un son HF ou très HF, de caractère émergent, localisable. Il se rapproche et peut être confondu avec le « sirènement » par son caractère harmonique-moteur et émergent, bien que le sifflement électrique soit plus HF que le « sirènement ». Il se rapproche du « sifflement aéroacoustique » par son caractère HF et localisable. Bien qu'il soit jugé moins inacceptable et plus cohérent avec le véhicule électrique que le « sifflement aéro », il reste considéré comme gênant et pénible. Le sifflement électrique est principalement causé par le fonctionnement des machines électriques et aux vibrations induites par leur champ électromagnétique.

#### 2.4.2.5 | Bourdonnement

Le bourdonnement est un son BF voire très BF ( $\approx 40$  Hz) ce qui le rend difficile à restituer sur des systèmes audios classique, avec un niveau assez élevé. Il possède un caractère harmonique-moteur et peut être discontinu en phase d'accélération. Il est oppressant et exerce une pression acoustique sur les tympans, comme la sensation d'être sous l'eau, il entraîne beaucoup de fatigue. Le bourdonnement rappelle les moteurs thermiques et souligne un défaut. C'est un bruit solidien émis par un ou plusieurs panneaux de l'habitacle. Le bourdonnement se rapproche du battement d'ailes de bourdon et du « tambourinement » dans sa sonorité.

#### 2.4.2.6 | Ventilation

Le son de ventilation correspond à un bruit large bande, plutôt directif et MF, auquel peut se superposer des éléments émergents type « sirènement » ou « sifflement ». Son caractère agressif dû à la présence de HF est, comme le niveau sonore, lié à la vitesse de rotation du pulseur. C'est un son qui contribue à l'effet cocon mais dont un niveau sonore trop élevé et l'agressivité, peuvent nuire à cet effet et rendre le son désagréable. Les sons de ventilation sont produits par le flux d'air traversant l'ensemble du système (pulseur, conduits, aérateurs), et varient en fonction des réglages (vitesse, orientation, température...).

#### 2.4.2.7 | Chargement

Le chargement correspond à un son large-bande, légèrement fluctuant en niveau, MF ou HF et plutôt directif. Un écoulement très turbulent crée un effet de flap-flap, aussi appelé bouffées de chargement. Un écoulement très directif se rapproche d'un son de « fuite ». Le chargement est un son habituel, assimilable à un bruit de fond, son niveau sonore augmente avec la vitesse, ce qui le rend inconfortable et impacte l'intelligibilité. Dominant à partir de 80 km/h, il est produit par l'aspect turbulent de l'écoulement d'air qui va exciter des parties de la caisse comme le rétroviseur, les montants de baie ou le pare-brise.

#### 2.4.2.8 | Hullement

Le hullement est un son composé de deux raies proches dont la modulation d'amplitude rapide ( $\approx 10$  Hz) est très semblable à un battement et évolue en fonction de la vitesse du véhicule. Sa fréquence peut varier d'un hullement à un autre, allant des BF aux MF. Son émergence par rapport à un bruit large-bande (roulement) le rend identifiable. Il est généralement constant en fréquence, amplitude, et vitesse de modulation dans le temps, ce qui le rend fatigant. Le hullement est produit par la résonance d'une cavité, celle du pneu.

### 2.4.2.9 | Fuite

La fuite est un son très fluctuant et très directif, plutôt HF et élevé en niveau. Il est émergent par rapport au bruit de fond et se caractérise par des bandes fines plutôt que par un bruit-large bande. Il est notamment plus directif, plus HF et plus identifiable mais moins fluctuant que le « chargement ». Il est plus fluctuant et tout aussi émergent que le « sifflement aéroacoustique ». Sa directivité et son niveau, qui varient en fonction de la vitesse du véhicule, sont des aspects dérangement voire inacceptables de la fuite. Ils nuisent à l'effet cocon ainsi qu'au caractère rassurant et isolant de l'habitacle. La fuite est un son d'origine aérodynamique dû à un défaut d'étanchéité.

### 2.4.2.10 | Hachage

Le hachage est un son HF voire très HF ce qui le rend parfois inaudible, faible en niveau, et émergent. Il a un caractère harmonique-moteur ou inversement harmonique-moteur ; sa hauteur peut diminuer lorsque le régime moteur augmente. Il peut être assez gênant, voire perçant ou très pénible, suivant les réglages et la sensibilité auditive de chacun. Il est comparable au son des lignes de haute tension électriques ou aux acouphènes. Le hachage est un son d'origine électromagnétique qui correspond au découpage de la tension continue en signal sinusoïdal.

### 2.4.2.11 | Tambourinement

Le tambourinement est un son mal défini, très BF, élevé en niveau et très erratique. On dit qu'il tabasse, comme un tambour ou des timbales. Le tambourinement correspond à un « roulement » dont le caractère sourd a été poussé à l'extrême avec une composante percutante très erratique. Il provoque une gêne physiologique et une sensation de surpression des oreilles rappelant le « bourdonnement ». Le tambourinement apparaît à faible vitesse sur une plage bien définie entre 10-50 km/h et sur certains types de pistes avec des sollicitations particulières (ex : mauvaises routes, route pavée).

## 2.4.3 | Résumé de la section

### Validation et finalisation

Une fois la première version du lexique construite, celle-ci est soumise aux experts déjà sollicités pour les interviews. Un retour quantitatif leur est demandé, ce qui permet d'évaluer numériquement la qualité des différents éléments de notre lexique. Le premier élément à évaluer, les définitions aux travers des modalités de description : causale, réduite et hédonique. Le deuxième élément : la qualité des exemples sonores pour illustrer le terme défini. Le troisième élément évalué est la pertinence du terme dans le lexique final. À la suite de ce retour quantitatif, un retour qualitatif est aussi obtenu pour avoir les moyens d'améliorer le lexique. Ce retour qualitatif est mené sous la forme de discussions sur les points ayant été évalués négativement par les participants.

À l'issue de ces échanges, quelques modifications sont faites, telles que le remplacement du terme « sifflement », par deux termes « sifflement électrique » et « sifflement aérodynamique » pour éviter la confusion entre les deux, ainsi que la modification de certains extraits sonores peu pertinents. La version finale du lexique obtenue à la suite de ces retours est alors disponible en ligne, en français et en anglais, suivant ce lien : <https://speak.ircam.fr/lexique/ev-descriptor/>

## 2.5 | Conclusion de chapitre

Ce chapitre démontre la possibilité et l'intérêt de construire un lexique autour d'un environnement sonore nouveau. La contribution originale de ce travail est le schéma complet autour de la création de ce lexique qui passe par trois points clefs. Le premier : la collecte de descriptions des sons obtenus et les éléments contextuels autour de leurs apparitions, à partir d'interviews semi-directives. Concrètement, cela a consisté en 12 interviews d'une durée de une heure avec des experts du domaine de l'acoustique automobile. Le second : la construction du lexique. En l'occurrence, 11 termes, qui seront les entrées de notre lexique sont alors identifiés dans la liste des termes. Chacun de ces termes est ensuite défini et illustré. Enfin : la validation de ce lexique a été réalisée par le même panel d'experts ayant contribué à sa création. Ceci est réalisé par un retour quantitatif, en répondant à un questionnaire fermé envoyé par mail, et un retour qualitatif en discutant autour du questionnaire rempli.

Ce lexique a pour vocation de décrire la perception des différents sons dans l'habitacle d'un véhicule électrique à l'aide d'un outil naturel : le langage. Ce lexique retranscrit les jugements portés sur ces sons à travers une description contextuelle. Il permet ainsi de mieux appréhender les attentes généralement formulées vis-à-vis d'un environnement sonore, sur la base d'une description fondamentale. Ces jugements et attentes étant susceptibles d'évoluer dans le temps, ce lexique n'a pas vocation à être figé et pourra être amené à évoluer. Le facteur humain est un élément important de cette méthodologie, mais qui se rend aussi responsable de cette variabilité. La principale difficulté réside donc dans l'adoption d'une approche à la fois répétable et objective, permettant de répliquer ce travail si nécessaire.

Un élément clef de ce type de lexique est aussi d'être rendu accessible au plus grand nombre. Ce point est atteint en s'assurant que les définitions soient les plus claires et les plus complètes possible, au travers d'extraits sonores et définitions minutieusement choisies, qui ont été reformulées en fonction des retours d'experts du domaine. D'un point de vue technique, cette accessibilité est garantie par le partage du lexique construit ici sur l'environnement *Speak* dédié,<sup>7</sup> hébergé sur les serveurs de l'IRCAM. Ainsi, la construction du lexique s'insère dans un mouvement de verbalisation du son plus large que la seule problématique du véhicule électrique ou de la composition sonore. En effet, l'environnement *Speak* contient une quantité croissante de nouveaux lexiques dédiés à la verbalisation du son dans différents domaines de l'acoustique.

D'autre part, ce type de réalisation, en plus d'atteindre les objectifs fixés de description du son d'un environnement précis, est un premier élément qui s'intègre dans un projet plus grand : la construction d'une nouvelle culture du son, spécifique au véhicule électrique et détachée du véhicule thermique. La généralisation de ces outils en dehors des métiers de l'acoustique pourrait ainsi amener l'écoute au cœur de l'utilisation du véhicule électrique. Le partage de ce lexique en interne, ainsi que les discussions informelles avec les différents experts interviewés, permet d'avoir un premier retour d'expérience. La version actuelle du lexique permet à des nouveaux arrivants dans le groupe Renault, de se familiariser rapidement avec les différents sons du véhicule et une partie du jargon technique.

Cette culture du son et de l'écoute, en pleine mutation dans le secteur automobile, ouvre la voie à la création de nouveaux environnements sonores. Cela offre des possibilités infinies concernant les sons susceptibles d'être intégrés à cet environnement. Une nouvelle population serait alors progressivement amenée à prendre une place de plus en plus importante dans le monde de l'automobile : les compositeurs et les designers sonores.

---

<sup>7</sup> <https://speak.ircam.fr/>





## 3. Création de sons ajoutés pour l'acoustique augmentée

*« Tout d'abord, l'acte de design n'est pas un flux incessant d'événements créatifs ; c'est un travail hautement organisé et structuré, qui est seulement occasionnellement interrompu par de soudaines idées et intuitions. Ensuite, on ne peut pas tenir pour acquis que ces rares événements créatifs sont nécessairement mystérieux et au-delà de toute compréhension » [117]*

### 3.1 | Introduction de chapitre

Le travail de description présenté dans le chapitre précédent met en évidence que, malgré la diminution du niveau sonore dans l'habitacle des véhicules électriques, le son conserve une place centrale dans l'expérience perceptive de l'utilisateur, comme en témoigne la richesse des descripteurs sonores identifiés. Ce lexique démontre également que cette réduction de niveau sonore n'a pas éliminé tous les sons indésirables générateurs d'inconfort. En revanche, elle a libéré un espace sonore réel au sein de l'habitacle. La définition du confort proposée dans le chapitre dédié à l'état de l'art (Chap. 1) suggère que l'ajout de sons pourrait contribuer positivement au confort, en introduisant et en définissant la notion d'acoustique augmentée.

L'acoustique augmentée nécessite cependant de passer par un processus de design sonore et de création, guidé par une intention : celle d'améliorer le confort sonore. La question centrale de ce chapitre est la suivante : peut-on guider le design de l'acoustique augmentée afin de créer des sons améliorant le confort ? Nous verrons alors premièrement, la manière dont le contexte a été appréhendé à partir du travail d'interviews précédent, permettant d'identifier plusieurs situations sonores dans lesquelles le concept d'acoustique augmentée est le plus pertinent. Deuxièmement, Nous présenterons trois typologies de masquage pour réaliser cette augmentation de la scène acoustique, chacune présentant des aspects et des fonctions différentes. Enfin, le synthétiseur développé pour concrétiser cette approche est décrit, ainsi que six masques sonores conçus dans ce cadre — deux pour chaque typologie.

## 3.2 | Choisir un environnement sonore

### 3.2.1 | Interview : partie sur le confort

#### 3.2.1.1 | Les limites du lexique construit

Le chapitre (Chap. 2) précédent nous permet d'obtenir une description de la scène sonore de l'habitacle du véhicule électrique. Cette description a été construite, à dessein, pour l'identification et la description verbale de sources sonores dans l'habitacle du véhicule. Ce type de description, en plus d'être révélatrice de ce qui est perçu par les individus indépendamment de ce qu'on est capable de mesurer, nous permet d'avoir plusieurs niveaux de description du son : réduit, causal et hédonique.

En revanche ce lexique et ces descriptions ne permettent pas à eux seuls de déterminer la manière de les traiter. En effet il manque à cette description de nombreux éléments contextuels. Ces éléments peuvent être très dépendants du modèle de véhicule sur lequel on travaille, des cas d'utilisation, des technologies existantes ou même des stratégies de marketing de l'industriel. De ce fait, même si le lexique remplit correctement son rôle et permet de décrire et reconnaître un Roulement par exemple, la description seule ne permet pas de savoir quels éléments rendent la scène inconfortable dans un contexte. Ainsi dans un retour du type « *Le sirènement sur le véhicule XY est particulièrement gênant en phase de décélération, aux alentours de 50 km/h* », indique les éléments contextuels qui nous donnent les pistes pour améliorer le confort, et c'est le lexique qui nous permet de nous assurer que le son responsable est bien un sirènement. La question reste donc entière d'un point de vue pratique quant à la manière d'améliorer le confort sonore à l'aide du design, mais nous disposons d'un outil supplémentaire pour arriver à nos fins à travers ce lexique.

#### 3.2.1.2 | Le lexique comme outil pour s'approprier le contexte

C'est donc pour cela qu'une partie des échanges des interviews menées est dédiée aux questions relatives au confort et au contexte autour des sons abordés, les questions rédigées pour guider l'interview sont reportées tableau 3.1. A partir de ces échanges, on est capable de développer cette dimension contextuelle une fois le lexique abouti et d'avoir des éléments pour mieux comprendre comment traiter un son dans un contexte. Cette deuxième partie du questionnaire permet une verbalisation libre autour du sujet du confort. Contrairement à la première partie où l'interviewer était assez présent, au travers de relances, pour aider le participant à verbaliser au maximum autour de la description du son, cette partie est plus librement guidé par le participant. On remarque en particulier des questions plus longues et plus complexes, faisant ici exception aux consignes de rédaction d'un questionnaire semi-directif évoqué dans le précédent chapitre. Cette exception s'explique par la nécessité d'aborder un sujet nouveau et complexe dans l'interview : l'ajout de sons pour améliorer le confort.

Ces détails contextuels ajoutés ne font en revanche pas à proprement parler partie du lexique puisqu'ils ne permettent pas la reconnaissance ou la description des sons. Ces détails proviennent d'échanges et questions spécifiques concernant l'amélioration du confort sonore et l'ajout de sons dans l'habitacle. Une analyse sémantique des échanges, comme ce qui a été réalisé dans le chapitre précédent est alors difficilement applicable, la raison étant que ces informations fortement contextuelles sont difficilement réductible en des concepts ou des d'occurrences. Par conséquent des extraits des échanges, représentant une idée entendue plusieurs fois, et appuyant un point avancé son isolés. Par exemple, on peut noter la différence de jugement qu'il sera fait d'un son de ventilation en fonction du confort thermique de l'utilisateur, de l'efficacité du système ou du niveau sonore dans l'habitacle de la voiture. Ainsi l'utilisateur priorisera plutôt le confort thermique que le confort acoustique, et considérera plutôt une

	Questions
Q7	Selon vous quel sera la place du confort sonore dans votre travail ?
Q8	Comment le définiriez-vous ?
Q9	Y a-t-il un son de l'habitacle en particulier que vous qualifieriez comme inconfortable ? Un autre confortable ?
Q10	Pourrions-nous désigner un son qui serait diffusé par le système audio de la voiture qui viendrait corriger le caractère inconfortable que vous venez d'identifier sans dénaturer la nature du son, de telle sorte que l'utilisateur rattache toujours ce son à son origine/son utilité ?

**Table 3.1.** Deuxième partie du questionnaire, plus ouverte, sur la notion du confort et de l'amélioration du confort.

ventilation bruyante comme plus efficace :

**MD :**

*« Potentiellement, on va être moins critique sur un bruit de dégivrage parce qu'on a mis la ventile à fond et que la seule chose qu'on attend c'est que ça dégivre vite, alors que si on est en position confort et qu'on a juste envie d'avoir une température constante mais pas d'être gêné par le bruit là on va être plus critique sur le bruit. »*

...

**MD :**

*« Finalement pour que tu aies une impression d'efficacité, il faut que tu l'entendes [NdA : La ventilation]. Si tu mets ta ventile, que tu entends rien, tu as l'impression qu'elle ne marche pas. Donc ce flux, t'as besoin de l'entendre. Par contre il faut qu'il soit maîtrisé dans sa sonorité et qu'il ne soit pas ... enfin qu'il soit plutôt rond quoi. »*

### 3.2.2 | Synthèse des éléments contextuels

L'ensemble des commentaires et précisions faites est difficilement exploitable tel quel. Pour faciliter l'exploitation de ces informations primordiales, une session de travail est effectuée avec Louis-Ferdinand Pardo et Eléonore Vialatte. À partir des interviews, du lexique et de l'expertise sur les sons mentionnés, les dimensions contextuelles les plus importantes en vue d'une « augmentation » de la scène sonore sont identifiées dans cette session. L'augmentation réside alors dans le fait d'utiliser ce concept d'acoustique augmentée, et donc de concevoir un design sonore à ajouter à cette scène.

Le tableau 3.2 complète ces informations en résumant les réponses aux questions suivantes :

— **Le son est-il pilotable ?** On parlera de son pilotable si l'on peut obtenir des paramètres du véhicule qui nous permettent de prévoir le comportement du son. Par exemple le sirènement ou sifflement électrique d'un moteur font l'objet de modélisation et mesures lors de la conception et l'industrialisation d'un véhicule. De ce fait, en connaissant la vitesse du véhicule, l'enfoncement de la pédale d'accélération ainsi que quelques autres paramètres techniques comme la réponse vibratoire du bloc-moteur ou le nombre de paires de pôles du moteur électrique, l'industriel est capable de simuler la signature sonore qu'aura ce sirènement. On parle alors de son pilotable puisque l'on peut anticiper le comportement du son à travers des paramètres accessibles. A l'inverse, un son typiquement non-pilotable est le chargement aérodynamique, pour lequel quelques facteurs sont connus, comme la vitesse ou la surface des vitres, mais le son dépend aussi grandement d'autres paramètres non contrôlable, comme la direction du vent ou les turbulences présentes dans l'air au moment du passage du véhicule.

- **L'ajout d'un masque est-il envisageable ?** Cette question traite de l'avis parmi les témoignages recueillis, quant à l'acceptabilité d'un son ajouté pour « augmenter » le son problématique à traiter. Cet avis est recueilli à priori avec une brève description du concept, mais sans illustration ou cas concret où cela est appliqué. Cela permet de s'assurer que les premières personnes concernées, à savoir les professionnels de l'automobile trouvent ce son ajouté acceptable. Par exemple le Chargement est défini comme « *Un son habituel, assimilable à un bruit de fond, dont le niveau sonore augmente avec la vitesse, ce qui le rend inconfortable et impacte l'intelligibilité.* ». De par sa définition, mais aussi des retours obtenus des spécialistes, on en déduit qu'il se prête mal à cette méthodologie d'acoustique augmentée. En effet, les retours recueillis suggèrent que le son lui-même est accepté et attendu mais c'est son niveau sonore qui est problématique. Par conséquent, ajouter un son ne fera qu'augmenter le niveau mais aura peu de chance de changer son acceptabilité puisque ce dernier est déjà accepté dans une certaine mesure.
- **Une solution technique existe-t-elle ?** Contrairement aux deux premières questions, cette information ne résulte pas en une différence de traitement. S'il existe ou non une solution technique ne change pas la manière dont le son ajouté sera conçu. En revanche elle permet de classer par ordre d'importance les sons sur lesquels travailler. Par exemple le roulement, ou tambourinement que l'on retrouve dans le lexique peut voir sa partie la plus basse fréquence traitée à l'aide du RNC [4, 118]. Bien que cette technologie ne soit pas universellement répandue à cause de son coût et ses limitations techniques, elle change les attentes en terme de confort sur le son en question et sur les manières de traiter ce son. De tels sons ne seront donc vraisemblablement pas traités par acoustique augmentée.
- **Le son est-il accepté ?** Cela revient à savoir si le son est cohérent avec les attentes de l'utilisateur et si l'existence même du son ou ses caractéristiques acoustiques posent un problème. On reporte la réponse à cette question dans le tableau suivant par un oui ou un non, mais cette réponse est la plus susceptible de varier en fonction du contexte et des situations. Un son de sirènement peut être acceptable dans un contexte, par exemple sur une bretelle d'insertion, où ce son traduit une montée en régime du moteur et est rapidement masqué par d'autres sources sonore liées à la vitesse du véhicule. Il peut devenir non accepté dans le même véhicule dans un contexte de conduite différent, tel que des embouteillages avec une accélération et freinage successifs faisant apparaître ce sirènement de manière intermittente. Cela rappelle que ce tableau peut évoluer en fonction des cas d'étude et est seulement une des manières d'intégrer le concept avec le lexique.

Les réponses à ces quelques questions nous permettent, une fois le son décrit et identifié, de guider le design sonore. En effet la question de l'aspect et la fonction du son ajouté est plus facile à cerner.

### 3.2.3 | Identification de situations sonores propices à l'acoustique augmentée

A partir de ces éléments contextuels supplémentaires il est désormais possible d'identifier en plus des objets sonores sur lesquels travailler, un contexte et un cas d'utilisation qui leurs sont liés. Nous identifions donc des « situations sonores » caractérisées par un ou plusieurs « objets » sonores à l'origine d'un ressenti négatif ainsi qu'un cas d'utilisation particulier. On liste trois situations sonores d'intérêt pour lesquelles l'acoustique augmentée peut répondre à une absence de confort.

#### 3.2.3.1 | Situation sonore 1

Sons : Ventilation + Sifflement aérodynamique. Cette situation correspond à un cas d'utilisation où le conducteur est passif dans l'habitacle, voiture à l'arrêt et sans autre source sonore que la ventilation, qui possède un défaut aéroacoustique. La scène sonore est donc constituée de 2 sons : un son de ventilation et un sifflement aéroacoustique. Dans cette situation, le son de ventilation est habituellement considéré

Nom	Pilotable	Masquable	Accepté	Attributs	Solution	
Roulement	En partie	Non	Oui	Niveau, Large-bande, BF, Régulier	RNC	
Sirènement	Oui	Oui	En partie	Fluctuant, MF, Cohérent, Discontinu, Pénible	Non	
Sifflement Aéroacoustique	En partie	Oui	Non	Aigu, Émergent, Directif, Stable, Désagréable	Oui	
Sifflement Électrique	Oui	Oui	En partie	HF, Strident, Localisable, Gênant, Pénible, Cohérent	Non	
Bourdonnement	En partie	Non	Non	BF, Niveau	RNC	
Ventilation	Oui	Non	Oui	Agressivité, Large-bande, Cocon, Niveau	Non	
Chargement	En partie	Oui	En partie	Oui	Niveau, MF, Fluctuant, Habituel	Non
Hululement	En partie	Oui	Non	Constant, Modulant, BF, Émergent, Identifiable, Fatigant	Non	
Fuite	Non	Oui	Non	Émergent, Fluctuant, Directif, Cocon, Étanchéité	Oui	
Hachage	Oui	Oui	Non	très HF, Harmonique-moteur, Pénible, Percant	Non	
Tambourinement	Non	Non	Non	Niveau, Erratique, BF	RNC?	

**Table 3.2.** Résumé des informations contextuelles supplémentaires aidant à la conception de l'acoustique augmentée.

comme acceptable et attendu, le sifflement comme inacceptable et désagréable. L'enregistrement utilisé dans la suite pour cette situation est disponible sur la plateforme d'écoute SoundCloud [ici](#).

### 3.2.3.2 | Situation sonore 2

Sons : Sirènement électrique + Roulement. Cette situation correspond à un utilisateur en phase d'accélération après un démarrage, le sirènement électrique étant causé par le couple du moteur électrique nécessaire à l'accélération. Dans cette situation le comportement de la voiture est donc dynamique et l'utilisateur actif. Dans cette situation les deux sons présents sont donc le roulement et le sirènement électrique du moteur électrique, chacun évoluant avec la vitesse du véhicule. Le sirènement est décrit comme harmonique-moteur avec une hauteur proportionnelle à la vitesse du véhicule. Le roulement quant à lui est décrit comme ayant seulement le niveau qui augmente avec la vitesse. Le sirènement est ici le son « défaut » bien que son caractère ne soit pas aussi inacceptable que pour le sifflement aérodynamique de la situation sonore 1.

### 3.2.3.3 | Situation sonore 3

Sons : Sifflement électrique + Chargement. Cette situation correspond à un utilisateur conduisant la voiture à vitesse constante sur autoroute. Pour cette vitesse, le son aérodynamique de Chargement est habituellement prédominant, auquel peut se superposer un sifflement électrique dû au fonctionnement du moteur. Le son de Chargement est susceptible de varier de manière aléatoire alors que le sifflement électrique sera plutôt constant. Dans cette situation, le son de chargement est plutôt accepté et cohérent avec la situation, bien que certaines caractéristiques de celui-ci, telles qu'une trop grande irrégularité ou un niveau trop élevé puissent le rendre désagréable, le sifflement est lui susceptible d'être inconfortable de par son caractère émergent, aigu et constant.

Dans la suite nous travaillerons avec la situation sonore la plus simple : la situation sonore 1 constituée d'un son de ventilation et d'un son de sifflement aéroacoustique créé par un défaut dans un conduit du système d'aération. Cette situation est stationnaire, ce qui facilite la conception et le design de masque, et a été observée et enregistrée en situation réelle dans un véhicule électrique auquel nous avons facilement accès. Ceci a permis d'enregistrer différents paramètres de ventilation ainsi que de pouvoir réécouter le défaut en situation réelle.

## 3.2.4 | Résumé de la section

### Choisir un environnement sonore

La première étape de ce travail dédié à l'acoustique augmentée consiste à sélectionner un environnement sonore associé à un cas d'utilisation précis, que nous appelons « situation sonore ». Cette situation sonore inclut, en plus des informations acoustiques, des informations contextuelles, relatives à l'apparition des bruits. Ces informations contextuelles sont déduites des interviews semi-directives. Dans la suite du manuscrit, une seule situation sonore sera étudiée : celle constituée d'un son de ventilation avec un sifflement, dans une situation de véhicule à l'arrêt ou l'utilisateur est passif dans l'habitacle.

## 3.3 | Typologies de masquage

Ce paragraphe détaille la méthodologie suivie pour la création de sons créés, appelés masques, que nous appliquerons à une situation sonore avec un contexte et des objets sonores précisés dans la section précédente. Il est essentiel d'être à même de caractériser la fonction et l'aspect que ces masques devront avoir. Pour cela nous détaillons dans la suite comment nous allons travailler avec 3 typologies de masquage appelées « Énergétique », « Attentionnel » et « Habillage », ayant chacun un aspect et une fonction différente.

### 3.3.1 | Fonction du masque

#### 3.3.1.1 | Masque Énergétique

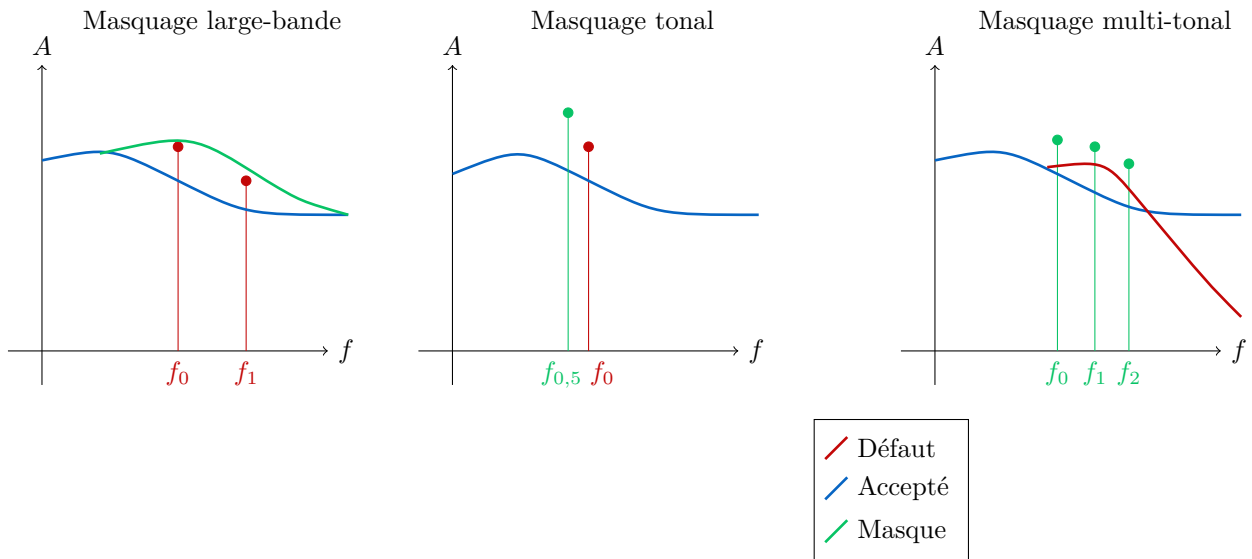
Ce type de masque a pour fonction de faire disparaître un objet sonore, jugé indésirable dans une situation sonore. Cette idée est inspirée des travaux de Zwicker et Fastl [119] dont le fonctionnement est aussi détaillé par Moore [120] et se rapproche de l'expérience présentée dans l'étude menée par Brocolini et al., mentionnée plus tôt [43]. Le terme de masquage énergétique est notamment utilisé dans des études d'intelligibilité et de compréhension de la parole dans le bruit [121, 122]. Ces études donnent bien le même nom pour parler du même phénomène, à savoir l'ajout d'énergie dans la même bande critique que le signal d'intérêt, et étudient alors l'impact du masque sur l'intelligibilité. Ici le signal à masquer n'est pas un signal de parole, mais plutôt un signal tonal, de fréquence et niveau constant, tel que présenté dans la situation sonore 1. On s'attend alors à retrouver ici les résultats connus sur la détection d'un signal masqué par un signal situé dans une zone fréquentielle proche.

#### 3.3.1.2 | Masque Habillage

Ce type de masque a pour fonction d'enrichir le son jugé indésirable, non pas pour le rendre imperceptible, mais pour modifier les propriétés acoustiques de celui-ci. Il n'est pas possible de modifier directement la source du son, puisque celle-ci n'est pas contrôlée, il est en revanche possible d'ajouter un son complétant le spectre du son indésirable. Il en résulterait alors une modification des propriétés réduites de ce son. On peut retrouver dans la littérature des dispositifs qui rentreraient ici dans la typologie de masquage Habillage. On peut citer par exemple l'ajout d'harmoniques ou de sous-harmoniques pour les moteurs thermiques [123, 124] que l'on peut retrouver dans des technologies comme le *Engine Sound Enhancement* (ESE) qui imite, dans un véhicule avec un moteur de taille réduite, la signature acoustique de moteurs réputés sportifs ou confortables. Le terme « habillage » ou « masquage par habillage » n'est pas mentionné tel quel.

#### 3.3.1.3 | Masque Attentionnel

Ce type de masque a pour fonction de détourner l'attention d'un son jugé indésirable. Le masque en question a donc pour vocation à être entendu comme un objet à part entière et supplémentaire dans la scène sonore. Ce type de masquage est librement inspiré de plusieurs thématiques de recherche, notamment les études sur la parole dans du bruit qui traitent alors de « masquage informationnel » [121, 125]. Ce masquage informationnel étudie l'effet d'une voix sur l'intelligibilité d'une autre voix. La voix ajoutée parasite alors la compréhension de la parole sans effectuer de masquage énergétique, ce qui nous permet de penser qu'il est alors possible d'effectuer un masquage similaire sur des signaux qui ne sont pas des voix. Le terme masquage informationnel n'a pas été repris puisqu'il est plutôt destiné à l'intelligibilité, ce qui l'écarte de notre sujet, on lui préfère le terme « masquage attentionnel », comme on peut le retrouver pour caractériser des installations sonore dans des paysages sonores urbains [126].



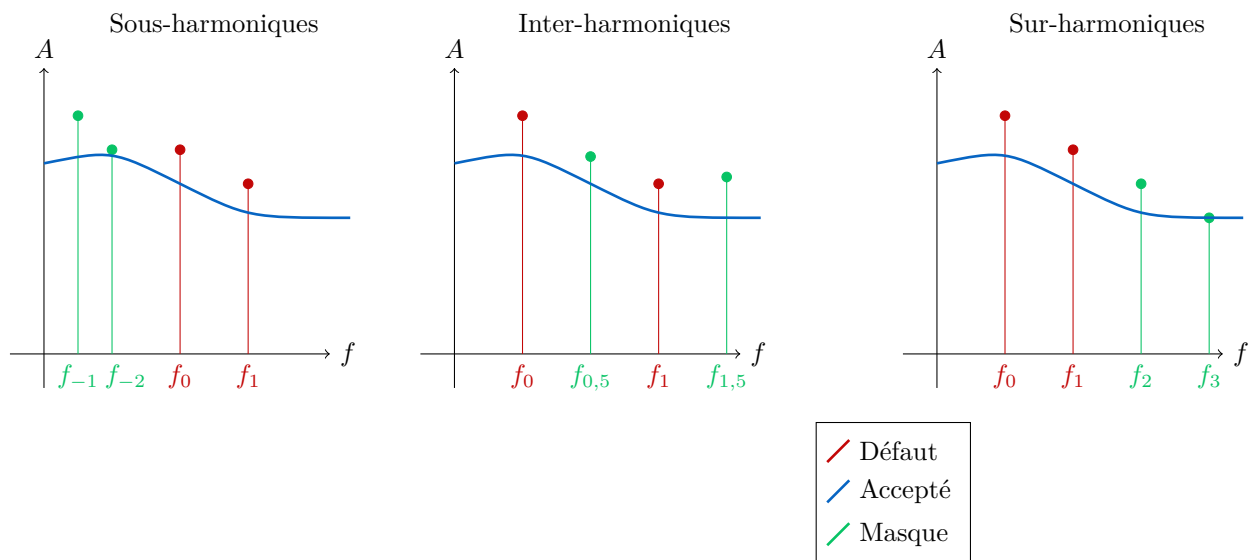
**Figure 3.1.** Différentes approches possibles pour la typologie Énergétique pour différents types de défauts. On observe que le niveau du masque est systématiquement plus élevé et plus grave pour efficacement masquer un signal indésirable. Le caractère tonal ou large-bande du masque sera plutôt guidé par l'aspect souhaité par le designer.

### 3.3.2 | Aspect du masque

#### 3.3.2.1 | Masque Énergétique

L'aspect du masque énergétique est nécessairement guidé par l'aspect du son indésirable. Pour parvenir à le masquer, une manière de procéder et de donner au masque une intensité sonore supérieure au son indésirable dans chaque bande de Bark. Ceci fixe une enveloppe fréquentielle de niveau minimum en fonction du son indésirable [119]. Dans le cas de la situation sonore 1, le sifflement est caractérisé par un partiel unique, et donc le masquage n'a besoin d'être calculé que par rapport à une seule bande de Bark centrée sur cette émergence.

L'aspect de ce masque est aussi guidé par les sons acceptés de l'habitable puisque ces sons déjà existants peuvent masquer une partie du spectre du son indésirable. La fonction de ce masque est de faire disparaître un objet sonore indésirable, le risque est cependant que le masque soit lui-même perçu comme indésirable. De ce fait le masque ajouté a vocation à se fondre dans les sons acceptés : on cherche ainsi à éviter un son qui pourrait être perçu comme un nouvel objet sonore. Pour effectuer un masquage énergétique, les travaux de Zwicker et Moore [119, 120] nous permettent d'identifier différentes sous-stratégies de cette typologie, figure 3.1. En fonction du défaut à masquer on peut imaginer procéder à un masquage par un son large-bande comme sur la figure de gauche, ou par un son harmonique comme pour la figure du centre et de droite. Le choix de cette sous-stratégie peut aussi être guidée par le son accepté, en choisissant un masque qui se fond le mieux avec ce dernier. Dans notre cas, la situation sonore 1 contient un son accepté très large bande et constant, et un défaut émergent avec un seul partiel, on adopte alors une sous-stratégie par masquage large-bande (Fig. 3.1 gauche). Le masque aura alors la capacité à effectivement ajouter de l'énergie dans la même bande critique de Bark que le défaut, pour le masquer, tout en ayant un aspect similaire au son accepté.



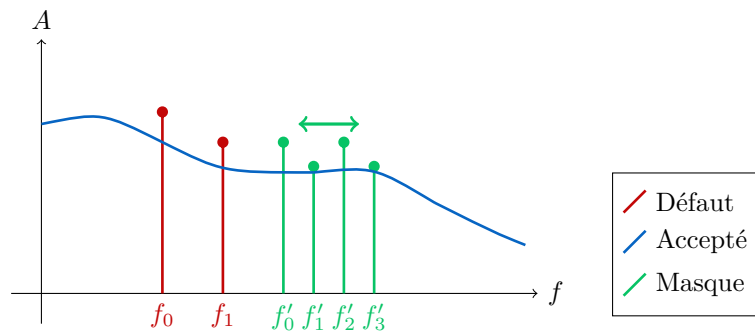
**Figure 3.2.** Différentes approches possibles pour la typologie Habillage, chacune consistant à inclure le son défaut comme un partiel d'un son global différent. A gauche : la sous-stratégie sous-harmoniques, consiste à ajouter des partiels plus grave, pour inclure le défaut dans un son harmonique plus mat. Au centre : la sous-stratégie inter-harmonique consiste à ajouter des partiels intermédiaire, rappelant les spectres de moteurs thermiques. à droite, la stratégie consiste à enrichir le son en lui ajoutant des harmoniques.

### 3.3.2.2 | Masque Habillage

L'aspect de ce masque est très fortement guidé par le son indésirable à améliorer. Cependant la manière de l'améliorer est très riche. Pour ce type de masque il a été choisi de s'inspirer de choix de design sonore ou d'ajout de sons déjà existants dans l'automobile avec l'ajout d'harmoniques. Le but étant de rester le plus proche possible de l'univers sonore auquel pourrait s'attendre un utilisateur. L'aspect est donc guidé par une volonté de garder des sons au spectre assez harmonique, proche des sonorités de fonctionnement des moteurs électriques ou thermiques. Dans cette typologie de masquage, le masque vient se mélanger au son indésirable : le nombre d'objets sonores de la situation n'est donc pas censé évoluer. La figure 3.2 schématise alors différentes sous-stratégies possible dans la typologie habillage. La stratégie sous-harmonique, représentée à gauche sur la figure, se prête bien à notre situation puisque le défaut est constitué d'une première harmonique  $f_1$  avec une fondamentale  $f_0$  indiquées en rouge sur la figure 3.2. La sous-stratégie inter-harmonique correspond à ce qui est fait dans le cas de l'ESE de moteurs thermiques, où des harmoniques impaires peuvent être ajoutées à un son de moteur qui produit initialement des harmoniques paires. Ces harmoniques impaires sont souvent le résultat d'une architecture de moteurs comportant 3, 6 ou 12 cylindres, souvent connotés comme sportif ou confortable. On peut ainsi prédire précisément les signatures harmoniques caractéristiques en fonction de l'architecture complète d'un moteur, du nombre de cylindre, du cycle d'allumage des pistons etc..<sup>1</sup> Enfin la sous stratégie en sur-harmoniques se prêtera mieux pour enrichir un son basse fréquence pauvre en harmoniques.

Un autre choix de son ajouté que l'on retrouve régulièrement dans le véhicule électrique est l'utilisation de l'illusion de Shepard-Risset. Ce type d'illusion donne l'impression d'un son de hauteur croissante indéfiniment. Elle est notamment utilisée pour sonifier un comportement dynamique rappelant la montée harmonique d'un moteur thermique [127]. Ce type de design semble le plus adapté pour des sons liés intuitivement au groupe moteur, comme le Sirènement, Hachage ou Sifflement Électrique.

<sup>1</sup> <https://architectures-et-sons-de-moteurs.blog4ever.com/>



**Figure 3.3.** Illustration du concept exploité pour les masques attentionnels. Le masque est présenté avec une plus grande richesse harmonique, un centroïde spectral plus élevé et une modulation, pour représenter les possibilités pour attirer l'attention

### 3.3.2.3 | Masque Attentionnel

L'aspect de ce masque est par conception le plus libre, puisqu'ici la typologie a vocation à créer un objet sonore à part entière. Ce nouvel objet sonore est pensé pour attirer l'attention et la détourner d'un son indésirable. Les études sur la capture attentionnelle montrent que la saillance d'un stimulus, présenté dans un corpus, est augmentée par la variation de paramètres tels que la brillance, la rugosité [128] ou les variations temporelles des pulsations d'un son [129]. Les masques attentionnels sont alors construits sur ce principe, avec des paramètres de brillance, rugosité ou un comportement temporel très différent de la scène dans laquelle il est présenté. L'objectif est d'augmenter sa saillance, et donc sa capacité à capturer l'attention. La figure 3.3 illustre ce principe avec un masque, en vert, pour lequel les fréquences des partiels  $f'_0, f'_1, f'_2, f'_3$  n'ont pas de liens avec ceux du défaut  $f_0, f_1$  et n'effectue pas de masquage énergétique non plus. Ici il n'y a pas de sous-stratégies clairement identifiées, les possibilités sont par nature plus grandes. On sait aussi que dans un paysage sonore, une source identifiée clairement [130, 131] est plus susceptible d'avoir un impact sur la perception (agréabilité, niveau perçu ...) que si la source n'est pas identifiée. Par conséquent, les masques attentionnels sont conçus comme des objets sonores à part entière, ces masques n'essaient pas de se dissimuler dans la scène sonore, mais bel et bien d'être entendus comme un objet existant intentionnellement.

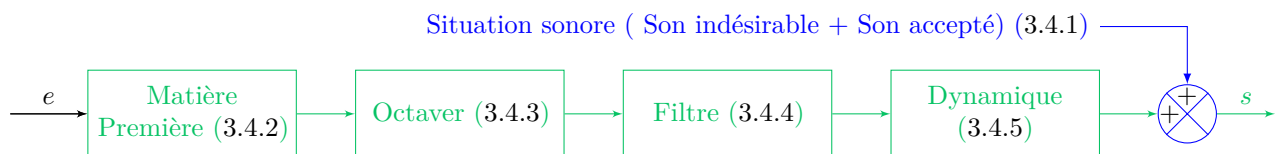
### 3.3.3 | Résumé de la section

#### Typologie de masquage

Pour réaliser notre solution d'acoustique augmentée, avec l'intention d'améliorer le confort sonore dans une scène présentant un défaut, trois typologies de masquage sont identifiées avec différents aspects et intentions. La première typologie, appelée masquage énergétique, a pour intention d'effacer la perception du défaut à l'aide de masques supérieurs en niveau dans les bandes critiques où le défaut existe, facilement accessible avec des bruits large-bande type bruit blanc. La deuxième typologie de masquage est appelée habillage et consiste à « compléter » le spectre du défaut, pour changer ses propriétés acoustiques, par exemple en ajoutant des harmoniques. La troisième typologie de masquage est appelée masquage attentionnel et consiste à ajouter un « objet sonore » dans la scène, avec ses propres spécificités, censé détourner l'attention du défaut.

## 3.4 | Construction d'un synthétiseur

Pour composer les masques souhaités il est nécessaire de construire son propre synthétiseur. Ceci s'explique par la particularité de l'approche que nous suivons ici puisque nous souhaitons travailler sur une situation sonore précise. La situation sur laquelle nous travaillons est la situation sonore 1 détaillée dans la section précédente (Sec. 3.2.3), constituée de deux objets sonores : un son de ventilation et un sifflement aérodynamique. Nous construisons ainsi notre synthétiseur à l'aide de l'environnement de programmation graphique Max-MSP.<sup>2</sup> Son but est de permettre d'obtenir les différents types de masquage pour les différentes situations sonores potentielles, bien que nous nous focalisons sur une seule situation dans ce chapitre. La figure 3.4 résume par un schéma bloc le fonctionnement de ce synthétiseur. Nous représentons deux flux distincts qui sont additionnés en bout de chaîne pour constituer le son final : un flux dédié entièrement à la synthèse du masque et un flux indépendant dédié à la reconstitution de la situation sonore.



**Figure 3.4.** Schéma-bloc résumant le fonctionnement du synthétiseur. Chaque bloc est détaillé dans une des sous-sections suivantes. L'entrée, en bleu, représente la situation sonore indépendante des blocs créant les sons, en vert.

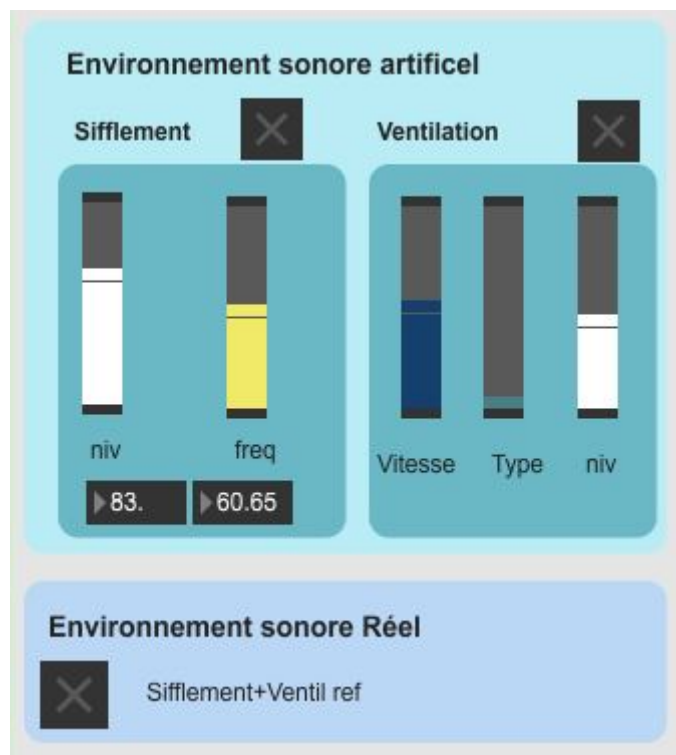
Nous orientons notre choix de synthèse vers une synthèse paramétrique [49]. Ce qui permet d'avoir un gradient de masques avec différents paramétrages, ou que l'on pourrait optimiser dans un deuxième temps selon plusieurs critères [132]. Pour garder trace des différents paramètres utilisés, le synthétiseur construit fonctionne par bloc en série, appliquant un effet après l'autre où chaque effet peut être modifié et appliqué indépendamment.

### 3.4.1 | Situation sonore modulable

Le flux audio dédié à la reconstitution sonore est indépendant du flux lié à la synthèse pour simuler l'environnement sonore de la voiture. Ce flux reconstitue la situation sonore 1 étudiée ici. Pour reconstituer cette scène deux options sont possibles. La première consiste à enregistrer une situation réelle en voiture et la rejouer telle quelle. Cette option a le mérite d'être la plus réaliste possible mais manque de contrôle possible des caractéristiques sonores du signal. La deuxième option consiste à recomposer les sources sonores après les avoir enregistrées ou synthétisées séparément. Cela permet d'avoir un grand contrôle des paramètres acoustiques au prix d'un réalisme diminué. Ce type de reconstruction est en partie possible grâce à la constitution d'une bibliothèque de sons, de différents types de véhicules électriques, accumulés lors de la construction du lexique (Chap. 2). Les sources étant ainsi séparées il est alors aisé de changer directement des paramètres acoustiques spécifiques de la situation sonore, tels que le niveau ou la hauteur du sifflement dans le cas de la situation considérée ici.

On représente figure 3.5 une capture de l'interface Max permettant ce paramétrage, dans laquelle le choix des deux options évoquées est possible grâce à des interrupteurs. On peut repérer la première option par l'interrupteur « Environnement sonore réel ». La deuxième option, plus souple, est nommée « Environnement sonore artificiel » et possède alors un interrupteur pour chaque source. La figure permet de voir les quelques paramètres choisis pour reconstituer la situation sonore artificiellement. Pour le sifflement, on a un paramètre pour la fréquence et un pour le niveau sonore. Pour la ventilation,

<sup>2</sup> <https://cycling74.com/products/max>



**Figure 3.5.** Capture d'écran du bloc dédié à la reconstitution d'une situation sonore sur le logiciel Max-MSP.

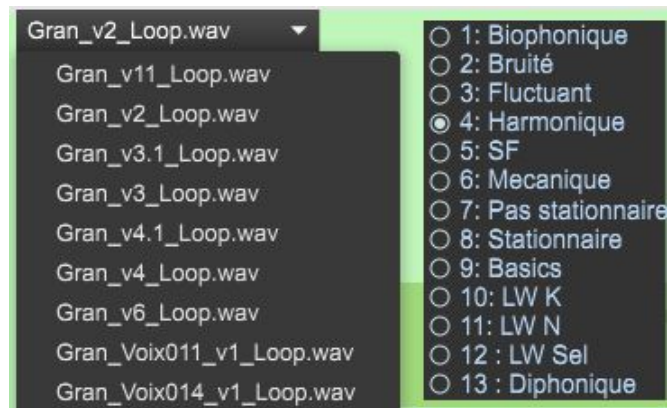
on trouve : un paramètre de niveau, un pour la vitesse de ventilation et un pour le type de ventilation, ce qui correspond à l'orientation donnée au flux d'air pulsé dans le véhicule. On remarque qu'un son synthétisé, ici le sifflement, est plus facilement guidé par des paramètres réduits alors qu'un son enregistré, ici la ventilation, est plus facilement guidé par des paramètres causaux.

### 3.4.2 | Matière première

Le deuxième bloc est orienté entièrement vers la synthèse de masque, le son additionnel à ajouter à la situation sonore identifiée. Il a été choisi de travailler à partir de sons élémentaires, créés pour l'occasion. La figure 3.6 affiche la manière dont on accède à ces matières premières via l'interface Max. Ces matières ont en partie été créées grâce à l'aide fournie par des personnes ayant une formation en design sonore. La principale créatrice de ces matières est Eléonore Vialatte, diplômée en design sonore de l'ESAD-TALM, lors d'un stage de 6 mois au sein de l'équipe PDS à l'IRCAM encadré par l'auteur principal.

Ces sons ont été conçus dès le départ pour se superposer au son de ventilation. Pour ce faire, plusieurs catégories de sons possédant un imaginaire ou des références communes ont été ciblées. Le tableau 3.3 reprend les catégories proposées ainsi que leur description. On peut retrouver ces catégories ainsi que les matières premières sonores composées dans la capture d'écran du synthétiseur (Fig. 3.6). Eléonore Vialatte a ainsi composé près d'une soixantaine de matières premières à exploiter. A celles-ci s'ajoutent une dizaine de matières premières composées par l'auteur ainsi que quelques-une fournies par le designer sonore Laurent Worms. Les matières premières composées et utilisées pour ce travail sont accessibles sur la plateforme SoundCloud, dans l'album *Matières premières* » en cliquant [ici](https://soundcloud.com/matthieu-duroyon-404568848/sets/matieres-premieres)<sup>3</sup>.

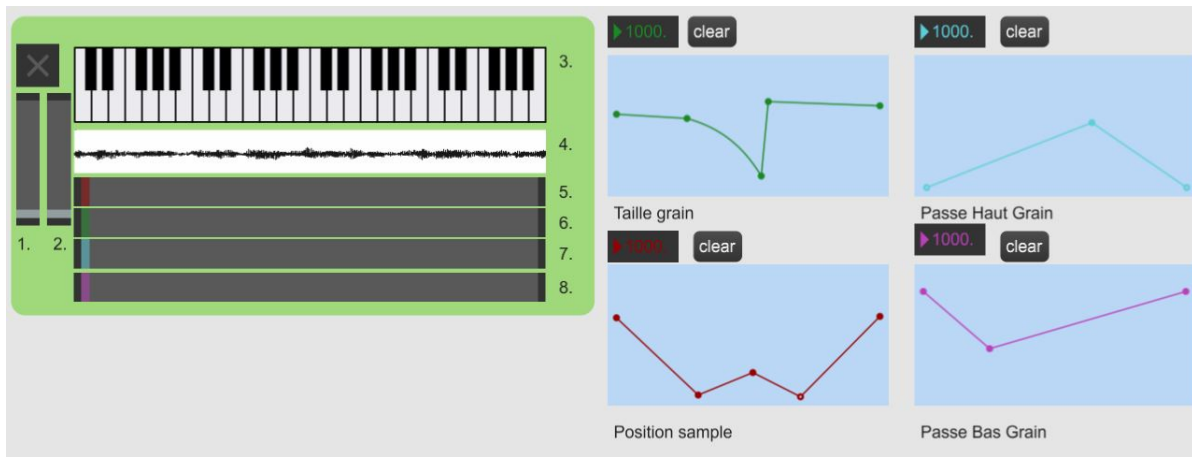
<sup>3</sup> Adresse complète : <https://soundcloud.com/matthieu-duroyon-404568848/sets/matieres-premieres>



**Figure 3.6.** Capture d'écran du bloc dédié à la sélection d'une matière première. La capture présente les différentes matières premières pour la catégorie nommée « Harmonique » par Eléonore Vialatte.

Axes de création	Description
Mécanique	Qui s'inspire de l'univers des machines (chaufferie, soufflerie, métal ou véhicules mécaniques type train).
Harmonique	Qui est composé d'une ou plusieurs tonalités, avec des harmoniques ou non et qui évolue de manière cyclique avec plus ou moins de variations dans le temps.
Biophonique	Qui fait référence à des sons d'environnements animal ou végétal (mare de grenouilles, insectes) ou évènement météorologiques (vent)
Bruité	Qui se rapproche des caractéristiques sonores de la ventilation.
Imaginaire (Sc.Fi.)	Qui renvoie au monde des machines futuristes, ou de l'espace.
Stationnaire	Qui ne fluctue pas ou très peu dans le temps et évolue de manière régulière. Constant en fréquence et/ou en niveau.
Fluctuant	Qui fluctue peu à beaucoup et qui évolue dans le temps de manière régulière ou irrégulière, en fréquence et/ou en niveau. Qui est plus riche.

**Table 3.3.** Description des catégories de matières premières de son telles que présentées par la designeuse sonore Eléonore Vialatte



**Figure 3.7.** Capture de l'interface Max pour l'utilisation du synthétiseur granulaire. Le curseur 1, correspond au  $x_1$  présenté sur le schéma et contrôle le sens de lecture de chacun des grains. Le curseur 2 contrôle la distribution temporelle des grains. Le curseur 3 contrôle la vitesse de lecture des grains et donc la hauteur. Le curseur 4 est une représentation visuelle de l'échantillon et des grains qui sont lus. Le curseur 5 contrôle la position du grain. Le curseur 6 contrôle la taille du grain et 7 le volume général. Sur la droite on peut voir les courbes d'automation permettant de commander dans le temps l'évolution de ces paramètres.

### 3.4.2.1 | Synthèse granulaire

Pour créer ces matières premières, un outil largement utilisé dans ce chapitre par l'auteur ainsi que par Eléonore Vialatte est la synthèse granulaire. Ce type de synthèse est choisi pour sa capacité à produire une variété de sons très riches à partir de peu d'échantillons, pouvant à la fois être très large-bande ou très harmonique. Cette synthèse, s'appuyant sur un principe connu depuis longtemps [133], a notamment été utilisée dans des œuvres musicales par Xenakis, qui propose ainsi *Analogique A* et *Analogique B*, deux compositions réalisées en 1959 utilisant un procédé de synthèse granulaire analogique. Cette synthèse est en particulier adaptée pour composer de la musique concrète, puisque le principe est de recomposer des sons à partir de grains issus d'enregistrements, appelés « quanta de sons » par Xenakis. La synthèse granulaire utilisant l'informatique se développe surtout à partir des travaux de Roads [134], influencé par les compositions de Xenakis. Il faut cependant réellement attendre le développement massif de l'informatique dans les années 2000 pour qu'elle prenne une place importante dans l'industrie du son, notamment dans le design sonore dans les jeux vidéos, comme par exemple ce que propose Siddiq [135]. Concrètement, ce type de synthèse permet de créer une grande variété de sons à partir des quelques échantillons de bases et donc peu de mémoire morte, au prix de nombreuses opérations informatiques, mais qui sont facilement accessibles aux processeurs modernes.

Dans ce type de synthèse, le choix d'un échantillon de base est extrêmement important, puisque ça sera cet échantillon qui, par sa richesse, permettra d'obtenir différents sons. Le choix de ce sample se base beaucoup sur l'expérience ou le ressenti du designer. Parmi les retours d'expérience, il est conseillé ici d'avoir un échantillon pas trop long ( $\approx 5s$ ) pour être capable de le mémoriser, idéalement de niveau sonore et de hauteur pas trop dynamique. Ce caractère stationnaire permet d'éviter des fluctuations en hauteur et en niveau causées par la lecture des grains positionnés à différents endroits dans l'échantillon. Il est intéressant d'avoir des changements de timbre assez importants dans le sample pour avoir une richesse de sons à disposition. Par exemple : plusieurs voix chantées en chœur, un orchestre qui s'accorde ou encore une voie parlée produisant plusieurs sons phonétiques offrent ce type de richesse.

La synthèse granulaire dispose de nombreux paramètres. L'article fondateur de cette synthèse [134] identifie 12 paramètres :

— Point de départ du grain.

- Durée du grain.
- Forme d'onde initiale + son évolution dans le temps.
- Centre de fréquence initiale + son évolution dans le temps.
- Largeur de bande + son évolution dans le temps.
- Densité des grains + son évolution dans le temps.
- Amplitude du grain + son évolution dans le temps.

Ces 12 paramètres sont valables pour une synthèse granulaire faite à partir de signaux créés par des synthétiseurs analogiques, tels que des formes d'ondes simples ou des bruits larges bandes. L'évolution des synthétiseurs granulaires permet de réaliser cette synthèse avec des échantillons plus riches pour lesquelles les paramètres de forme d'onde, centre de fréquence et largeur de bande ne suffisent plus à les caractériser complètement. En effet, les échantillons peuvent désormais être arbitrairement complexes et long. On peut alors programmer d'innombrables nouveaux paramètres, tels que l'enveloppe temporelle avec laquelle est lu chaque grain, la distribution aléatoire du point de départ de lecture du grain ou encore des filtres appliqués à chaque grains pour ne citer que quelques exemples. Dans notre cas nous jouerons alors sur 8 paramètres qui apparaissent sur le schéma 3.8 et la capture 3.7, listés ici :

1.  $x_1$ , la densité de grain.
2.  $x_2$ , la distribution aléatoire du point de départ du grain.
3.  $x_3$ , la vitesse de lecture du grain.
4.  $s_4$ , le sample choisi.
5.  $x_5$ , la position moyenne de départ des grains.
6.  $x_6$ , la taille du grain.
7.  $f_7$ , fréquence de coupure d'un filtre passe-bas appliqué à chaque grain.
8.  $f_8$ , fréquence de coupure d'un filtre passe-haut appliqué à chaque grain.

La capture d'écran 3.7 du synthétiseur granulaire construit sur la base d'un synthétiseur déjà existant, créé par Christophe Poovey,<sup>4</sup> illustre la manière dont les paramètres présentés dans le schéma sont contrôlés. On peut voir les paramètres principaux avec la représentation visuelle du sample qui est lu, ainsi qu'un clavier MIDI guidant la vitesse de lecture et donc la hauteur des grains lus. A cela s'ajoute quatre paramètres : taille du grain, position du grain ainsi que la fréquence de coupure d'un filtre passe-haut et passe-bas ajouté à cette synthèse. Cette matière première construite par synthèse granulaire peut être variable dans le temps, à l'aide de pistes d'automation représentées graphiquement, qui permettent à certains paramètres de varier dans le temps.

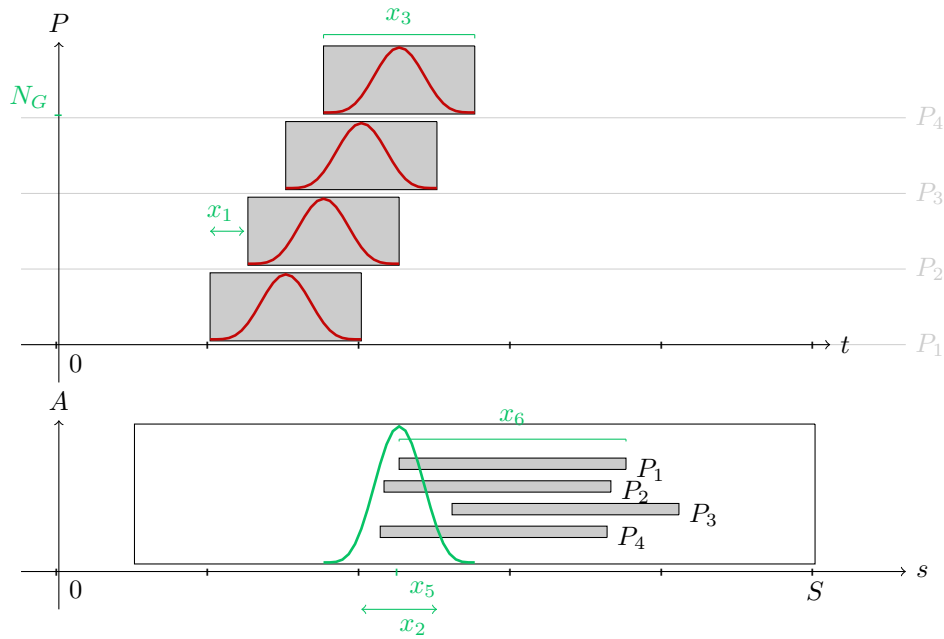
La description de ce bloc illustre bien le nombre de possibilités illimitées offert par ce type de synthèse et la difficulté de s'approprier entièrement un tel outil de création. Certaines matières premières ont ainsi pu être créées à partir de cet outil, telle que Preset10Granv2, que l'on peut écouter dans l'album *Matières premières* disponible en cliquant [ici](#).<sup>5</sup>

#### 3.4.2.2 | Lecture en boucle

Pour le bon fonctionnement du synthétiseur, il est essentiel que les sons composés soient capables de jouer en boucle puisqu'ils seront amenés à être lus à des vitesses différentes, comme pour une synthèse

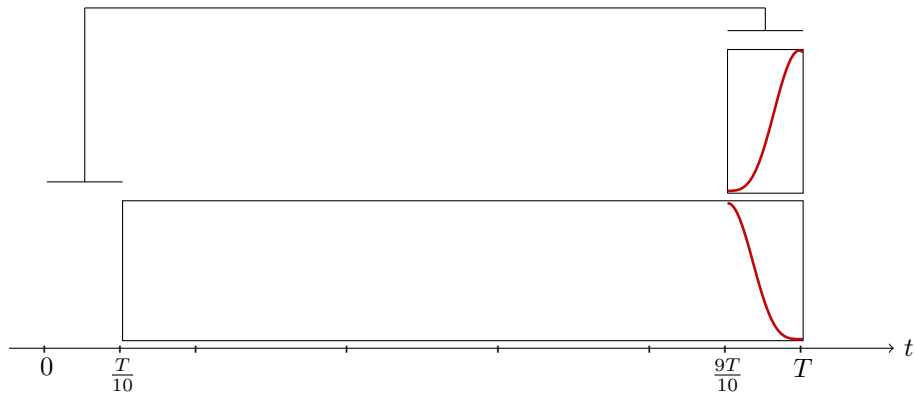
<sup>4</sup> <https://github.com/composingcap/grainflow>

<sup>5</sup> Adresse complète : [https://soundcloud.com/matthieu-duroyon-404568848/preset10gran2?in=matthieu-duroyon-404568848/sets/matieres-premieres&si=3e3bad02c63f40ef9d038ba4f1e6f636&utm\\_source=clipboard&utm\\_medium=text&utm\\_campaign=social\\_sharing](https://soundcloud.com/matthieu-duroyon-404568848/preset10gran2?in=matthieu-duroyon-404568848/sets/matieres-premieres&si=3e3bad02c63f40ef9d038ba4f1e6f636&utm_source=clipboard&utm_medium=text&utm_campaign=social_sharing)

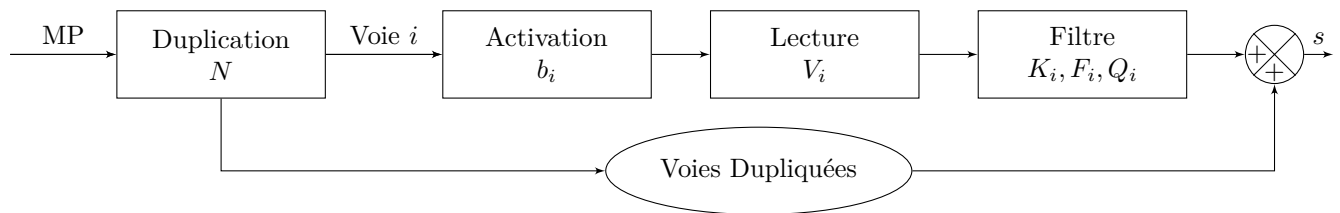


**Figure 3.8.** Schéma illustrant le fonctionnement d'une synthèse granulaire et des différents paramètres dans le synthétiseur construit ici. Le graphique du bas représente la longueur du grain  $x_6$ , la position moyenne  $x_5$  et la distribution aléatoire de départ des grains  $x_2$  dans un sample de durée  $S$ . Le graphique du haut représente le processus de synthèse granulaire : la lecture successive des  $N_G$  grains, avec un espacement temporel  $x_1$ , une vitesse de lecture  $x_3$  et l'enveloppe représentée en rouge.

par table d'ondes où la hauteur d'un son joué est contrôlée par sa vitesse de lecture. Dans le cas où ils ne bouclent pas naturellement avec eux-même, ils sont alors modifiés pour pouvoir être joués en boucle. La modification est faite automatiquement au travers d'une routine Python, suivant le schéma 3.9. Le principe consiste à réaliser un fondu entre le début et la fin de l'audio.



**Figure 3.9.** Reconstruction d'un audio pour permettre une lecture en boucle fluide. On superpose les 10 premiers pourcent aux 10 derniers avec un fondu à iso-niveau. L'extrait en est réduit de 10 % mais peut être maintenant lu en boucle.



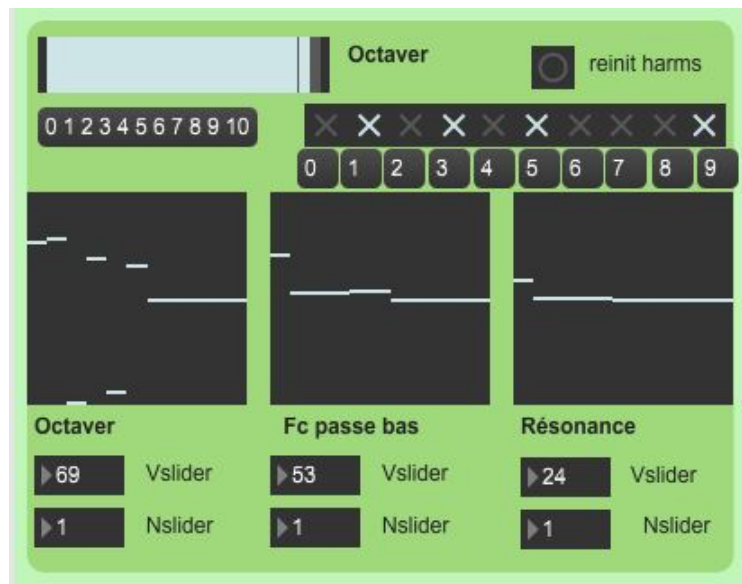
**Figure 3.10.** Schéma-bloc détaillant le fonctionnement du bloc Octaver. Ce schéma laisse apparaître les paramètres  $b_i$  qui contrôlent l'activation de chaque piste,  $V_i$  la vitesse de lecture et donc la hauteur et  $K_i, F_i, Q_i$  les paramètres du filtre résonant. Chacun de ces paramètres sont indépendants en fonction des différentes voies dupliquées  $i$  mais ont les mêmes effets.

### 3.4.3 | Octaver

Une fois la matière première composée et choisie en fonction du type de masquage que l'on souhaite réaliser (Énergétique, Habillage, Attentionnel), il convient d'adapter la matière première au défaut exact présent dans la situation, ainsi qu'à la typologie de masquage suivie. En effet ces matières premières ont été conçues pour être à leurs tours modelées. Pour cela le synthétiseur contient un bloc appelé Octaver. Le principe de ce bloc est de travailler la matière première comme s'il s'agissait d'un partiel, composant un son plus riche. L'idée est donc de superposer plusieurs pistes lues à des vitesses différentes, représentant les différentes harmoniques de ce son plus riche. Le terme harmonique est abusif ici, puisque les différentes voies de la matière première composées peuvent être elles-mêmes harmoniques, on nomme donc ce bloc du nom de l'effet « Octaver », qui se rapproche de la technique implémentée ici. De cette manière, on peut créer plusieurs octaves de notre matière première choisie dans différents canaux et les recomposer ensuite. Ce système permet à la fois de composer des sons harmoniques très riches ou des sons large-bande avec un profil fréquentiel contrôlé, en fonction de la matière première et du nombre d'octaves choisies pour notre synthèse.

Le sous-schéma-bloc 3.10 explicite le fonctionnement précis du bloc Octaver dont l'interface Max est représenté ci-dessous. Une fois la matière première sonore choisie, on peut dupliquer celle-ci un nombre  $N$  de fois souhaité. Ces voies sont ensuite traitées indépendamment avant d'être ré-assemblées à la fin du bloc. Sur une voie  $i$  dupliquée, on peut maintenant choisir ou non de l'activer via le booléen  $b_i$ . Cette option est surtout utile dans la conception du filtre pour permettre d'écouter une à une les voies modifiées. Si la voie est activée, on peut maintenant choisir la vitesse de lecture de celle-ci  $V_i$ . La valeur de  $V_i$  est habituellement prise comme un entier positif, d'où le nom du bloc « Octaver », mais rien nous empêche ici de prendre des valeurs négatives ou rationnelles, même si ces possibilités ne seront pas explorées dans ce chapitre. Enfin, sur chacune des voies, avant de les recomposer, on applique un filtre résonant de gain  $K_i$ , de fréquence de coupure  $F_i$  et de facteur de qualité  $Q_i$ , ce filtre sert majoritairement de passe-bas d'ordre 2 en évitant ainsi que trop de HF ne se superposent de manière incontrôlée et rendent le résultat final trop bruité. La majorité du timbre d'un masque est ainsi donnée par le choix de la matière première et la liste de  $3N$  valeurs paramétrant les  $N$  voies du bloc Octaver. La capture d'écran du bloc Octaver figure 3.11 représente comment sont paramétrés ces éléments. La matière première de la catégorie « Harmonique », ainsi modifiée sert au design du masque nommé Habillage 1 dans la section 3.5.3.

On observe sur la figure 3.11 la construction de 11 octaves, allant de 0 à 10, ou seulement 4 de ces pistes, O1, O3, O5, O9 sont activées. La capture d'écran laisse aussi apparaître la possibilité de venir filtrer chacun de ces canaux avec un filtre résonant. Ces filtres spécifiques à chaque canal permettent d'éviter une saturation des HF causée par la superposition de plusieurs octaves d'une matière première sonore qui peut parfois déjà être assez riche en HF. Dans le cas de notre exemple, cela résulte en un son



**Figure 3.11.** Capture d'écran du bloc Octaver. Celui-ci fonctionne en dupliquant un nombre de fois souhaité une piste et en choisissant la vitesse de lecture de chacune de ces pistes.

assez riche harmoniquement et plutôt BF. Pour ce masque, les paramètres présentés dans le schéma bloc 3.10 sont :

$$\begin{aligned}
 N &= 9, \\
 B &= [1, 0, 1, 0, 1, 0, 0, 0, 1], \\
 V &= [1, *, 3, *, 5, *, *, *, 9], \\
 K &= [0, *, -10, *, -13, *, *, *, -29], \\
 F &= [800, *, 2400, *, 4100, *, *, *, 6800], \\
 Q &= [0.5, *, 0.5, *, 0.5, *, *, *, 0.5],
 \end{aligned}$$

### 3.4.4 | Filtre

Une fois le cœur du masque composé à partir de l'enrichissement d'une matière première sonore grâce à l'octaver, nous appliquons maintenant des modifications plus classiques. On se restreint ici uniquement à l'application de filtres. Une version plus évoluée du synthétiseur aurait pu implémenter ici des effets plus complexes tel que de la réverbération, du délai ou de la compression. Dans le but de garder ce synthétiseur assez simple et surtout de diminuer le nombre de paramètres, cette partie n'a pas été implémentée.

Les filtres utilisés se limitent au nombre de trois, puisque seulement un passe-bas, un passe-haut et un filtre coupe-bande seront utilisés. Ils ne sont appliqués qu'une fois le masque entièrement composé, notamment pour faire des ajustements à posteriori sur un masque, dont une composante tonale est trop forte, ou une plage de fréquence trop chargée. Cela permet de réaliser ces ajustements sans re-paramétrer l'ensemble du masque. On peut par exemple citer l'exemple du masque attentionnel A1 : les retours formulés après une première écoute faisaient état d'une composante tonale trop marquée. Un filtre coupe bande a donc été ajouté à posteriori pour couper la composante en question tel qu'illustré sur la figure 3.12.



**Figure 3.12.** Filtre coupe-bande appliqué en dernier lors de la création du masque attentionnel A1. On peut lire sur la capture une valeur de facteur de qualité  $Q = 0.05$ , un gain  $K = -11dB$  et une fréquence de coupeure  $F = 2465Hz$ .

### 3.4.5 | Aspect dynamique

Pour l'aspect dynamique du masque, un glissando de Shepard-Risset est applicable aux différents masques composés. Cette volonté a été motivée par la présence assez fréquente de ce type d'effet dans les designs sonores liés aux véhicules électriques, de par sa capacité à évoquer un mouvement, et notamment une montée en régime d'un moteur sans pour autant augmenter indéfiniment la hauteur réelle [136].

La figure 3.13 illustre la manière technique de réaliser ce type d'effet à partir d'un seul canal dupliqué 4 fois dont on est capable de modifier la vitesse de lecture (et donc la hauteur). Le fonctionnement du Shepard-Risset se base sur 2 principes fondamentaux. Le premier : une croissance par morceaux de la hauteur, où la discontinuité de la fonction hauteur est synchronisée avec le zéro de la fonction amplitude. Le second : une superposition de plusieurs pistes avec un déphasage de  $\frac{T}{4}$ , où  $T$  est la période de la fonction hauteur et amplitude. La superposition de ces pistes avec un déphasage de  $\frac{T}{4}$  assure une amplitude totale constante si l'amplitude de chaque canal est guidée par une fonction sinusoïdale. Cette figure illustre aussi les différents paramètres dont on a accès en utilisant cet effet, à savoir : la longueur de la fenêtre  $T$ , la tessiture du glissando  $\alpha$  ainsi que la fréquence de départ  $F_{Start}$  qu'aura ce glissando, duquel on pourra déduire la fréquence moyenne susceptible de recevoir le plus d'énergie au cours du glissando  $\hat{f} = F_{Start}(\frac{\alpha+1}{2})$ . Ces paramètres doivent être minutieusement choisis en prenant en compte leurs impacts :

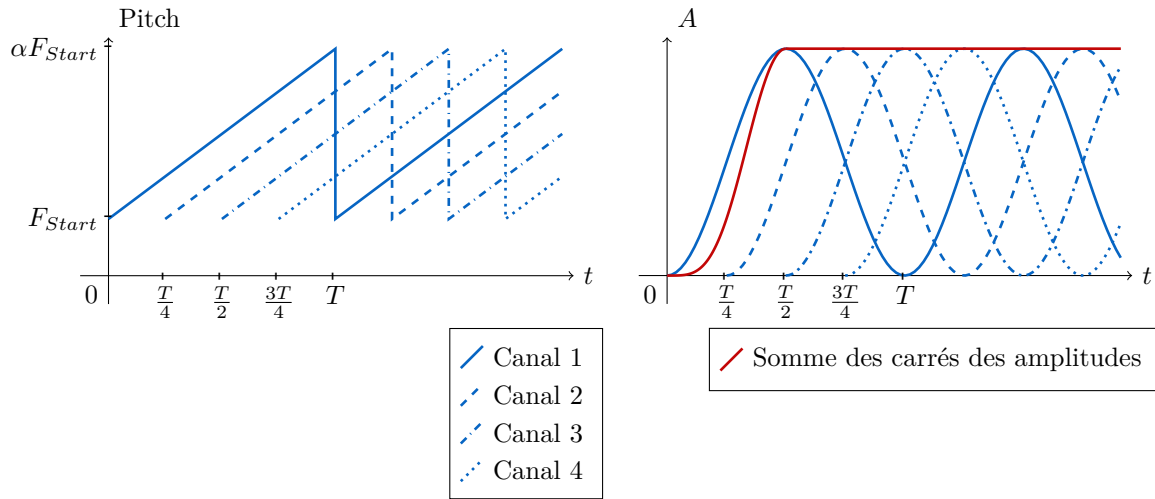
- Le choix d'une fenêtre  $T$  trop grande implique une inertie assez lente de l'effet. Pour que l'illusion soit efficace, il faut qu'on soit en régime établi avec plusieurs canaux se superposant comme représenté sur la figure 3.13. De ce fait une modification du son entrant dans un de ces canaux nécessitera d'attendre  $\frac{T}{2}$  avant que plusieurs canaux se superposent et rendent l'illusion effective.
- Un choix de tessiture  $\alpha$  faible causera des battements assez fort pour des signaux tonaux. En effet, on peut observer sur la figure 3.13 qu'à l'instant  $\frac{T}{2} + \frac{T}{8}$  le canal 1 et 2 ont une amplitude égale, puisque l'amplitude des deux pistes données par des sinus déphasés est égale.

$$f_1 = F_{Start} + (\alpha - 1) \frac{t}{T} F_{Start},$$

$$f_2 = F_{Start} + (\alpha - 1) \frac{t - \delta}{2T} F_{Start}, \quad \delta = \frac{T}{4}$$

$$\frac{f_1}{f_2} = \frac{T + (\alpha - 1)t}{T + (\alpha - 1)(t - \delta)}$$

Comme la croissance de la hauteur est linéaire on peut en déduire les valeurs exactes du rapport



**Figure 3.13.** Schéma de fonctionnement d'une illusion de Shepard-Risset. Ce schéma fait apparaître les paramètres principaux que l'on manipule dans ce synthétiseur, la tessiture  $\alpha$ , la fréquence de départ  $F_{Start}$  et la période  $T$ .

des fréquences à l'instant  $t = \frac{T}{2} + \frac{T}{8}$

$$\begin{aligned} \frac{f_1}{f_2} &= \frac{T + (\alpha - 1)\frac{5T}{8}}{T + (\alpha - 1)\frac{3T}{8}} \\ &= \frac{8 + 5(\alpha - 1)}{8 + 3(\alpha - 1)} \iff \alpha = \frac{(5\frac{f_1}{f_2} - 3)}{5 - 3\frac{f_1}{f_2}} \end{aligned}$$

On montre ainsi que, si  $\frac{f_1}{f_2} = 2^{1/12}$ , ce qui correspond à une seconde mineure, alors  $\alpha = 1.26$ . Si l'on souhaite obtenir une tierce majeure, on souhaite une différence de 5 demi-tons, soit un rapport de fréquence  $\frac{f_1}{f_2} = 2^{5/12}$ , ce qui nous donne  $\alpha = 3.69$ . Par conséquent le paramètre  $\alpha$  peut être choisi pour obtenir, ou éviter, un intervalle musical précis en plus de guider l'intensité de la montée harmonique. Cet arpège musical apparaîtra donc avec une périodicité  $\frac{T}{4}$ , à chaque superposition de canaux. En revanche cet intervalle n'est pas constant au cours du temps, puisque le ratio  $\frac{f_1}{f_2}$  varie lui aussi. Cette considération n'est plus nécessaire pour une piste bruitée pour laquelle la perception de hauteur et de note n'est plus pertinente, ce qui nous permet d'avoir un effet de glissando très progressif avec par exemple  $\alpha = 0.5$  sans avoir de problème de sons perçus comme dissonants.

- Une fréquence de départ  $F_{start}$ . Ce paramètre, combiné avec la tessiture  $\alpha$  permet de choisir la bande de fréquences, centrée en  $\hat{f} = F_{start}(\frac{\alpha+1}{2})$  qui recevra le plus d'énergie dans le cas où l'on cherche à effectuer un masquage énergétique ou positionner l'effet conçu dans une plage fréquentielle très précise.

### 3.4.6 | Résumé de la section

#### Construction d'un synthétiseur

Pour réaliser les différentes typologies de masquage, un synthétiseur est construit. Ce synthétiseur fonctionne en utilisant ce qu'on nomme des matières premières, celles-ci peuvent être des matériaux sonores très élaborés, composés par des designers sonore pour ce but, ou être simplement un signal sonore simple à traiter, comme un bruit blanc. Nous avons exploré en détail ici les possibilités offertes par la synthèse granulaire pour construire ces matières premières. A partir de cette matière première, une série de modifications paramétriques par bloc est effectuée pour obtenir de cette matière sonore la typologie de masquage souhaitée. Les blocs constitutifs sont : un bloc octaver, permettant de changer la hauteur d'une matière première lue en boucle, ou d'en créer des « octaves » pour enrichir le masque. Un bloc filtre, permettant d'appliquer des filtres classiques en sortie du bloc octaver. Et un bloc dynamisme, permettant d'ajouter un effet dynamique en utilisant l'illusion du glissando de Shepard-Risset.

## 3.5 | Composition des masques

Cette section détaille d'un point de vue plus technique la composition des masques sonores utilisés ainsi que la situation sonore. La figure 3.14 représente le spectre de la situation sonore telle que reconstruite, avec la ventilation et le sifflement. Le défaut de sifflement a été augmenté en niveau lors de la reconstruction de la scène pour le rendre très saillant.

Le sifflement se caractérise par une émergence de près de 20 dB par rapport au bruit de fond, une fréquence fondamentale de 3 kHz ainsi qu'un caractère légèrement fluctuant. Le son de ventilation est très large-bande, avec une pente de -5 dB/octave très proche du bruit brownien. Cette pente n'est pas parfaitement respectée, on remarque en particulier dans le spectre un creux s'étalant de 500 à 1000 Hz, ainsi qu'une bosse aux alentours de 2000 Hz, ces spécificités seront exploitées lors de la construction des masques. D'un point de vue temporel, le son de ventilation est quand à lui très constant. Tous les masques dont la composition est détaillée ici sont disponible à l'écoute et au téléchargement dans l'album *Masques Seuls* en cliquant [ici](#).<sup>6</sup>

### 3.5.1 | Masque Énergétique 1 (E1)

Le premier masque composé est probablement le plus simple, il a pour vocation de faire disparaître le sifflement en rajoutant le moins d'énergie possible et se fondant avec le bruit de fond : la ventilation. En reprenant la description des sons proposée par Carron et al. [106], disponible sur la plateforme *SpeaK*<sup>7</sup> : *Word4Sounds*, on propose une description réduite du masque. Le son ajouté sera donc bruité, non dynamique, naturel continu et constant.

Pour réaliser ce masque, la matière première utilisée sera un bruit blanc classique. Les modifications apportées à cette matière première par le bloc Octaver consistent à construire 2 canaux, chacun avec un filtre résonant centré sur une fréquence différente. La fréquence de résonance du premier filtre est centrée autour de 1000 Hz sur le creux du spectre de la ventilation évoquée précédemment, le deuxième

<sup>6</sup> Adresse complète : [https://soundcloud.com/matthieu-duroyon-404568848/sets/masques-seuls/s-BuNWHDOZAJT?si=9c536f51c4f5419aa75aa3c4a0057be7&utm\\_source=clipboard&utm\\_medium=text&utm\\_campaign=social\\_sharing](https://soundcloud.com/matthieu-duroyon-404568848/sets/masques-seuls/s-BuNWHDOZAJT?si=9c536f51c4f5419aa75aa3c4a0057be7&utm_source=clipboard&utm_medium=text&utm_campaign=social_sharing)

<sup>7</sup> <https://speak.ircam.fr/lexique/lexique-ircam/>

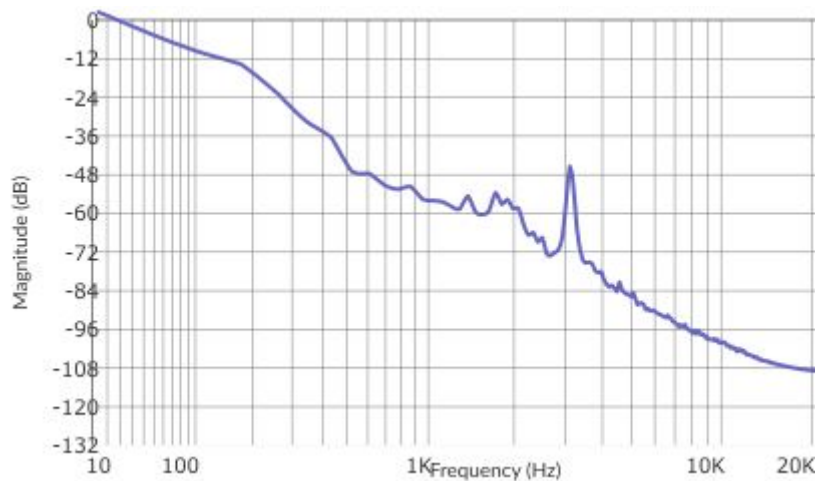


Figure 3.14. Spectre fréquentiel de la situation sonore incluant le son de ventilation et le sifflement.

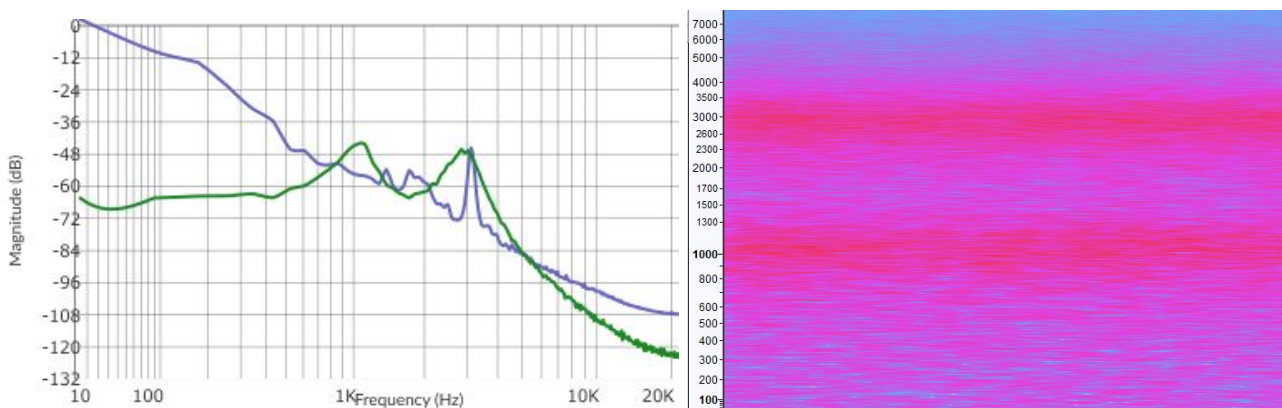


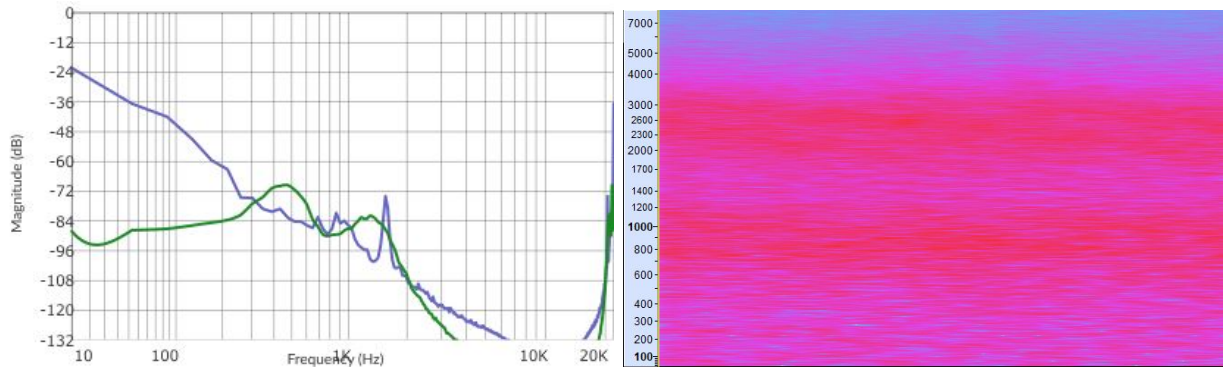
Figure 3.15. A gauche, Spectre du masque E1 (en vert) superposé à la situation sonore (en bleu). Le spectre est calculé avec une fft de 512 points puis moyenné sur une fenêtre glissante de 200 points de mesure, cela permet de réduire la composante aléatoire du bruit et faciliter la lecture du graphe au prix d'une précision temporelle faible. On peut clairement apercevoir l'ajout d'énergie dans la bande de Bark du défaut à masquer ( $3kHz$ ) et dans le creux du son de ventilation ( $\approx 1kHz$ ) afin de baisser le centroïde spectral. A droite, spectrogramme du masque seul qui met en évidence le caractère stationnaire du masque

filtre est quant-à-lui centré sur le défaut aux alentours de 3000 Hz pour effectuer le masquage. Cette composition en deux bandes spectrales a pour objectif de diminuer le centroïde spectral du son, parfois aussi appelé « agressivité » du son par les acousticiens automobiles, en n'ajoutant pas trop d'énergie dans les HF (Fig. 3.15). Différentes déclinaisons du masque E1 pourront ainsi être construites avec différents équilibres. Les autres blocs ne sont pas exploités pour concevoir ce masque.

### 3.5.2 | Masque Énergétique 2 (E2)

Ce second masque énergétique est assez similaire au premier dans ses intentions, à cause des contraintes imposées par ce type de stratégie. En revanche, le masque est pensé pour être plus dynamique que E1. La description verbale réduite de ce masque sera : bruité, dynamique, naturel continu et légèrement descendant.

La composition de ce masque est similaire au premier, car il utilise la même matière première et le même paramétrage du bloc Octaver. La différence réside ici dans l'utilisation du bloc dynamique pour ajouter une légère modulation temporelle d'amplitude de période 2s ainsi qu'un glissando de Shepard-Risset descendant. Pour le glissando, le paramètre de tessiture ( $\alpha$ ) est fixé égale à  $-1,5$  et la



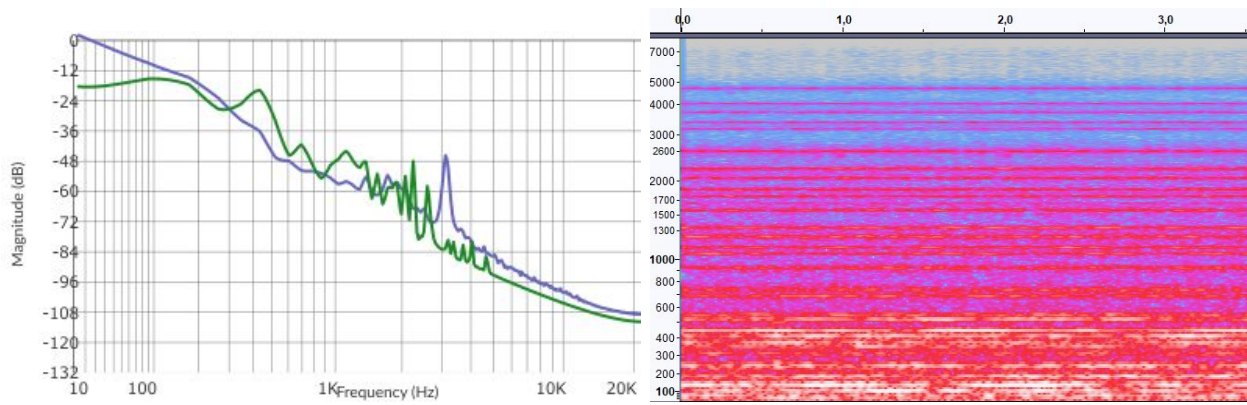
**Figure 3.16.** A gauche, Spectre du masque E2 (en vert) superposé à la situation sonore (en bleu). Le spectre a été calculé avec les mêmes paramètres que précédemment et ne laisse donc pas apparaître le caractère dynamique du masque. Le spectrogramme du masque E2 avec une fenêtre fréquentielle de 16384 points, ce qui permet d’avoir une bonne résolution fréquentielle et faire apparaître visuellement l’effet du Glissando.

taille de fenêtre  $T = 8s$ . Ces choix de paramètres permettent une variation lente et descendante de la hauteur, qui vise à créer un son plus apaisant tout en étant dynamique. Le choix de la fréquence de départ  $F_{Start}$  garanti que le maximum d’énergie du filtre est bien superposé au sifflement pour le masquer, malgré l’aspect dynamique du masque. La représentation spectrale adoptée pour le masque E1 n’est pas adaptée ici puisque elle ne permet pas de représenter l’évolution temporelle du masque. La figure 3.16 représente donc le spectrogramme qui permet de mieux apprécier l’évolution temporelle du masque. Le code couleur nous permet aussi de remarquer la similitude avec le masque E1, avec deux bosses d’énergies du spectre centrée aux alentours de 1000 Hz et 3000 Hz.

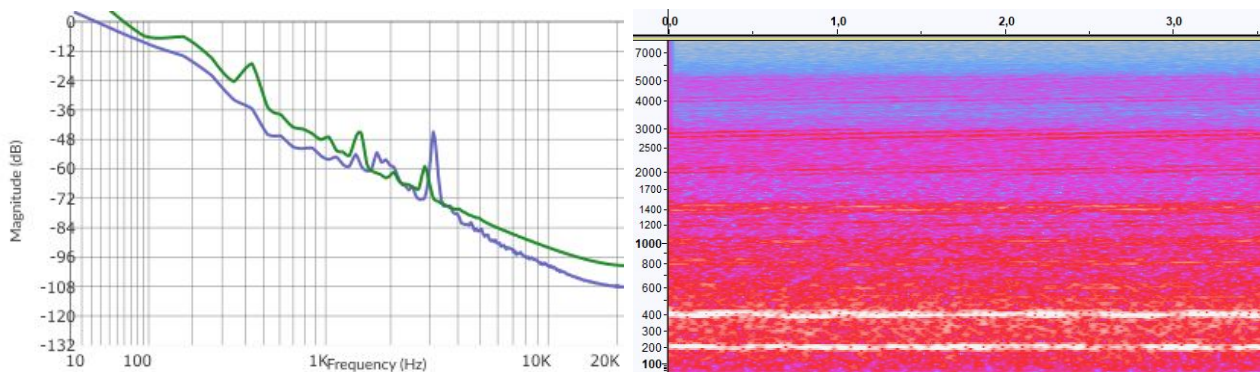
### 3.5.3 | Masque Habillage 1 (H1)

Ce premier masque de la stratégie Habillage a pour objectif d’inclure le son de sifflement dans un son plus large, inspiré de l’univers de la science-fiction. Pour cela la matière première utilisée, composée par Eléonore Vialatte, est catégorisée comme un son imaginaire (Sc.Fi) et harmonique (Tab. 3.3). Ce masque tente d’inclure le sifflement dans un son que l’on pourrait rattacher à un fonctionnement de machinerie fictive. Par construction le masque sera assez brillant pour pouvoir inclure le sifflement ( $\approx 3kHz, 20dB$ ) comme un harmonique de ce son, avec un risque de devenir strident. Il est nécessaire d’être prudent sur la construction de ce son là pour ne pas rendre ce caractère brillant trop important. En reprenant les termes du lexique Words4Sounds de SpeaK, la description du masque correspond à un son tonal, métallique, légèrement fluctuant, avec une composante discontinu dans les HF.

La matière première utilisée est ici assez harmonique. Cependant, par construction les matières premières restent assez simples, en l’occurrence celle utilisée ici est BF et possède peu d’harmonique. Ainsi le bloc Octaver est utilisé pour « étendre » le son aux MF et HF pour obtenir un timbre avec suffisamment d’énergie aux alentours de 3 kHz pour inclure efficacement le sifflement (Fig.3.17). Combiné à ce bloc, le bloc Filtre ici coupe l’énergie du masque juste sur la même bande de fréquence que le sifflement pour que ce dernier prenne le rôle d’une harmonique « fantôme » du masque.



**Figure 3.17.** A gauche, spectre du masque H1 superposé à la situation sonore. On remarque notamment le nombre de sous-harmoniques construites au sifflement ainsi que de très légères harmoniques au delà des 3 kHz. A droite, spectrogramme.

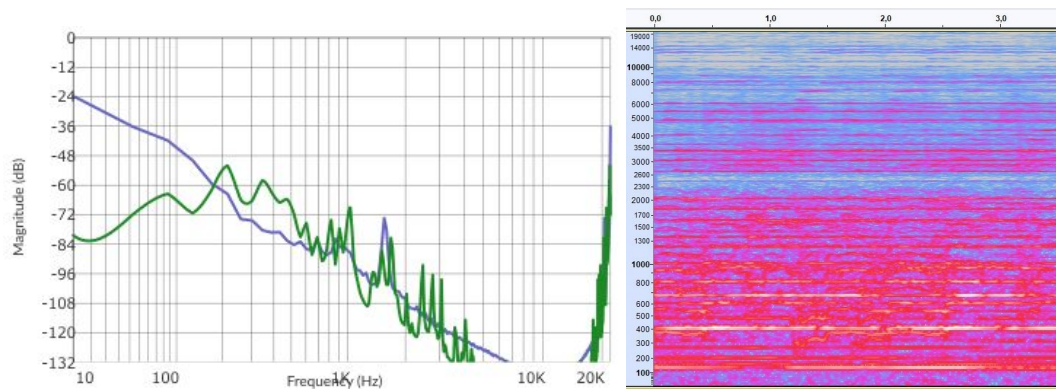


**Figure 3.18.** A gauche, spectre du masque H2. On peut observer en vert assez nettement sur le spectre, 3 des 4 canaux créés par le bloc Octaver et notamment leur position dans les creux du son de Ventilation, et un en superposition avec le sifflement à habiller.

### 3.5.4 | Masque Habillage 2 (H2)

Le masque dans lequel le sifflement doit se noyer est cette fois-ci plus proche de l'univers automobile. Le son ajouté ici a pour vocation de rappeler le son théorique que pourrait faire une voiture. Le son ajouté est donc assez tonal, dynamique, constant, légèrement ascendant et dissonant.

Ici la matière première sélectionnée possède par son timbre un aspect proche du son imaginaire que pourrait faire un véhicule électrique. La modification du bloc Octaver va consister à dupliquer la piste en 4 canaux, chacun avec une vitesse de lecture respectivement de 1, 3, 5, 7 (Fig. 3.18). Cela permet d'élargir spectralement le masque en gardant l'essence de la matière première, initialement composée en BF. Le choix des octaves impaires est motivé par la volonté de garder un timbre assez proche de ce que l'on peut retrouver dans les motorisations classiques de véhicule ou les harmoniques impaires ont un rôle important [123]. La hauteur du fondamental, ainsi que le bloc Filtre sont calibrés pour s'assurer que la raie causé par le sifflement se superpose bien avec une raie de masque pour l'inclure dans ce son. Enfin le bloc dynamique ajoute un glissando pour garder cette composante dynamique ascendante rappelant une montée en régime d'un moteur. Les paramètres du glissando sont de  $T = 5ms$  et  $\alpha = 1.18$ , ce qui nous donne un intervalle musical proche d'une seconde mineure laissant apparaître un léger battement, et une croissance très légère de la hauteur, l'effet dynamique est donc très léger pour ne pas donner l'illusion d'un mouvement du véhicule qui est inexistant dans cette situation sonore.



**Figure 3.19.** Spectrogramme du masque A1 avec une fenêtre fréquentielle de 4096 points, ce qui permet d’avoir une bonne résolution fréquentielle et bien faire apparaître visuellement les événements sonores assez court. On remarque la signature assez caractéristique de la synthèse granulaire notamment dans les HF où l’on voit l’apparition puis la disparition de grains.

### 3.5.5 | Masque Attentionnel 1 (A1)

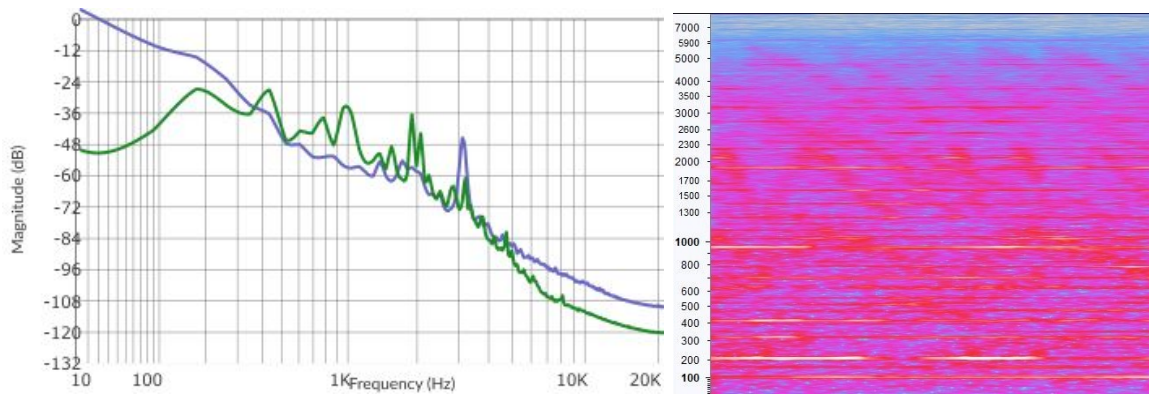
Pour appliquer la stratégie de détournement d’attention escomptée avec cette typologie de masquage, il a été choisi ici de composer un masque très dynamique, avec de nombreux événements sonores éphémères. En reprenant le lexique Word4Sounds de SpeaK, le masque construit ici sera riche, dynamique et assez brillant.

Dans le cas présent, la matière première prise est constituée de plusieurs voix chantées *a capella*. Ces voix chantées ont en premier lieu été modifiées en parallèle par le bloc Synthèse Granulaire, de manière à les rendre peu reconnaissables et notamment enlever une partie de leur musicalité. Cela est notamment possible en sélectionnant des grains de taille assez importante  $\approx 500ms$ , ce qui permet de reconnaître la hauteur et le timbre du son [133], mais ces grains sont sélectionnés aléatoirement sur une portion d’environ deux secondes. Comme le chant évolue lentement cela permet d’avoir des grains suffisamment similaires entre eux pour ne pas être dissonant, mais tout de même différents pour enrichir le son et le rendre difficilement reconnaissable. Ici aussi la représentation fréquentielle ne permet pas de représenter fidèlement l’apparition de ces nombreux événements sonore, on y ajoute alors le spectrogramme figure 3.19 en complément. Le spectrogramme reflète assez clairement l’application d’un filtre coupe-bande mentionnée plus tôt (Sec.3.4.4) de facteur de qualité  $Q = 0.05$ , un gain  $K = -11dB$  et une fréquence de coupure  $F = 2465Hz$ . Ici le spectrogramme affiche le masque seul sans le son de ventilation et le défaut qu’il augmente, ce qui améliore la lisibilité du graphe mais ne rend pas compte des éléments extérieurs qui viendront donc compléter les creux du spectrogramme.

### 3.5.6 | Masque Attentionnel 2 (A2)

Le deuxième masque de la stratégie attentionnelle vise à créer une ambiance sonore entendue et assimilée comme un nouvel objet sonore et pas seulement des événements ponctuels comme pour A1. L’aspect réduit du masque, décrit à l’aide du lexique Word4Sounds, sera dynamique, riche, mat.

Contrairement aux autres masques, la matière première sonore utilisée ici est déjà aboutie puisqu’elle provient de sons fournis par Laurent Worms. De ce fait le son composé est déjà pensé pour appartenir à l’habitacle de la voiture, même s’il n’est pas composé pour cette scène sonore spécifique. Cette matière sonore a été choisie comme pour le masque A1, pour son caractère assez riche et dynamique. Encore une fois, le défaut que l’on souhaite traiter ici est assez HF, et la ventilation charge déjà les BF, de



**Figure 3.20.** A gauche, représentation fréquentielle du masque A2. Le spectre affiche un instantané du masque et donc ne restitue pas correctement le caractère temporel du masque, il est cependant suffisant pour illustrer la volonté d'avoir l'énergie plutôt MF ou BF pour garder un son le plus mat possible. A droite spectrogramme du masque A2 avec une fenêtre fréquentielle de 4096 points pour compléter le spectre.

ce fait le bloc Octaver et Filtre servent ici à déplacer l'énergie dans les bandes de fréquences moins chargées comme le creux identifié vers  $\approx 500\text{Hz}$  figure 3.20. Ce masque, de par sa typologie, ne tente pas de faire du masquage énergétique, de ce fait on n'ajoute pas ou peu d'énergie dans les bandes de fréquences du sifflement  $\approx 3000\text{Hz}$  et au delà, ce qui permet de garder un son relativement mat. Le spectrogramme 3.20 permet de mettre en évidence le caractère dynamique du masque, conformément à la typologie de masquage adoptée.

### 3.5.7 | Résumé de la section

#### Composition des masques

Dans cette section, deux masques sont composés pour chaque typologie, appelées E1, E2, A1, A2, H1, H2, respectivement pour la typologie Énergétique, Attentionnelle et Habillage. Les matières premières utilisés, ainsi que les masques composés peuvent être écoutés sur la plateforme SoundCloud, à l'adresse associée à ce manuscrit en cliquant [ici](#)<sup>a</sup> et en parcourant les différents albums à disposition.

<sup>a</sup> Adresse complète : <https://soundcloud.com/matthieu-duroyon-404568848/sets/matieres-premieres>

## 3.6 | Conclusion de chapitre

Ce chapitre détaille le processus suivi pour la création de sons dédiés à être ajoutés à un environnement pour opérer le concept d'acoustique augmentée. La première étape consiste à caractériser les situations sonores sur lesquelles travailler, en s'appuyant sur une description des sons basée sur le lexique élaboré au chapitre précédent, ainsi que sur le contexte, les habitudes, et les réactions des utilisateurs lors de leur apparition. Cette analyse contextuelle, spécifique au domaine de l'automobile, a été rendue possible grâce aux questions posées lors des interviews sur le confort. La deuxième étape, qui constitue la contribution originale de ce travail, repose sur l'identification de trois typologies de masquage inspirées de la littérature, nommées masquage énergétique, habillage, et masquage attentionnel. Chacune de ces typologies présente une fonction et un aspect différent, guidés par les éléments contextuels identifiés.

La création sonore est permise par la structuration des processus et le développement des outils numériques associés. Parmi l'ensemble des masques créés avec cet outil, seuls six, qui semblaient les meilleurs, ont été gardés et présentés dans ce chapitre. Le synthétiseur construit n'est cependant pas spécifique à la seule situation sonore traitée, et a été pensé pour être réutilisable pour des situations sonores plus complexes. La logique de fonctionnement par bloc permet d'adapter la synthèse à de nouvelles consignes de design ou de nouvelles situations.

Par construction, les trois typologies présentées sont différentes dans leurs intentions. En particulier, les masques énergétiques proposés sont très neutres et ont une expressivité limitée. Les masques de la typologie habillage, en revanche, laissent un peu plus de place à une forme d'expressivité. Mais c'est surtout la typologie attentionnelle qui offre une plus grande liberté de création, mais elle est complexe à mettre en œuvre, car elle nécessite un ensemble de réglages fins des blocs du synthétiseur présenté, ou fait appel à des matières premières plus élaborées. Il est nécessaire d'avoir une maîtrise des outils de création en synthèse sonore, ainsi qu'une capacité à élaborer des propositions singulières en design sonore pour répondre aux enjeux de création de masquage de type attentionnel. Ce constat montre la place importante que pourrait avoir le designer sonore à l'avenir dans le monde automobile.

Le chapitre suivant portera plus en profondeur sur l'évaluation quantitative que l'on fera de ces masques et si ces évaluations sont en accord avec les intentions des différentes typologies, à savoir, améliorer le confort.



## 4. Évaluation des solutions conçues avec la méthodologie *Best-Worst Scaling*

### 4.1 | Introduction de chapitre

Les chapitres précédents ont porté sur la description de l'environnement sonore de l'habitacle du véhicule électrique en proposant un lexique construit à partir des interviews d'acousticiens du domaine ; puis dans un second temps, sur l'ajout de sons : l'acoustique augmentée, à une situation sonore bien identifiée et bien décrite. La question qui sera traitée dans ce chapitre est celle de la quantification de l'effet produit par cette acoustique augmentée.

Cette étude perceptive a donc pour but de mesurer les effets des différentes typologies de masque, notamment par rapport à leur fonction respective, précisée dans le chapitre sur leur conception (Chap. 2). Cette évaluation doit aussi être cohérente avec la notion de confort et sa définition proposée dans le premier chapitre (Chap. 1) qui est rappelée ici : « *Le confort sonore est ce qui contribue à une sensation agréable dans l'utilisation d'un objet, en termes de retour sonore, cohérent avec les attentes de l'utilisateur et l'environnement sonore déjà existant.* ».

La forte dimension contextuelle du confort rend son évaluation par une estimation directe délicate. Pour faciliter son évaluation, la situation sonore sur laquelle l'étude sera menée, a justement été choisie pour son contexte d'utilisation simple et bien cadré. De plus, l'analyse des interviews sur la description des sons et du confort nous permet déjà d'avoir une hypothèse forte : le sifflement jugé inacceptable et inconfortable est responsable de l'inconfort ressenti dans la scène, alors que le son de ventilation est attendu et n'est pas responsable de l'inconfort. Ce cadre posé au précédent chapitre permet de réduire, ou à défaut, de bien caractériser les éléments contextuels. Ce cadre maîtrisé nous permet alors de restreindre l'évaluation du confort à deux jugements reliés à la définition. Le premier : jugement d'agrément, aussi appelé agréabilité, perçu de la scène dans cette situation d'utilisation de la ventilation. Le second : le jugement de détection ou la détectabilité, d'un son inattendu et peu harmonieux avec l'environnement sonore, le sifflement. Ces deux jugements, ainsi que l'impact de chacun des masques sur ceux-ci, seront alors évalués dans ce chapitre.

La problématique guidant alors ce chapitre est la suivante : les masques composés selon les typologies présentées permettent-ils d'impacter les deux aspects suivants de la situation sonore traitée : l'agrément et la détectabilité ? Dans un premier temps, nous verrons les différents types de test qu'il est possible d'entreprendre, et pourquoi nous avons choisi d'utiliser la méthodologie *Best-Worst Scaling* (BWS). Dans un second temps, nous verrons comment a été utilisée la synthèse paramétrique présentée au chapitre précédent pour la conception de cette étude. Nous verrons enfin les résultats tirés de cette expérience et les discussions qui s'en suivent.

## 4.2 | Méthodes d'évaluation perceptive

### 4.2.1 | Les évaluations indirectes

L'utilisation de méthodes d'évaluation perceptive est récurrente en acoustique, et plus précisément en psychoacoustique, pour établir des relations entre des grandeurs physiques et des grandeurs sensorielles en audition. Une première catégorie d'évaluation utilisée est l'évaluation indirecte, pour la mesure de seuils de détection ou seuils de discrimination, appelés *Difference Limen* (DL). C'est ce type de test qui permet de quantifier, par exemple, la différence entre deux niveaux sonores perçus, depuis les travaux de Fechner [137], ou encore la différence perçue entre deux fréquences [138], ou entre deux centroïdes spectraux [139]. Ces travaux permettent alors d'établir des valeurs de variations minimales d'une grandeur acoustique pour être juste perçue, appelées *just-noticeable difference* (jnd). Ces seuils de perception sont alors clefs dans la conception d'évaluations d'un second type : les évaluations directes. Ces valeurs de jnd seront alors réutilisées dans la suite pour s'assurer que la différence entre deux stimuli soit bien perçue.

### 4.2.2 | Les évaluations directes

L'évaluation directe consiste à directement demander au participant d'évaluer une grandeur perçue pour un stimulus. Dans le cas de notre étude, une méthode adaptée est alors l'évaluation par échelles d'intervalles bornées par un DS, qui consiste à demander l'évaluation d'un stimulus entre un descripteur et son opposé. L'utilisation de cette méthode, simple en apparence, nécessite quelques précautions [140] :

- Contrôler les paramètres des stimuli tels que le niveau sonore, la hauteur et la durée du son.
- Choisir un nombre approprié d'échelles dans le cas où plusieurs dimensions sont évaluées.
- Choisir des descripteurs pertinents.
- Définir la signification des descripteurs utilisés pour les échelles.

Ce que révèlent ces prérequis pour ce type d'expérience, c'est la difficulté de choisir les descripteurs à évaluer. Le cas idéal est celui où la population interrogée est déjà familière des descripteurs, par formation par exemple [99], ce qui permet de s'assurer qu'ils soient pertinents et clairs pour les participants. Une autre possibilité est d'avoir des échelles déjà bien établies dans le domaine d'intérêt [141, 142], cela pose tout de même la question de la traduction, ou des habitudes d'utilisation des termes dans le cas où les résultats sont établis pour une population et réutilisés sur une autre pour laquelle des différences significatives peuvent apparaître [143, 144, 145]. Une autre solution est aussi de procéder à une sélection des meilleurs descripteurs en amont de l'expérience parmi une liste présentée [30, 146, 147]. Dans notre cas, la démarche s'approche plutôt de cette dernière solution puisque le travail de construction du lexique et l'analyse lexicale a justement permis de choisir des descripteurs adaptés au contexte que nous souhaitons étudier.

Ainsi l'évaluation directe par DS est une méthodologie adaptée au type de mesure souhaité ici. En revanche un des défauts de l'évaluation directe est sa plus grande sensibilité aux erreurs d'appréciation, notamment lorsqu'il est difficile de quantifier une perception, ou lorsque les valeurs absolues de deux stimuli sont proches l'une de l'autre selon la dimension évaluée. Susini et al. [148] montrent par exemple la variabilité que peut avoir une évaluation DS en fonction des consignes, et le faible degré de différenciation entre différentes conditions expérimentales trop proches.

### 4.2.3 | La méthode de comparaisons par paires

Pour affiner la discrimination perceptive entre deux sons, un paradigme d'évaluation par paires peut être adopté [149]. Ce type d'évaluation est particulièrement adaptée pour effectuer des *Multi-Dimensional Scaling* (MDS) [150, 151, 152], qui consistent à juger la dissimilarité entre des paires de sons présentées. Cette dissimilarité est alors traitée comme une distance euclidienne pour en déduire un espace latent expliquant les principales dimensions perceptives utilisées pour juger les différences [153]. La MDS est une alternative aux échelles sémantiques précédentes, puisque celle-ci ne requière aucune hypothèse sur les descripteurs à utiliser, seule la similarité entre les stimuli est jugée.

Cette comparaison par paires permet en effet de relever des différences fines entre les stimuli, mais au prix d'un nombre élevé d'évaluation. Chaque paire de stimuli possibles est évaluée, pour un nombre  $N$  de stimuli, on obtient  $\frac{N(N-1)}{2}$  évaluations par participant.

Dans le cas de notre évaluation, les dimensions perceptives à évaluer sont restreintes puisque seules deux jugements nous intéressent ici : le jugement de détection du défaut et le jugement d'agrément de la scène sonore. En revanche le nombre de stimuli est grand puisque les typologies identifiées au chapitre précédent peuvent contenir de nombreux masques différents. Une comparaison par paires serait difficilement réalisable : pour seulement deux masques par typologie et deux conditions expérimentales (avec et sans défaut) ce qui fait  $2 \times 3 \times 2 = 12$  sons, soit 66 paires à évaluer. Si trois masques sont utilisés par typologie au lieu de deux, le nombre de paires à évaluer augmente à 153.

## 4.3 | Le choix de la méthodologie *Best-Worst Scaling*

### 4.3.1 | Le principe

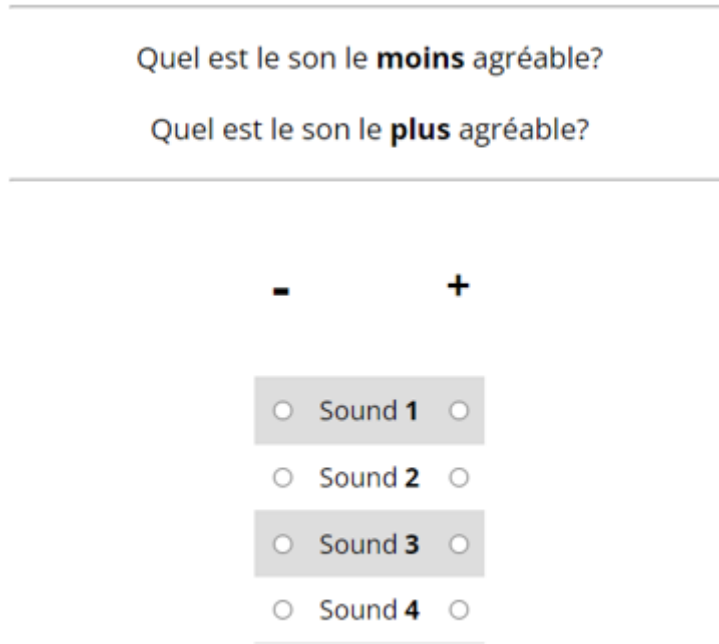
Une alternative à la comparaison par paires mais qui permet un jugement relatif des sons est la BWS, utilisée régulièrement dans des études économiques ou de marketing [154]. Cette méthodologie consiste à proposer au participant un groupe de sons, souvent quatre [155], parmi lesquels le participant doit choisir le stimulus qui correspond le mieux et le moins bien à un critère à évaluer. Pour illustration, la figure 4.1 présente l'interface utilisée pour ce paradigme dans l'étude menée ici.

### 4.3.2 | Les algorithmes de scoring

La problématique de la BWS réside dans la manière d'estimer une valeur numérique pour les stimuli évalués à l'issue de l'expérience. Une méthode assez simple consiste à attribuer à chaque stimulus un score en fonction du nombre de fois qu'il a été choisi le pire ou le meilleur. Cette solution, bien que fonctionnelle, présente de nombreux défauts. En effet, il est alors possible que des stimuli n'aient jamais été sélectionnés et n'aient donc ainsi pas de scores. De plus, cette solution ne permet pas d'utiliser toute l'information déduite d'un jugement BWS puisque dans une évaluation telle que présentée sur la capture 4.1, les 5 inégalités suivantes pour un essai en sont déduites :  $i \in [1, N]$ , où  $b_i$  est le meilleur stimulus,  $w_i$  le pire, et  $s_{1,i}, s_{2,i}$  les deux non sélectionnés, alors :

$$b_i > s_{1,i}, b_i > s_{2,i}, b_i > w_i, w_i < s_{1,i}, w_i < s_{2,i} \quad (4.1)$$

L'utilisation d'algorithmes de *turnament scoring* présentés par Hollis [155] permet de proposer des valeurs numériques aux sons évalués en utilisant les 5 inégalités déduites d'un essai. Ces algorithmes



**Figure 4.1.** Capture d'écran d'une interface BWS à quatre sons. L'interface a été réalisée en Javascript grâce à la librairie JsPsych.

permettent de donner un score à chaque stimulus en fonction du nombre de "matches" gagnés ou perdus, tel que le Elo [156] utilisé dans de nombreux tournois sportifs internationaux. Ces algorithmes sont alors cruciaux dans ce type d'analyse et l'étude exhaustive de Hollis permet de quantifier précisément les avantages et inconvénients de chacun.

Dans notre cas, le choix se porte sur l'algorithme *value learning*, qui offre le meilleur compromis entre la qualité des résultats et la résilience au bruit de mesure [155]. Cet algorithme fonctionne en calculant un score  $V_A$  pour un stimuli  $A$  de manière incrémentale à partir des différents matchs  $i$  auxquels il a pris part. La règle de calcul utilisée pour une itération faite à partir du match  $i$  est la suivante :

$$V_{A,i+1} = V_{A,i} + \alpha\beta(\gamma - V_X), \quad (4.2)$$

avec les paramètres suivant :

$$\begin{aligned} V_X &= \text{Score du stimulus adverse dans le match,} \\ \alpha_{A,i,x} &= \frac{V_X(1 - V_{A,i})}{V_{A,i}(1 - V_X) + V_X(1 - V_{A,i})}, \text{ appelée saillance,} \\ \beta_i &= \frac{0.025}{i}, \text{ appelée taux d'apprentissage,} \\ \gamma &= 0 \text{ en cas de défaite, } 1 \text{ sinon ; appelée valeur observée.} \end{aligned}$$

La saillance  $\alpha$  traduit l'importance que représenterai la préférence d'un stimulus  $A$  sur un autre  $X$  lors de son match  $i$ , par rapport aux scores déjà calculés sur tous les matchs précédents. Celle-ci augmente lorsque  $V_{A,i} < V_X$ , autrement dit lorsqu'une victoire est inattendue selon les score déjà estimés. De plus, le taux d'apprentissage  $\beta_i$  diminue avec le nombre d'itérations ce qui permet une convergence facilitée des résultats. Ces paramètres là ont été conservés tels quels et sont donc les mêmes que Rosi et al. ainsi que Hollis, dans leur implémentations des fonctions de scoring.

L'utilisation pratique de ces algorithmes dans une expérience perceptive liée au son est alors testée par Rosi et al. [157] et prouve ainsi la qualité similaire des résultats entre une étude par DS et par BWS, en montrant un gain de temps et de confort lors de la réalisation de la tâche pour le participant. Cette facilité accrue pour le participant est ici un argument de poids, puisque la tâche de détectabilité du sifflement dans la scène sonore est présumée difficile. A cela s'ajoute la volonté dans notre étude de travailler avec un large corpus de son ( $N > 150$ ). Ces deux éléments ont donc motivé ici l'utilisation d'une BWS plutôt qu'une autre méthode.

### 4.3.3 | La conception d'une expérience BWS

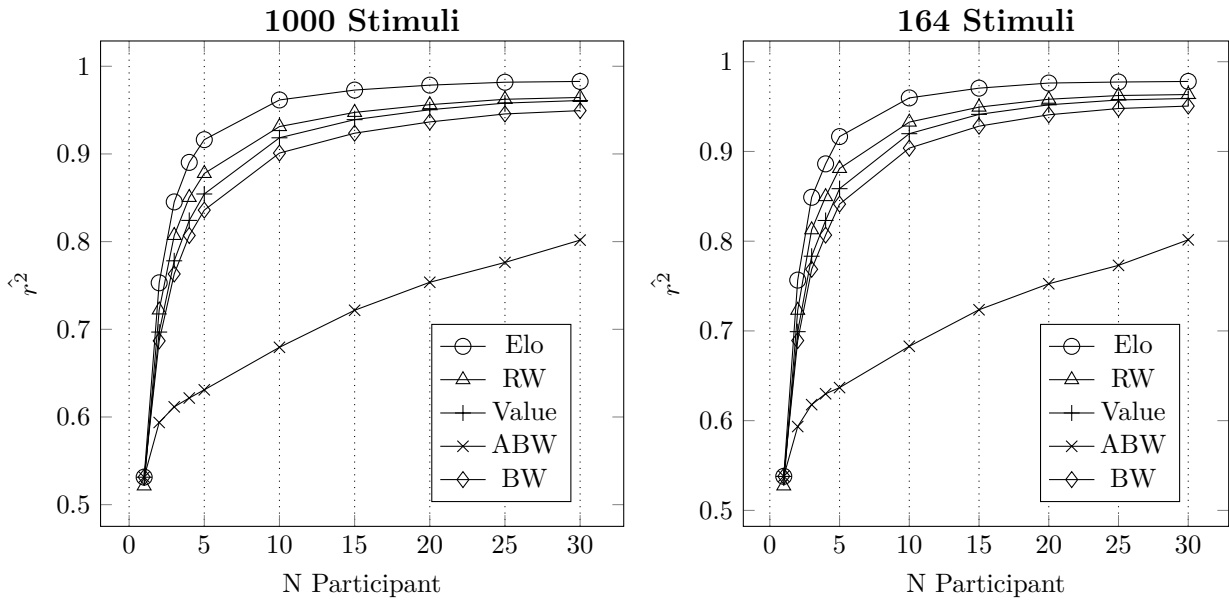
Ce type d'algorithme décuple alors le potentiel d'informations qu'il est possible d'extraire d'une expérience BWS. Cependant l'utilisation de ces algorithmes nécessite une étape supplémentaire dans la conception de l'expérience pour garantir une symétrie dans les données. En effet le dénombrement du nombre de P-uplets possible parmi  $N$  sons nous donne :  $\frac{N!}{P!(N-P)!}$ , pour  $N=100$  sons à évaluer avec des 4-uplets on obtient près de  $4 \times 10^6$  essais à effectuer. Évaluer l'ensemble des P-uplets n'est alors pas possible. Pour obtenir un maximum d'information avec l'ensemble des participants ainsi qu'une symétrie dans les données récoltées, Hollis [155] propose de contrôler les séries de stimuli présentés à chaque participant en respectant les contraintes suivantes :

- Chaque participant doit entendre une et une seule fois chaque son : cela contraint le nombre de stimuli  $N$  à être un multiple de  $P$ , le nombre de choix de la BWS. Cela a aussi pour conséquence d'avoir  $\frac{N}{P}$  essais.
- Une paire de sons apparue une fois ne peut pas ré-apparaître. Chaque participant jugera  $\frac{N}{P} \frac{P(P-1)}{2} = \frac{N(P-1)}{2}$  paires. Sachant que l'ensemble des paires possibles pour les  $N$  stimuli est  $\frac{N(N-1)}{2}$ , cette contrainte ne peut être respectée que si le nombre de participants  $K < \frac{N-1}{P-1}$ . Dans le cas où le nombre de participants est trop grand, certaines paires seront forcément représentées. Augmenter le nombre de participants ne dégradera pas la qualité des scores calculées, mais dans ce cas, cela peut dégrader la symétrie entre les scores de tous les sons, certains auront alors eu plus d'information pour en déduire un score.
- Tous les N-uplets doivent être différents. Comme évoqué précédemment, le nombre de N-uplets est de  $\frac{N!}{P!(N-P)!}$ , et la contrainte précédente assure aussi que celle-ci soit respectée donc cette condition n'est pas contraignante.

Ces conditions imposent alors d'avoir une liste de séquences de stimuli prévue à l'avance, fixée par le nombre de stimuli à tester et le nombre de participants. Chaque participant aura une liste de stimuli précise à évaluer pour permettre d'extraire le maximum d'information de la procédure.

### 4.3.4 | La question du nombre de participants

Classiquement, pour une évaluation DS par exemple, le nombre de participant nécessaire peut être calculé pour un test en fonction de la puissance statistique souhaitée. La puissance statistique représente la probabilité de correctement rejeter une hypothèse par un test statistique. Celle-ci se calcule à partir du risque de première espèce (communément appelé la p-value), l'écart entre les moyennes de deux populations testées, l'écart-type dans ces populations et leur tailles. Une puissance de 0.8 est communément admise comme une bonne valeur de puissance. Le calcul peut donc être inversé pour en déduire un nombre minimum de participants si les autres grandeurs sont connues. Ce calcul est en particulier possible si la variance inter-participant du phénomène mesuré est connue. Les fonctions de scoring ne permettent pas d'avoir une valeur par participant mais seulement une valeur globale. La variance inter-participant ne peut donc pas être calculée dans ce cas, et une puissance statistique ne



**Figure 4.2.**  $\hat{r}^2$  est la moyenne, sur 100 simulations, du coefficient de détermination  $r^2$  entre les scores estimés et les valeurs réelles simulées. La figure de gauche est réalisée pour une BWS à 1000 stimuli, celle de droite pour une BWS à 164 stimuli correspondant à l'expérience présentée dans ce chapitre. Les différentes fonctions de scoring sont calculées : Elo, RW (Rescorla Wagner), Value Learning (Value), Absolute Best Worst (ABW) et Best Worst (BW).

peut pas être déduite au test en fonction du nombre de participants.

Le choix du nombre de participants peut cependant être guidé à travers les simulations d'expériences BWS proposées par Hollis en fonction de l'algorithme choisi et du bruit dans la mesure. Les figures présentées par Hollis sont répliquées ici (fig. 4.2) avec les mêmes paramètres qu'il a choisi, ainsi qu'avec les paramètres de l'expérience que nous avons choisi de mener (164 stimuli). La très grande proximité entre les résultats nous permet d'attester que les figures obtenues fonctionnent aussi pour notre expérience malgré le choix d'un nombre de stimuli différent. On se référera donc à présent entièrement aux figures présentées par Hollis. Cette figure réalisée sans bruit dans la décision des participants illustre une convergence des méthodes à partir de 10 ou 15 participants. Les figures faites par Hollis dans le cas de participants non idéaux, soumis à un bruit aléatoire montrent alors une convergence des méthodes plus tardives aux alentours de 20 ou 25 participants pour l'algorithme *Value Learning*, qui fonctionne le mieux en condition bruitée. Le nombre de 25 participants sera choisi dans notre étude.

### 4.3.5 | Résumé de la section

#### Le choix de la méthodologie *Best-Worst Scaling*

À l'issue du chapitre détaillant les méthodologies d'évaluation perceptive à notre disposition, il a été choisi de prendre la *Best-Worst Scaling* (BWS) qui permet d'évaluer un large corpus de sons avec une plus grande facilité pour les participants. A cela s'ajoute qu'une telle méthodologie permet une évaluation relative des stimuli entre eux en les écoutant par groupe de quatre, ce qui permet d'évaluer plus fidèlement des stimuli à priori assez semblables, comme c'est le cas pour notre étude.

La BWS nécessite cependant une attention particulière lors de la conception de l'expérience, car l'utilisation de fonctions de scoring qui améliore l'information extraite de ce type d'évaluation s'accompagne d'une attribution précise des stimuli entre chaque participant. Ainsi, chaque participant passe une expérience légèrement différente, et c'est la recombinaison de l'ensemble de ces jugements qui permet aux fonctions de scoring de déduire une valeur numérique précise pour les stimuli.

## 4.4 | Méthode

### 4.4.1 | Schéma expérimental

Le choix de la méthodologie BWS nous permet de travailler avec un large corpus de stimuli. Cette caractéristique nous permet alors d'évaluer différents facteurs agissant sur les dimensions à évaluer. Le but de cette expérience est d'évaluer les 3 typologies de masquages présentées au chapitre précédent (Chap. 3) et l'importance de ces typologies relativement à des grandeurs acoustiques classiquement considérées. Pour cela, nous présenterons un nombre important de stimuli en vue d'une analyse statistique sur les groupes de sons, puisque l'analyse statistique sur les participants n'est pas possible avec les scores obtenus par la BWS. En particulier, le schéma expérimental choisi ici est un schéma factoriel complet qui nous permettra à terme de conclure sur l'effet de chacun des facteurs expérimentaux et de leurs interactions. Le premier facteur expérimental sera alors la typologies de masquage (E, A, H). Chaque typologie possède deux niveaux correspondant à des masques différents suivant la même typologie : (E1, E2) pour la typologie énergétique, (A1, A2) pour la typologie attentionnelle et (H1, H2) pour l'habillage. Nous avons déjà vu au précédent chapitre comment les six masques sont créés. A cela s'ajoute le facteur expérimental lié à la présence du défaut dans la situation sonore choisie, qui possède trois niveaux : une absence de sifflement (S0), un sifflement faible (S1) et un sifflement fort (S2). Cette gradation des valeurs du sifflement permet d'éviter d'avoir une tâche de détection pure, mais d'y inclure une quantification de la détectabilité et l'agréabilité lié au défaut. Puis deux facteurs expérimentaux, liés aux propriétés acoustiques des masques sont ajoutés : un paramètre de volume sonore comportant trois niveaux (V0, V1, V2) et un paramètre de brillance comportant lui aussi trois niveaux : brillant (B), neutre (F) et mat (D). Ces deux propriétés acoustiques sont fréquemment associés à des jugements de gêne ou d'agréabilité des sons, les inclure dans l'étude permettra de comparer la magnitude de ces facteurs par rapport à celle du facteur typologie. Les différents paramètres du schéma sont résumés dans le tableau 4.1. L'ensemble des stimuli composés, tels que présentés dans l'expérience sont disponibles sur le compte Soundcloud associé à ce manuscrit en cliquant [ici](https://soundcloud.com/matthieu-duroyon-404568848/albums)<sup>1</sup> où six albums peuvent être trouvés, du nom des masques (E1, E2, A1, A2, H1, H2) et contiennent chacun les stimuli avec toutes les conditions détaillées.

<sup>1</sup> <https://soundcloud.com/matthieu-duroyon-404568848/albums>

Typologie de Masque (M)	Brillance (B)	Volume (V)	Sifflement (S)
E1,E2	D, F, B	V0, V1, V2	S0,S1,S2
H1,H2	D, F, B	V0, V1, V2	S0,S1,S2
A1,A2	D, F, B	V0, V1, V2	S0,S1,S2

**Table 4.1.** Schéma expérimental factoriel complet

Cette paramétrisation permet, par le schéma factoriel complet de quantifier l'impact du choix des masques par rapport à des grandeurs psychoacoustiques tels que la sonie et la brillance. On a ainsi  $6 \times 3 \times 3 \times 3 = 162$  stimuli à évaluer dans chacune des deux expériences. De plus, les deux stimuli de références sans masques ont été ajoutés : ils correspondent à la situation sonore idéale, sans sifflement et la situation sonore non-idéale, avec le sifflement. Les algorithmes de scoring présentés et testés par Hollis [155] ont notamment prouvé une meilleure corrélation aux valeurs réelles en incluant deux joueurs fictifs (*Dummy Players*). Le premier est systématiquement évalué comme le meilleur dans chacun des essais où il est présenté, le deuxième est systématiquement le pire dans les expériences simulées dans l'étude. Ces joueurs fictifs servent alors à borner les valeurs de scores possibles et à augmenter la précision des scores. Dans le cas d'une expérience non simulée comme la nôtre, il est délicat d'inclure des "sons fictifs", les deux stimuli de références, idéal et non-idéal, rempliront cette tâche. On s'attend en effet, à ce que les scores obtenus pour ces stimuli encadrent tous les autres et puissent servir à étalonner les scores autour de ces deux valeurs de références.

Un total de  $6 \times 3 \times 3 \times 3 + 2 = 164$  stimuli, qui est divisible par 4 est constitué. La méthodologie BWS peut être utilisée de manière optimale avec 41 4-uplets de sons, en présentant une et une seule fois chaque stimulus. De plus, l'inégalité  $K < \frac{N-1}{P-1}$  est bien respectée puisque le nombre de participants est  $K = 25$  et que le nombre de sons est  $N = 164$ , il sera donc possible de correctement construire les séquences de stimuli à présenter aux participants en respectant les contraintes citées.

## 4.4.2 | Construction des stimuli

### 4.4.2.1 | Facteur Masque

Les six niveaux du facteur Masque ont été présentés dans le chapitre précédent (Chap. 3) : E1, E2, H1, H2, A1, A2.

### 4.4.2.2 | Facteur Défaut

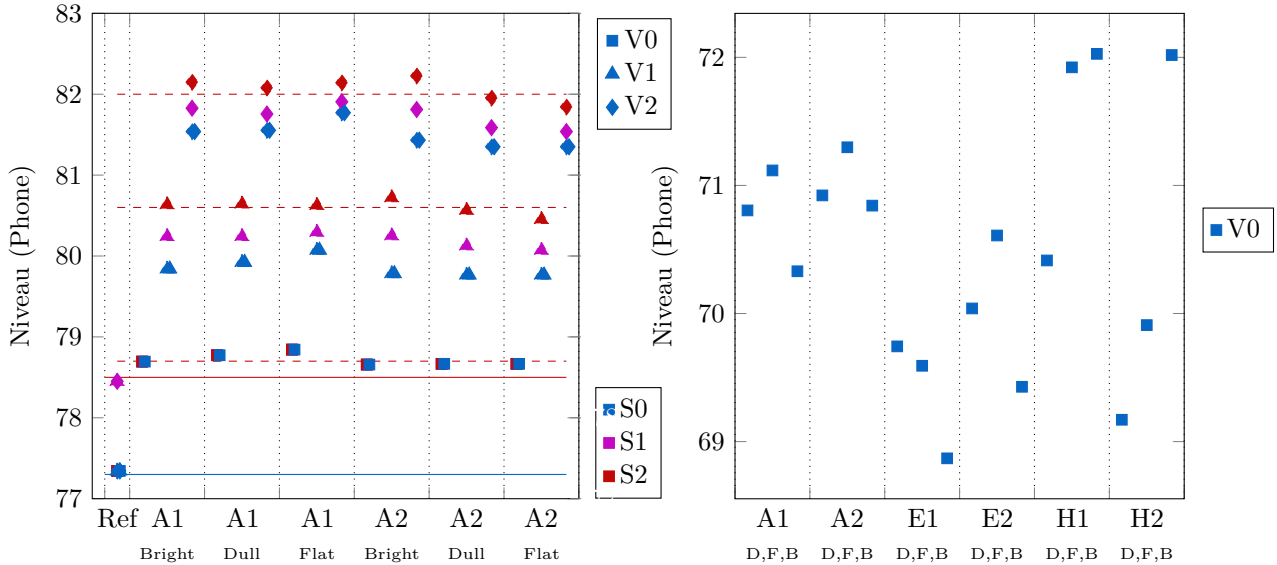
Les trois niveaux de sifflement sont réalisés en assemblant les masques composés à trois scènes avec un sifflement différent à l'aide du module "situation sonore" du synthétiseur. Chaque son de ventilation est le même, auquel est ajouté un son de sifflement avec une émergence de 15 dB (S2), avec une émergence de 10 dB (S1) ou sans émergence (S0).

### 4.4.2.3 | Facteur Volume

Le paramètre volume sonore agit fortement sur la perception, et notamment sur l'agréabilité, il est donc important que tous les stimuli présentés soient d'intensités sonores perçues égales au sein d'un même niveau (V0, V1, V2). Pour réaliser cette égalisation, il ne serait pas cohérent de modifier le volume global de chaque stimuli puisqu'ils sont composés de trois signaux différents :  $s_{\text{Ventilation}} + s_{\text{Sifflement}} + s_{\text{Masque},i}$ . Cela résulterait en un changement de situation sonore, puisque certains stimuli auraient des sons de

ventilations et de sifflement à des volumes différents. L'égalisation en sonie est obtenue en agissant uniquement sur le volume du masque. Elle a été réalisée en respectant l'équation 4.3.

$$N_{V0} = \text{Sonie}(s_{\text{Ventilation}} + s_{\text{Sifflement}} + x_i s_{\text{Masque},i}) \quad (4.3)$$



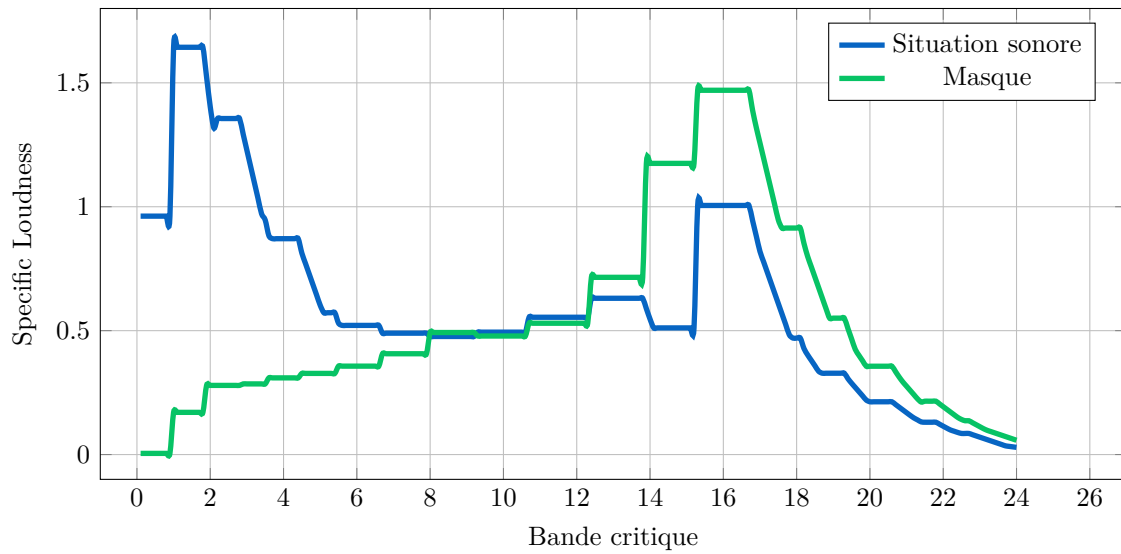
**Figure 4.3.** Niveau de sonie calculée des scènes complètes (à gauche) et des masques seuls (à droite) en Phones, pour différents facteurs expérimentaux. La figure de gauche détaille tous les facteurs expérimentaux pour deux masques A1, A2 en comparaison avec les deux scènes de référence sans sons ajoutés. La figure de droite représente les valeurs en sonie des masques seuls dans la condition (V0), qui sont différentes dans chaque condition pour que les scènes globales soient à iso-sonie.

Pour chaque masque  $i$ , on cherche la valeur de  $x_i$  qui égalise la sonie mesurée de la scène totale par rapport à un niveau de référence  $N_{V0}$  (Eq. 4.3). Ce procédé est réalisé numériquement, à l'aide de la bibliothèque Mosquito [158]<sup>2</sup> qui calcule le niveau perçu à l'aide de la norme ISO532-1. Ainsi, toutes les scènes sonores de niveau V0 sont à iso-sonie  $N_{V0}$  en ayant le même signal de ventilation et de sifflement, mais chacun des masques ajoutés a un volume différent. La figure 4.3 représente pour illustration la sonie des différentes déclinaisons de la scène acoustique avec le masque A1. L'augmentation du niveau sonore des scènes avec masques par rapport aux scènes de référence sans défauts ni masques est notable. Comme le niveau du sifflement varie dans les conditions S0, S1, S2, conformément au schéma expérimental, un écart est visible sur le graphique pour ces différentes conditions liées au sifflement. La figure de droite affiche le niveau des masques seuls, cela permet de démontrer à posteriori l'intérêt d'une égalisation de la scène globale et pas des masques seuls, puisque pour une même iso-sonie fixé, V0 dans le cas de la figure, les niveaux des masques ajoutés sont très différents en fonction de leurs paramètres expérimentaux.

Le choix du niveau d'iso-sonie  $N_{V0}$  a donc été fait en s'assurant que tous les masques soient perçus. La sonie calculée pour chacun des masques dans la condition V0 est strictement supérieure à celle dans la scène de référence. De la même manière, pour choisir le volume maximum du masque, V2, celui-ci est calibré à partir du masque énergétique E1. Le but est alors de trouver le volume du masque E1 pour que le sifflement soit effectivement masqué. Il peut en effet être théoriquement calculé à partir des travaux de Zwicker et Moore [120, 159].  $N_{V2}$  est donc le niveau de la scène tel que la sonie spécifique du masque soit supérieur au sifflement dans la bande de Bark de celui-ci. Comme l'illustre la figure 4.4, ce volume n'est pas pris au minimum, mais légèrement plus élevé. Cela permet d'avoir un volume V1,

<sup>2</sup> <https://github.com/Eomys/MoSQITo>

intermédiaire à V0 et V2, ayant une différence en sonie supérieure à une valeur de jnd.



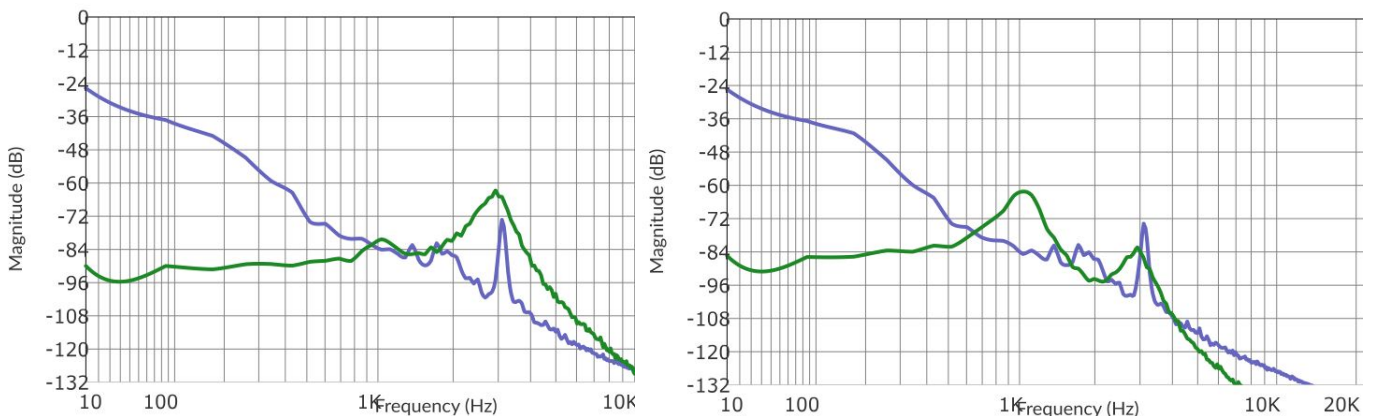
**Figure 4.4.** Représentation de la densité de sonie du masque et de la situation sonore en bande de Bark, pour expliciter le caractère masquant du masque énergétique E1 au volume V2, de brillance B.

Pour résumer, l'égalisation en sonie a été réalisée en modifiant uniquement le volume du masque en respectant ces 3 critères :

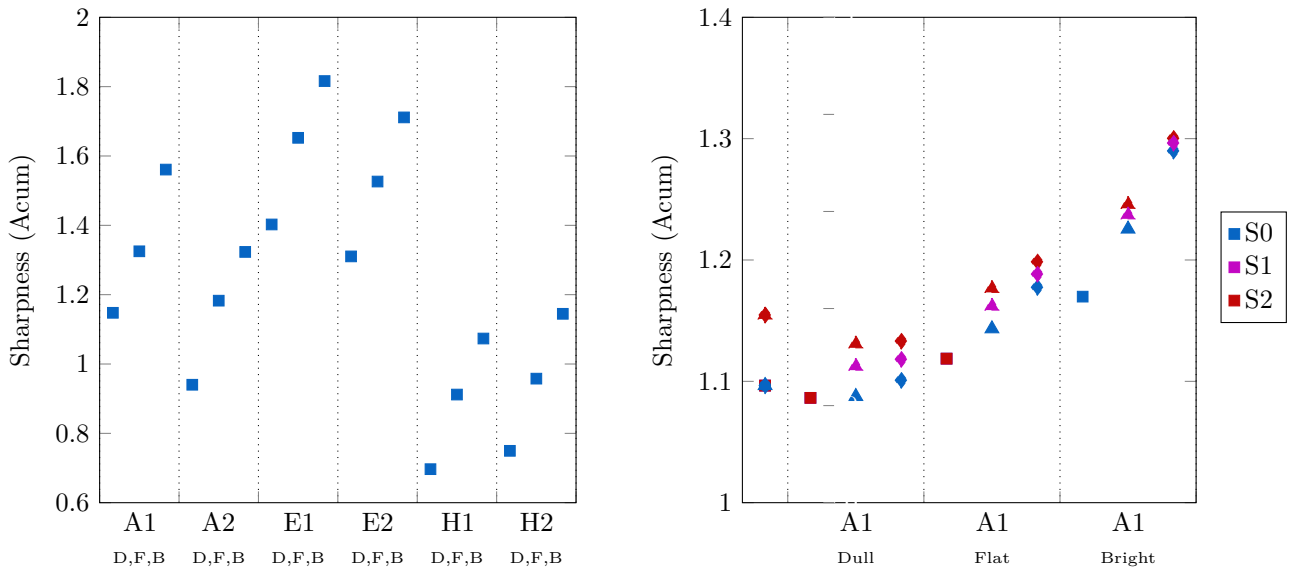
- Le volume minimum du masque doit lui permettre d'être perçu dans la situation sonore.
- Le volume maximum du masque, dans le cas du masque énergétique E1, doit lui permettre de théoriquement masquer le sifflement.
- Les scènes sonores aux volumes V0, V1, V2 doivent avoir au moins une valeur de jnd de différence les unes avec les autres, pour s'assurer que la différence est suffisamment importante.

#### 4.4.2.4 | Facteur Brillance

Le facteur Brillance est aussi important, car il s'agit d'une des dimensions les plus importantes dans l'évaluation du timbre d'un son. Le synthétiseur (Chap. 3) est notamment conçu pour pouvoir effectuer simplement ce type de modification sur le timbre. En effet, une fois un masque conçu, une variation en brillance perçue peut facilement être obtenue en changeant les paramètres du bloc Octaver, notamment le niveau de chacune de ces octaves. Le spectre de ces différents masques, en superposition du son de



**Figure 4.5.** Spectres des variations B *Bright* (à gauche) et D *Dull* (à droite) du masque énergétique E1. La différente répartition énergétique illustre l'impact du changement de paramétrage  $K_D$  et  $K_F$ .



**Figure 4.6.** Valeurs calculées de brillance par l'indicateur psychoacoustique calculant la *sharpness* sur les masques seuls à gauche pour l'ensemble des combinaisons des facteurs Masques (A1, A2, E1, E2, H1, H2) et Brillance (Dull, Flat, Bright) possibles. A droite la brillance est représentée pour la scène complète uniquement pour le masque A1, avec les déclinaisons de Défaut (S0, S1, S2) et de Brillance (Dull, Flat, Bright).

la situation, est affiché sur la figure de gauche (Fig. 4.5) pour la déclinaison brillante du masques, et sur la figure de droite pour la déclinaison mat. En reprenant la présentation faite du bloc Octaver, les trois déclinaisons du masque énergétique E1 selon le facteur Brillance, notées E1B, E1F et E1D sont réalisées avec les paramètres suivants, décrit au chapitre 3 :

$$\begin{aligned}
 N &= 2, \\
 B &= [1, 1], \\
 V &= [1, 3], \\
 K_D &= [0, -10], K_F = [0, 0], K_B = [-10, 0], \\
 F &= [1000, 3000], \\
 Q &= [0.9, 0.7],
 \end{aligned}$$

De la même manière que pour la sonie, l'indicateur psychoacoustique lié à la brillance, la *sharpness*, telle que présentée par [159] est calculé. La figure 4.6 représente les valeurs calculées de *sharpness* en acum, et permet aussi de s'assurer que les différences entre les masques sont équivalentes et de vérifier que l'espace des deux dimensions du niveau sonore et de la brillance sont correctement pavés. La partie droite de la figure représente les valeurs de *Sharpness* pour les scènes complètes et plus seulement pour le masque. Le timbre de la scène change faiblement, mais le timbre du masque oui, les modifications de résultats causés par ce facteur seront donc bien du à la modification du timbre du masque.

### 4.4.3 | Participants

Pour cette double expérience, 50 participants ont été recrutés ( 50 % d'hommes et 50 % de femmes) pour y prendre part, à raison de 25 participants par expérience. Cette étude a été approuvée par le comité éthique de l'Institut européen d'administration des affaires (INSEAD), et la collecte des données a été réalisée au sein du laboratoire de science comportementale de l'INSEAD Sorbonne Université. Le laboratoire a été aussi en charge de répartir les participants recrutés dans chaque expérience. Chaque participant était ainsi volontaire pour l'étude et rémunéré pour sa participation à un taux horaire fixé par l'institut. Chaque participant est aussi informé de l'intérêt et la durée approximative de l'étude au travers d'une feuille de consentement qu'il est invité à signer avant de prendre part à l'expérience. Le schéma 4.7 résume le déroulement de cette expérience.

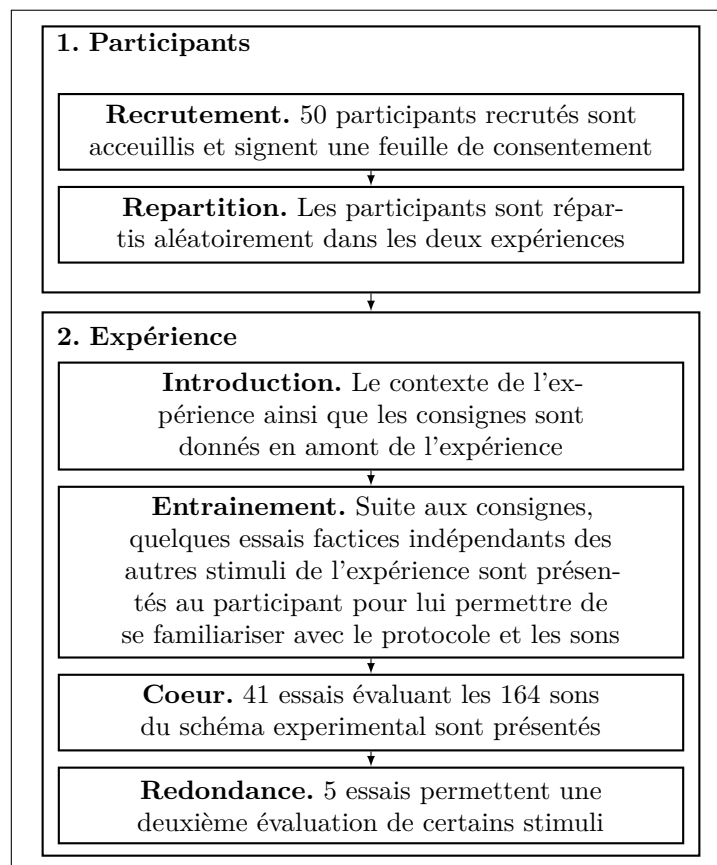


Figure 4.7. Déroulement de l'expérience

### 4.4.4 | Réalisation technique

Pour la réalisation de cette expérience, la librairie JsPsych a été utilisée.<sup>3</sup> Cette librairie développée en Javascript et en open-source l'a été à dessein pour les expériences psychophysiques [160] et est fréquemment utilisée dans la littérature scientifique. Elle dispose ainsi d'une communauté de chercheurs mettant à jour et proposant des bibliothèques à partir de cette librairie pour des expériences précises. Par exemple Matthieu Fraticelli,<sup>4</sup> a développé et proposé la réalisation de plusieurs paradigmes expérimentaux classiques avec JsPsych, dont une trame pour la BWS qui est reprise ici. On peut aussi citer les travaux de Donhauser et Klein [161] qui proposent en accès libre<sup>5</sup> une série d'applications expé-

<sup>3</sup> <https://www.jspsych.org/v7/>

<sup>4</sup> <https://github.com/MatthieuFra/jsPsychPDS>

<sup>5</sup> <https://peterdonhauser.com/publication/audio-tokens/>

mentales. Les schémas expérimentaux complets proposés ne seront pas réutilisés mais une partie des outils construits et accessibles dans cette bibliothèque sera réutilisée et modifiée dans notre expérience.

L'utilisation d'une librairie Javascript open-source telle que JsPsych nous permet alors de lancer nos expériences à partir d'un fichier html. Il est donc possible de réaliser l'expérience une fois implémentée sur n'importe quelle machine possédant un navigateur web. Cela permet une grande souplesse dans le déroulement de l'expérience. En particulier, il est possible de stocker l'expérience sur un serveur *Just Another Tool for Online Studies* (JATOS) en ligne<sup>6</sup> hébergé à l'IRCAM. Il est alors possible d'accéder à l'expérience depuis n'importe quelle machine seulement avec un lien :

— Déteçtabilité : <https://jatos.ircam.fr/publix/W0fQdZyJC21>

— Agréabilité : <https://jatos.ircam.fr/publix/zgO3LTXsvJ3>

Les liens partagés ici sont ainsi les mêmes liens que ceux effectivement utilisés par les participants. Dans notre cas, cette diffusion a deux intérêts principaux. le premier : dans un but pédagogique, permettre de montrer le déroulement de l'expérience à des personnes externes intéressées pas le projet. Le deuxième, est de faciliter la mise en œuvre de l'expérience dans l'institut chargé de recruter les participants, l'INSEAD. En effet tous les problèmes de compatibilités entre les machines, ou d'installation de l'expérience ont pu être évités en permettant le passage de l'expérience avec un simple lien et permettre une centralisation des données, puisque tout est stocké sur un serveur en ligne sécurisé.

#### 4.4.5 | Déroulement de l'expérience

Chaque participant est ensuite placé dans une cabine individuelle isolée acoustiquement de l'extérieur. Chaque cabine possède une station de travail similaire, avec le même casque audio Beyerdynamic DT 770 Pro, branché directement à la station de travail. Pour s'assurer que les niveaux d'écoute soient les mêmes pour chaque participant, en amont de l'expérience le niveau sonore de restitution est mesuré et calibré dans chaque cabine et pour chaque casque à l'aide d'un sonomètre B&K 2250 et d'un support spécialement conçu pour y poser les casques comme représenté sur la figure 4.8. De cette manière, le niveau sonore de chaque casque peut être calibré et réglé pour s'assurer que chaque participant passe l'expérience avec le même niveau à  $\pm 0.5$  dB.

---

<sup>6</sup> <https://www.jatos.org/Whats-JATOS.html>



Figure 4.8. Calibration du niveau de restitution du casque au sonomètre.

#### 4.4.6 | Résumé de la section

##### Méthode

Dans notre étude, deux expériences jumelles seront alors menées pour évaluer les masques conçus. La première visant à évaluer le jugement d'agrément perçu à l'écoute des situations sonores recomposées, et la seconde évaluant la détectabilité du défaut de sifflement, présent ou non dans la scène. Pour cela, un schéma expérimental factoriel complet est suivi. La situation sonore traitée dans cette étude est augmentée par des masques des trois typologies (A, E, H) possédant deux niveaux du facteur Masque (1, 2), auxquels sont ajoutés un facteur Volume ayant trois niveaux (V0, V1, V2), un facteur Défaut ayant trois niveaux (S1, S2, S3) et un facteur Brillance ayant lui aussi trois niveaux (D, F, B), pour un total de  $6 \times 3 \times 3 \times 3 = 162$  stimuli auxquels on ajoute les deux scènes de référence, avec et sans sifflement.

Deux expériences par BWS (accessibles aux liens suivants : [Detectabilité](https://jatos.ircam.fr/publix/W0fQdZyJC21),<sup>a</sup> [Agréabilité](https://jatos.ircam.fr/publix/zgO3LTXsvJ3)<sup>b</sup>) sont alors construites pour 25 participants chacune, ce qui permet d'évaluer les 164 stimuli. L'expérience se déroule à l'INSEAD au casque dans des cabines isolées, où le niveau de restitution a été calibré.

<sup>a</sup> <https://jatos.ircam.fr/publix/W0fQdZyJC21>

<sup>b</sup> <https://jatos.ircam.fr/publix/zgO3LTXsvJ3>

Stimuli	Expérience	Best	Worst	Total	Score
Idéal	DéTECTABILITÉ	0	20	25	0.231
Non-idéal	DéTECTABILITÉ	14	5	25	0.627
Idéal	AGRÉABILITÉ	18	2	25	0.748
Non-idéal	AGRÉABILITÉ	15	0	25	0.771

**Table 4.2.** Classement des scènes sonores idéales et non-idéales, pour vérifier leurs rôles de *Dummy players*

## 4.5 | Résultats

### 4.5.1 | Sons fictifs

L'intérêt de présenter des sons fictifs, *Dummy players* [155], dans la conception de l'expérience BWS a été présenté plus tôt (Sec. 4.4.1). Dans notre cas, il s'agit des sons associés à la scène idéale et non-idéale, n'ayant pas été augmentée. Les premiers résultats du test de BWS sont présentés dans le tableau 4.2, qui comportent le nombre de fois qu'une des deux scènes de référence, idéale ou non-idéale, a été sélectionnée comme la meilleure ou la pire dans les 4-uplets ou elles ont été présentées. A cela est ajouté le score final obtenu par l'algorithme de scoring détaillé plus tôt. L'ensemble des scores bruts obtenus pour chacun des stimuli et des expériences est présenté lui en annexe A (Annexe A).

Ce tableau ainsi que les scores présentées en annexe (Annexe A) en complément nous permettent de voir que le rôle de *Dummy players* des deux stimuli, qui avaient vocations à encadrer les résultats, n'est pas respecté. Dans le cas de l'expérience de détectabilité, on s'approche de ce qui était supposé puisque la scène idéale a été jugée 20/25 fois avec la "pire" détectabilité du sifflement, et la scène non-idéale 14/25 fois avec la meilleure détectabilité. Mais le caractère systématique du *Dummy players* qui permet la plus grande précision des algorithmes de scoring n'a manifestement pas été observé. Pour l'expérience d'agrément, l'échec est encore plus marquant, puisque de manière assez surprenante les deux scènes ont un score similaire.

### 4.5.2 | Comparaison inter-expérience

Le déroulement de l'expérience BWS requiert que chaque participant évalue une séquence de sons précise, différente de celles des autres participants. Suite au désistement de quelques candidats et à l'erreur dans l'attribution des numéros de participant aux remplaçants, certaines séquences précises dans l'expérience d'agrément n'ont pas été évaluées. Ainsi, l'expérience de détectabilité a bien été effectuée au complet avec les 25 séquences correctement évaluées par 25 participants différents. En revanche, deux séquences, correspondant aux participants numérotés 20 et 21 n'ont pas été collectées pour l'expérience d'agrément.

Les retours spontanés des participants ont révélé que l'expérience de détectabilité a été estimée comme fatigante et parfois difficile, alors qu'aucun retour de ce type n'a été formulé pour l'expérience d'agrément. La raison même de cette différence s'explique probablement par le fait que la tâche de détectabilité soit plus objective et que certains stimuli ne présentent aucun défaut à détecter, alors que le jugement d'agrément est plus subjectif.

Cette différence peut être observée dans les résultats collectés. L'expérience d'agrément a ainsi duré en moyenne  $\mu = 25.5$  s par essais et  $\mu = 32.4$  s pour l'expérience de détectabilité. Ces moyennes ont été calculées uniquement sur les essais enregistrés pour l'expérience, c'est-à-dire sans prendre en compte l'introduction à l'étude ou les essais d'entraînement. Les écarts-types calculés pour ces deux

séries de données sont  $\delta = 8.66$  s et  $\delta = 9.31$  s. Les variances sont suffisamment proches pour garder l'hypothèse d'homoscédasticité. Un test de Student bilatéral nous permet alors de rejeter l'hypothèse  $\mathbb{H}_0$  : "Les deux expériences ont la même durée" [t(46), p=1.33 × 10<sup>-2</sup>, d= 0.76]. Cela traduit donc un effet significatif qui différencie les deux expériences.

### 4.5.3 | Cohérence et conformité des participants

Cette différence significative sur la durée de l'expérience laisse penser qu'une des deux expériences a pu poser problème et que les résultats ainsi collectés ne sont pas pertinents. La cohérence et la conformité des participants sont alors calculées et comparées entre les deux expériences.

#### 4.5.3.1 | Cohérence inter-participant

Dans leur utilisation de la BWS, Rosi et al. [157] utilisent une mesure de conformité (traduit de l'anglais *compliance*) aussi utilisée dans [155] qui donne la proportion de jugement fait par un participant  $k$  en accord avec les résultats sur l'ensemble des participants. Cette conformité  $c_k$  est calculée pour chaque participant  $k$ , en comparant les cinq inégalités obtenues pour les  $N$  essais (Eq. 4.1), au score moyen obtenu avec l'ensemble des participants pour chacun de ces stimuli, noté  $\hat{s}_i$ . Pour un essai  $i \in [1, N]$ , où  $b_i$  est le meilleur stimulus,  $w_i$  le pire, et  $s_{1,i}, s_{2,i}$  les deux non sélectionnés. Les scores de conformités  $c_{i,j}, j \in [1, 5]$  sont calculés comme suit :

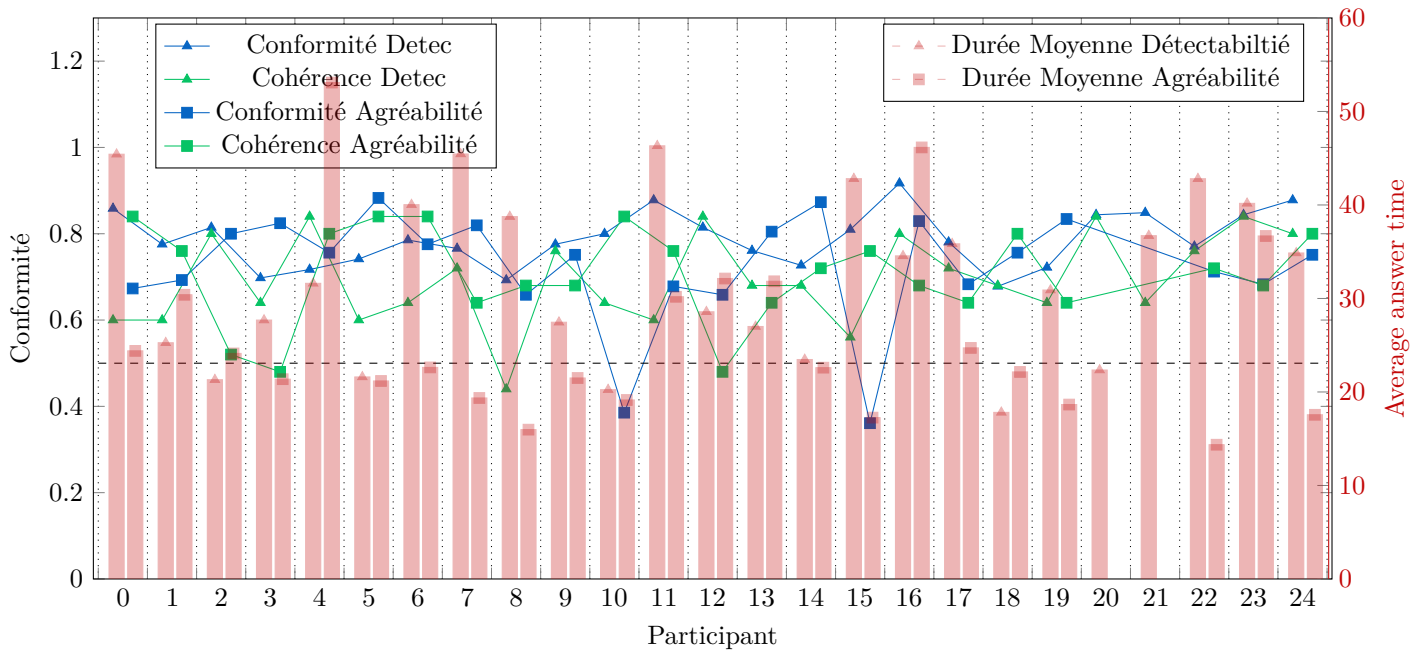
$$\begin{aligned} c_{i,1} = 1 &\iff (b_i > s_{1,i}) \wedge (\widehat{b}_i > \widehat{s}_{1,i}), \\ c_{i,2} = 1 &\iff (b_i > s_{2,i}) \wedge (\widehat{b}_i > \widehat{s}_{2,i}), \\ c_{i,3} = 1 &\iff (b_i > w_i) \wedge (\widehat{b}_i > \widehat{w}_i), \\ c_{i,4} = 1 &\iff (w_i < s_{1,i}) \wedge (\widehat{w}_i < \widehat{s}_{1,i}), \\ c_{i,5} = 1 &\iff (w_i < s_{2,i}) \wedge (\widehat{w}_i < \widehat{s}_{2,i}). \end{aligned}$$

La conformité pour un participant  $k$  se calcule alors à partir de la conformité moyenne pour chacune des  $5N$  inégalités mesurées au long de l'expérience (Eq. 4.4).

$$c_k = \frac{1}{5N} \sum_{i=1}^N \sum_{j=1}^5 c_{i,j} \quad (4.4)$$

Une conformité moyenne de 78.8 % est calculée pour l'expérience de détectabilité et 72.5 % pour celle sur l'agrément. La différence entre les deux expériences est statistiquement significative [t(44), p=3.36 × 10<sup>-2</sup>, d=0.64].

Si l'intégralité des essais d'un participant est conforme au score moyen, la conformité est de 100 %. En répondant au hasard, la conformité moyenne est alors de 50 %. Or, dans l'expérience d'agrément (Fig. 4.9), deux participants ont une conformité faible (0.38 et 0.36), nettement en dessous du hasard. Les participants avec une conformité plus faible que 50 % (deux participants pour le jugement d'agrément, zéro pour la détectabilité : 0.38 et 0.36) sont exclus. Une fois les données des participants avec une conformité trop faible écartées, la conformité moyenne calculée est de 78.8 % pour l'expérience de détectabilité et de 75.7 % pour l'expérience d'agrément. Le test statistique est répliqué, [t(44), p=0.133] et une fois "nettoyée", la différence de conformité trouvée n'est plus statistiquement significative. Cela montre que les deux éléments ayant exprimé un fort désaccord avec la moyenne expliquaient cette



**Figure 4.9.** Conformité et cohérence de chaque participant dans chacune des 2 expériences. On ajoute le temps moyen pris par le participant pour répondre aux jugements

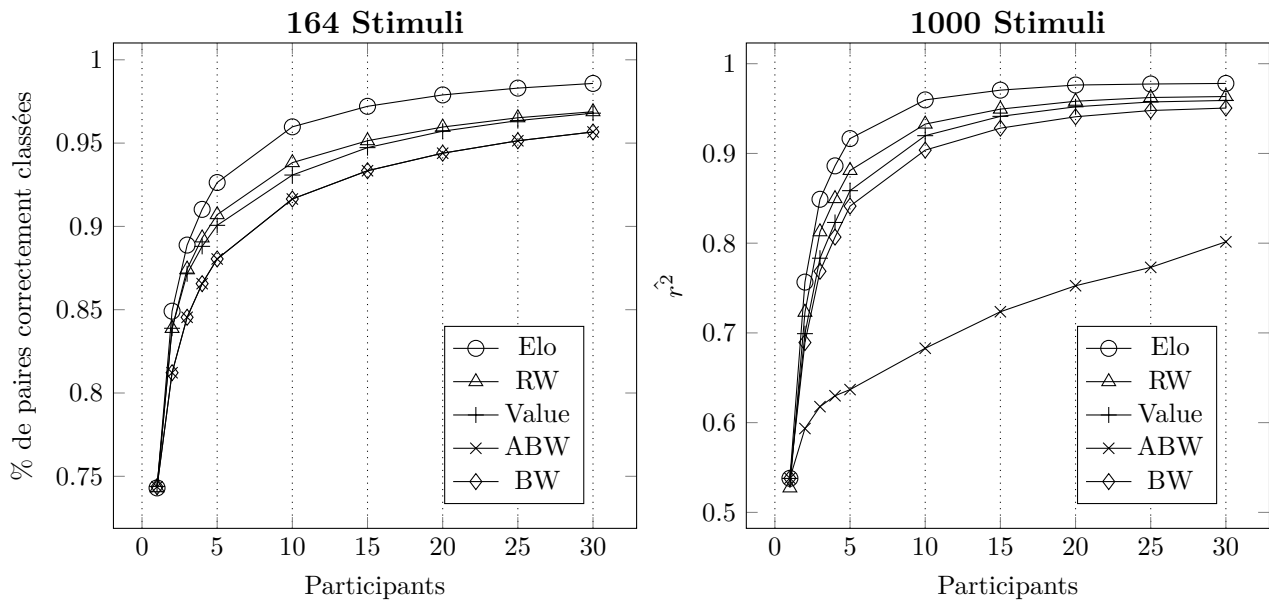
différence.

La faible littérature sur l'utilisation d'une méthodologie BWS pour des tests de ce type permet de comparer ces valeurs uniquement à l'étude de Rosi et al. [157], où 20 participants ont évalué 100 sons pour une conformité calculée de 83 %. La différence de conformité entre l'expérience d'agréabilité avec les valeurs de conformité moyennes pour l'expérience de détectabilité ou celles reportées de l'étude [157] peut s'expliquer par le fait que l'agréabilité est un jugement hédonique, pouvant être subjectif, alors que la détectabilité et le jugement de la brillance d'un son, utilisés par Rosi et al., sont plus objectifs.

#### 4.5.3.2 | Cohérence intra-participant

Ce calcul de conformité inter-participant ne permet cependant pas de conclure sur la qualité intrinsèque des réponses. Pour obtenir cette information, un calcul de cohérence intra-participant est requis. Pour cela, cinq 4-uplets de sons déjà présentés à chaque participant sont présentés à nouveau pour évaluation. Les réponses fournies sont alors comparées aux évaluations précédemment effectuées pour calculer la cohérence du participant. Cette cohérence est calculée de la même manière que la conformité, c'est-à-dire en comparant inégalité par inégalité.

La figure 4.9 illustre les scores de cohérence et de conformité de chaque participant et de chaque expérience, ainsi que le temps moyen pris pour évaluer un essai. L'axe des abscisses donne le numéro du participant, identique, pour les deux expériences. En effet, bien que le participant  $n$  de l'expérience de détectabilité est différent du participant  $n$  de l'expérience agréabilité, ils ont chacun évalué exactement la même séquence de stimuli.



**Figure 4.10.** Les deux figures sont réalisées pour une BWS à 1000 stimuli. La figure de droite calcule  $\hat{r}^2$ , la moyenne sur 100 simulations, du coefficient de détermination  $r^2$  entre les scores estimés et les valeurs réelles simulées. La figure de gauche représente le pourcentage de paires des stimuli correctement ordonnés avec les scores estimés par rapport aux valeurs réelles.

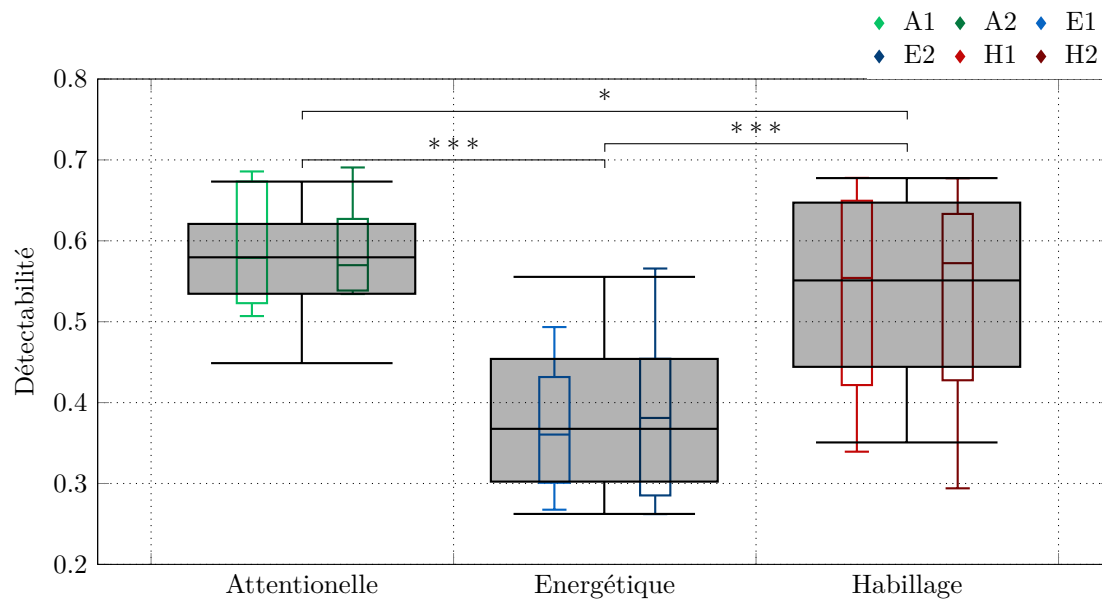
#### 4.5.3.3 | Une comparaison intra-participant avec le score individuel ?

La mesure de cohérence intra-participant permet d'avoir une première métrique. Cependant cette cohérence est seulement possible en présentant exactement le même 4-uplet au participant. Une telle mesure ne permet pas, par exemple d'identifier, parmi les réponses d'un seul participant une donnée aberrante (*outlier* comme cela pourrait être fait pour une mesure par *rating scale*).

La question posée ici est de savoir s'il est possible d'utiliser les fonctions de scoring présentées plus tôt sur un seul participant. Les résultats [155] obtenus sur simulation indiquent une faible corrélation entre les valeurs réelles et calculées par scoring pour un nombre faible d'essais, comme on aurait avec un seul participant.

Cependant, cette faible corrélation ne nous permet pas de conclure sur la qualité du tri effectué par l'algorithme sur le corpus de stimuli. Or, c'est précisément cette information qui est utilisée dans le calcul de conformité. En effet, pour vérifier qu'un participant est conforme avec lui-même, nous n'avons pas besoin d'une bonne approximation du score, mais seulement que la relation d'ordre soit la bonne (Eq. 4.4). Nous proposons alors de répliquer la figure présentée par Hollis en incluant cette fois-ci une métrique indiquant si l'ordre est respecté entre les données calculés par le score et les données réelles. La figure (fig. 4.10) compare donc le score de corrélation moyen obtenu sur des simulations avec le pourcentage de paires correctement ordonnées.

Ce que nous montre alors cette figure, c'est qu'avec les résultats d'un seul participant, on est capable au mieux d'avoir 75 % des classements corrects. Ainsi la mesure de conformité d'un participant serait bornée entre 50 %, si le participant répond au hasard, et 75 % si le participant répond parfaitement. Ce résultat révèle alors que l'utilisation du score individuel est difficilement exploitable pour l'analyse de cohérence intra-participant. Le graphique présenté montre cependant la bonne performance des algorithmes puisque 5 participants suffisent à obtenir le classement de 164 stimuli correct à 90 %.



**Figure 4.11.** Boîtes de Tukey pour l'expérience de détectabilité. Les boîtes de couleurs sont calculés sur les 27 stimuli propres à chaque masque. Les boîtes en gris englobent les résultats pour les deux masques au sein de chaque typologie notée en abscisse : Attentionnelle, Énergétique, Habillage.

Cette réflexion autour de la détection de données aberrantes s'articule autour de la volonté d'utiliser les informations de chaque individu dans l'analyse de la BWS et est plus amplement développé au chapitre 5.

#### 4.5.4 | Résultats sur l'expérience de détectabilité : pertinence des typologies

Une fois ce travail préliminaire achevé, qui permet de s'assurer de la qualité des données collectées a priori, l'analyse des résultats peut maintenant être menée. La figure en boîte de Tukey 4.11 illustre les scores obtenus par masque. Pour obtenir ces résultats, la boîte de Tukey est calculée sur les 27 stimuli dans chaque masque (E1, E2, A1, A2, H1, H2). Une deuxième représentation, superposée à la première et affichée en gris, est cette fois calculée par typologie. Elle agrège les résultats des 54 stimuli, regroupant les déclinaisons des deux masques de chaque typologie. Cette représentation met en évidence la proximité des résultats de chaque masque au sein d'une même typologie. On remarque en particulier le caractère apparemment différent entre chaque typologie pour diminuer la détectabilité du défaut.

Pour compléter cette représentation, nous effectuons une série de test de Student bilatéraux par paire visant à tester cette hypothèse reformulée ici : "Le Masque 1 de la typologie  $i$  a un score moyen différent du Masque 2 de la même typologie". Ces tests sont effectués sur le vecteur de valeurs constitué par les scores des 27 stimuli pour chaque masque. Les masques E1 et E2 ne sont pas statistiquement différents [ $t(52) = -0.039$ ,  $p = 0.969$ ], ni le masque H1 et H2 [ $t(52) = 0.745$ ,  $p = 0.459$ ], ni le masque A1 et A2 [ $t(52) = 0.266$ ,  $p = 0.791$ ] pour cette expérience.

Les différences entre les typologies sont maintenant explorées. Des tests de Student par paires avec les données regroupées par typologie sont maintenant effectués, ce qui augmente le nombre de degrés de libertés (DDL) du test. Les typologies A et E sont déterminées comme différentes [ $t(106) = -11.2$ ,  $p = 9.9 \times 10^{-20}$ ,  $d = -2.16$ ], E et H sont différentes [ $t(106) = -7.26$ ,  $p\text{-value} = 6.8 \times 10^{-11}$ ,  $d = -1.39$ ], et H et A sont différentes [ $t(106) = -2.07$ ,  $p\text{-value} = 4.2 \times 10^{-2}$ ,  $d = 0.39$ ].

Facteur	Somme des carrés	DDL	P-Value	Puissance
Masque	1.14	5	$2 \times 10^{-45}$	0.99
Volume	$4.2 \times 10^{-3}$	2	0.28	0.06
Brillance	0.02	2	$1.1 \times 10^{-3}$	0.10
Défaut	1.0	2	$6.0 \times 10^{-45}$	1.0
Masque×Vol	0.10	10	$1.3 \times 10^{-7}$	0.11
Masque×Brill	0.05	10	$1.3 \times 10^{-3}$	0.08
Masque×Déf	0.10	10	$4.0 \times 10^{-7}$	0.11
Brill ×Déf	0.09	4	$5.0 \times 10^{-9}$	0.17
Vol× Déf	0.06	4	$8.0 \times 10^{-7}$	0.14
Résidu	0.17	108		
<b>Facteur</b>	<b>2.75</b>	<b>161</b>		

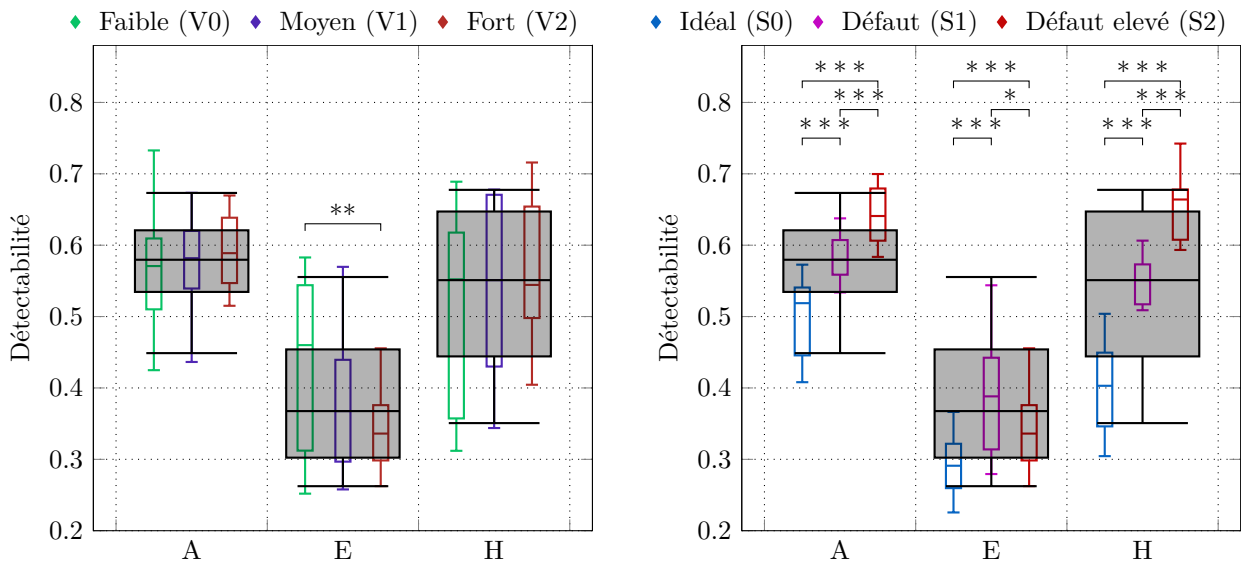
**Table 4.3.** Table ANOVA Pour l'expérience de détectabilité. Parmi les interactions, seuls les facteurs significatifs sont reportés dans la table.

Ces résultats sont interprétés comme une première validation des hypothèses autour de la construction des masques. En effet, il n'y a pas de différences significatives entre les deux masques au sein d'une même typologie puisque il n'y a pas de différences significatives trouvées. En revanche, les différences sont bien significatives pour le facteur typologie, ce qui permet de conclure qu'elles sont bien différentes les unes des autres et que ce facteur est pertinent. Ces différences significatives sont représentées sur la figure 4.11 par le symbole (\*\*\*) pour une p-value inférieure  $p < 10^{-3}$ , par (\*\*) pour une p-value inférieure à  $p < 10^{-2}$  et (\*) pour une p-value juste significative  $p < 0.05$ .

#### 4.5.5 | Résultats sur l'expérience de détectabilité : analyse des facteurs

Le schéma expérimental factoriel complet se prête à une analyse par ANOVA des résultats. Comme mentionné plus tôt, la table ANOVA, ou tout autre test statistique, ne peut pas être calculée en incluant les participants dans les DDL. De ce fait, il est possible de calculer la significativité des différences entre deux groupes de stimuli ayant suffisamment de degrés de libertés, mais ne peut pas être calculée entre deux stimuli seuls. La table ANOVA 4.3 nous permet de voir les facteurs, ainsi que les interactions de facteurs expliquant le plus de variance parmi les jugements des participants. Les paramètres expliquant le plus de variance sont alors identifiés : le facteur Volume (somme des carrés =1.14, soit 41.6 % de la variance totale expliquée), Défaut (36 %), et les interactions Masque × Volume (3.8 %) Masque × Défaut (3.1 %). Ces différents paramètres agissant sur la détectabilité sont maintenant représentés figure 4.12, en agrégeant les résultats par typologies de masquages. Les interactions entre facteurs, telles que Masque × Volume ou Masque × Défaut, sont également observables.

La table ANOVA montre le faible impact du volume sur la détectabilité sauf pour le masque énergétique (E), pour lequel il augmente le pouvoir masquant. L'interaction Masque × Défaut révèle que tous les masques n'interagissent pas de la même manière avec la présence du défaut. En particulier, les valeurs de détectabilité semblent limitées pour le masque attentionnel, obtenant des valeurs proches de détectabilité indifféremment de la condition Défaut. Pour le masque suivant la typologie habillage, celui ci semble ne pas avoir d'effet sur la détectabilité puisque la plage de valeur prise en fonction du défaut est proche de l'amplitude entre la scène idéale (score : 0.231) et la scène non-idéale (score : 0.267).



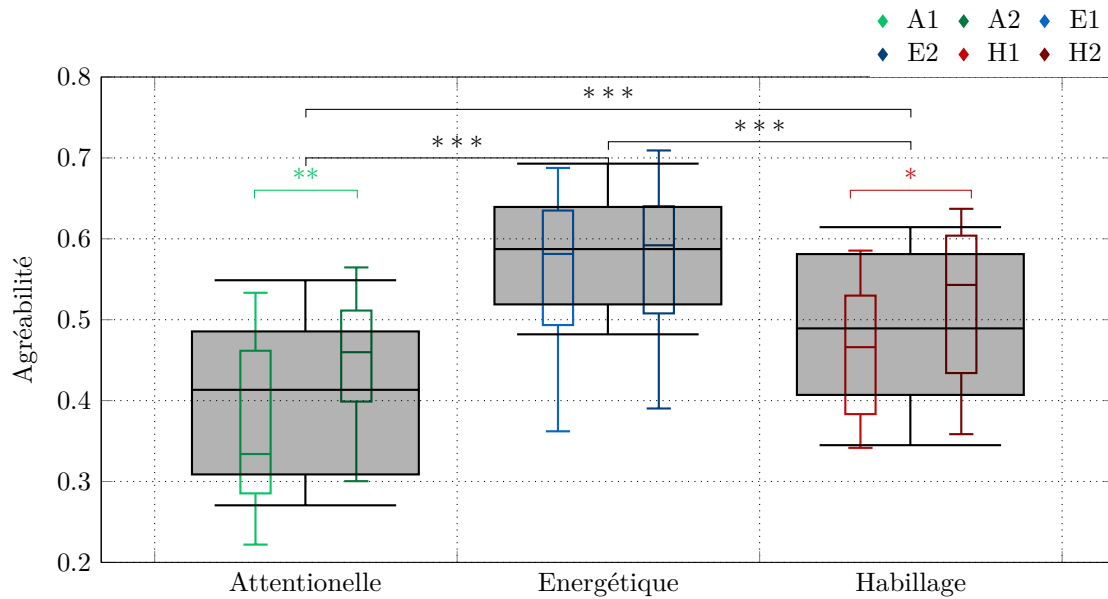
**Figure 4.12.** Boite à moustache pour l'expérience de déteçtabilité, la figure gauche permet de voir l'interaction du facteur Masque et Volume, la figure de droite permet de voir les interactions du facteur Masque avec le facteur Défaut.

#### 4.5.6 | Résultats sur l'expérience d'agrément : pertinence des typologies

De la même manière que pour l'expérience de déteçtabilité, la figure 4.13 représente en boite de Tukey les distributions des scores au sein de chaque catégorie de masque et de typologie pour l'expérience du jugement d'agrément. Contrairement à l'expérience précédente, une différence semble apparaître dans les masques au sein d'une même typologie.

Les calculs statistiques pour mesurer la pertinence d'avoir regroupé les différents masques par typologies de masquage sont à nouveau menés, et confirment que les masques E1 et E2 ne sont pas statistiquement différents [ $t(52) = -0.10$ ,  $p=0.92$ ]. En revanche, pour les autres masques, la proximité est moins vérifiée : les masques A1 et A2 sont perçus comme significativement différents [ $t(52) = 2.9$ ,  $p=5.0 \times 10^{-3}$ ,  $d = -0.79$ ] tout comme les masques H1 et H2 [ $t(52) = -2.08$ ,  $p=4.2 \times 10^{-2}$ ,  $d = -0.55$ ] pour cette expérience. Les différences entre les typologies sont également explorées. Les typologies A et E sont différentes [ $t(106) = 9.72$ ,  $p=10^{-16}$ ,  $d = 1.07$ ], E et H sont différentes [ $t(106) = 5.56$ ,  $p=10^{-7}$ ,  $d = -1.39$ ], et H et A sont différentes [ $t(106) = 4.06$ ,  $p=9.0 \times 10^{-4}$ ,  $d = 0.78$ ].

Contrairement à l'expérience précédente, les résultats statistiques ne permettent pas de confirmer ici l'hypothèse selon laquelle les masques dans la même typologie de masquage produisent un jugement d'agrément similaire. Pour la typologie Énergétique (E), aucun effet significatif est trouvé, donc l'hypothèse est valide. Pour la typologie H, un effet significatif est trouvé avec une faible p-value ( $p=0.04$ ), mais tout de même inférieure au seuil de significativité choisie de 0.05, et une taille d'effet elle aussi assez faible ( $d=0.55$ ), ce qui témoigne d'une certaine proximité dans les résultats malgré la significativité statistique de la différence. En revanche pour la typologie attentionnelle (A), la p-value est plus faible ( $p=5.0 \times 10^{-3}$ ) et la taille d'effet plus importante ( $d=-0.79$ ), ce qui traduit une différence notable entre les deux masques au sein de la typologie. Pour le jugement d'agrément, considérer indistinctement les deux masques de la typologie attentionnelle n'est donc pas possible.



**Figure 4.13.** Boîtes de Tukey pour l'expérience d'agréabilité. Les boîtes de couleurs sont calculés sur les 27 stimuli propres à chaque masque. Les boîtes en gris englobent les résultats pour les deux masques au sein de chaque typologie notée en abscisse : Attentionnelle, Énergétique, Habillage.

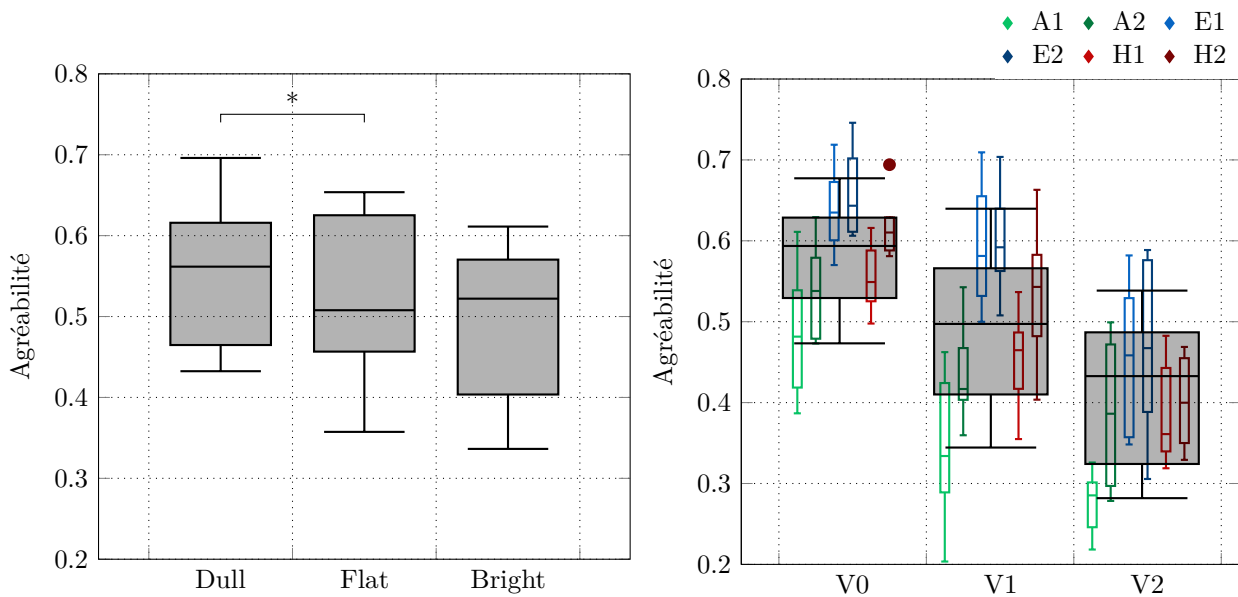
Facteur	Somme des carrés	DDL	P-value	Puissance
Masque	0.96	5	$4.3 \times 10^{-36}$	0.98
Volume	0.77	2	$2.5 \times 10^{-34}$	1.0
Brillance	0.17	2	$6.0 \times 10^{-13}$	0.52
Défaut	0.04	2	$1.8 \times 10^{-4}$	0.17
Masque $\times$ Volume	0.05	10	$2.6 \times 10^{-2}$	0.08
Résidu	0.24	108.0		
<b>Total</b>	<b>2.29</b>	<b>161</b>		

**Table 4.4.** Table ANOVA pour l'expérience d'agréabilité. Seuls les éléments significatifs sont reportés.

#### 4.5.7 | Résultats sur l'expérience d'agréabilité : analyse des facteurs

L'analyse ANOVA faite précédemment montre l'effet principal de TROIS facteurs sur l'agréabilité : le type de Masque (Somme des carrés = 0.96, soit 42 % de la variance expliquée), le Volume (33 %) et la Brillance (7 %) et l'interaction Volume  $\times$  Masque (2.1 %). La figure suivante (Fig. 4.14) représente plus en détail les distributions en fonction de ces facteurs et leurs interactions. En particulier, dans le cas où les résultats sont agrégés selon le facteur Volume, les interactions avec les différents masques sont présentés, ce qui permet de visualiser les différences qui ont été statistiquement calculées entre les différents masques au sein d'une même typologie.

Ce graphe illustre en particulier que le choix d'un masque adéquat peut être préféré à une diminution de niveau. Le masque E2 au volume V2 est ainsi significativement préféré au masque A1 au volume V1 [ $t(16) = 4.8$ ,  $p\text{-value} = 1.9 \times 10^{-4}$ ,  $d = 2.27$ ]. La différence de niveau entre les deux conditions V1 et V2, donnée par la figure 4.3 est d'environ 1.5 phone. La différence forte dans le jugement d'agrément du masque A1, qui possède des évaluations nettement plus basses que les autres masques, est aussi remarquable. Le faible impact du paramètre de Brillance, ainsi que l'absence d'interactions significatives entre la Brillance et d'autres paramètres montrent que le jugement d'agrément d'un masque est faiblement impacté par ce paramètre.



**Figure 4.14.** Boite à moustache pour l'expérience d'agréabilité, la figure gauche regroupe les résultats selon le facteur principal brillance. La figure de droite représente les résultats selon le facteur niveau sonore, en montrant l'interaction avec le type de masque.

#### 4.5.8 | Résumé de la section

##### Résultats

Les résultats se regroupent en 3 parties. La première partie vise à déterminer la qualité de l'expérience, et montre que, bien qu'une différence de durée existe entre les deux expériences, aucune différence significative n'apparaît dans la qualité des résultats selon les calculs de cohérence et de conformité des participants.

La deuxième partie des résultats permet de mettre en évidence la pertinence des typologies de masquage, en particulier dans l'expérience liée au jugement de détectabilité, où les différents masques d'une même typologie obtiennent des résultats similaires, mais les différentes typologies présentent des résultats distincts. Pour l'expérience sur le jugement d'agrément, le résultat est moins net puisque la typologie attentionnelle ne respecte plus cette observation.

Enfin la dernière partie analyse les facteurs expérimentaux et met en évidence le faible impact du facteur Défaut sur le jugement d'agrément. Cette analyse met aussi en évidence qu'un bon choix de typologie peut obtenir un score d'agréabilité similaire à un autre malgré une différence d'une jnd du facteur Volume.

## 4.6 | Discussion

### 4.6.1 | Le choix des sons fictifs, *dummy players*

La section 4.5.1 met en évidence l'échec des deux scènes de références comme *Dummy players*, notamment pour l'expérience d'agrément. Cela peut s'expliquer par la grande ressemblance entre les deux stimuli représentant la scène sonore originale avec et sans défaut, alors que les scènes augmentées disposaient par construction d'une grande variété de caractéristiques possibles.

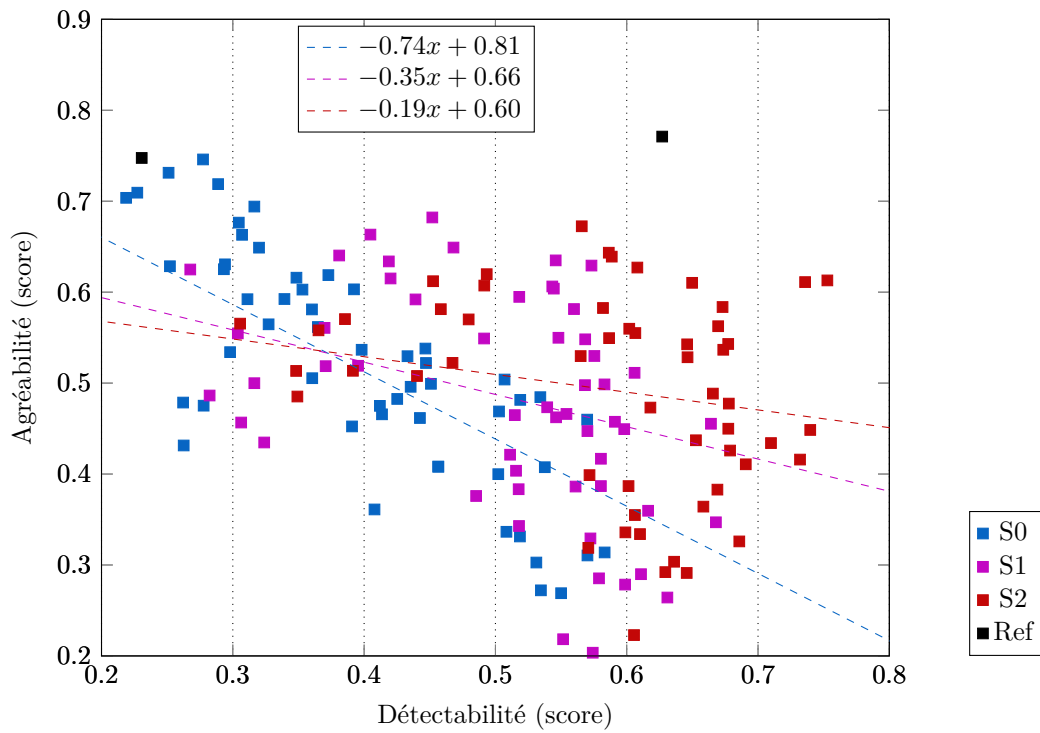
Ce que montre en particulier les résultats de l'analyse des facteurs réalisée, c'est justement que l'impact du facteur Défaut a été sur-estimé dans la conception de l'expérience. Le pourcentage de la variance du jugement d'agrément expliqué par ce paramètre est très faible ( $\approx 1.8\%$ ), comparativement aux autres facteurs tel que le volume du masque ( $\approx 33.6\%$ ), le type de masque ( $\approx 42.1\%$ ) ou même les éléments résiduels ( $\approx 10.5\%$ ). Ainsi le résultat surprenant, qui donne un score d'agrément supérieur à la scène non-idéale (0.771) par rapport à la scène idéale (0.748) s'explique par le fait que le paramètre différenciant ces deux scènes, a un effet très faible par rapport à tous les autres paramètres des masques par rapport auquel ils ont été comparés, et notamment plus faible que des éléments résiduels n'ayant pas été prédits par le modèle.

La construction de ces *Dummy players* à priori est donc difficile à mettre en place dans un jugement d'agrément de ce type. Une proposition pourrait alors être de forcer ces *Dummy players* à être systématiquement le meilleur ou le pire des stimuli dans chacun des 4-uplets où il est présenté en ignorant la réponse du participant. Cette solution permet de s'assurer que leurs rôles soient remplis et d'améliorer la convergence théorique des algorithmes. Cependant elle est peu souhaitable, puisqu'elle induit un déséquilibre dans les données récoltées par la BWS puisque chacun des autres stimuli présents dans ce 4-uplet n'auront alors pas été jugé par un participant.

### 4.6.2 | Lien entre détectabilité et agrément

Le cadre dans lequel se pose cette étude perceptive est principalement donné par l'analyse qualitative de la scène acoustique de l'habitacle, permise par le travail de lexique et du contexte qui en découle. À l'issue de ce travail, l'existence d'un défaut dans une scène a été identifiée comme indiscutablement problématique, notamment par les experts en acoustique automobile. Il est apparu nécessaire de quantifier le lien entre agrément de la scène et présence du défaut. Comme les mêmes stimuli ont été évalués par deux groupes différents suivant la même méthode BWS, les valeurs de scores sont alors obtenues avec exactement le même jeu de données sur un jugement différent. Les valeurs de détectabilité sont présentées en fonction de l'agrément sur un même graphique (fig. 4.15). Chaque point représente donc les valeurs de scores obtenues pour chaque stimulus. Ce que montre ce graphique, c'est que la corrélation entre ces deux jugements est assez faible mais significative [ $t(163)$ ,  $p = 1.58 \times 10^{-9}$ ,  $r^2 = 0.2$ ], de pente  $-0.41$ . C'est-à-dire que la perception d'un son de défaut impacte effectivement le jugement d'agrément d'une scène mais explique peu de variance. Ce résultat en termes de corrélation est cohérent avec le résultat de la table ANOVA évoquée plus tôt.

Notre attention est maintenant portée sur le lien entre ces deux jugements, d'agrément et de détectabilité en fonction de la condition Défaut. La question est de savoir si ce jugement de détectabilité est pertinent dans le cas d'une condition sans défaut (S0), ou s'il explique beaucoup plus de variance dans la condition (S2). Les trois conditions expérimentales (S0, S1, S2) relatives à l'existence du défaut sont repérées en couleur ainsi que les deux scènes de référence indiquées en noir et la corrélation que l'on trouve entre les deux variables évaluées et calculée. Pour les trois conditions, la



**Figure 4.15.** Corrélation obtenues pour les différentes conditions expérimentales liées au sifflement : S0, S1, S2.

corrélations est significative : pour la condition S0 [ $t(163)$ ,  $p = 9.7 \times 10^{-13}$ ,  $r^2 = 0.63$ ], de pente  $-0.97$ , pour la condition S1 [ $t(53)$ ,  $p = 8.3 \times 10^{-4}$ ,  $r^2 = 0.19$ ], de pente  $-0.52$ , pour la condition S2 [ $t(53)$ ,  $p = 4.9 \times 10^{-2}$ ,  $r^2 = 0.07$ ], de pente  $-0.27$ .

Le point mis en évidence par ce graphique est la valeur plus élevée du coefficient linéaire dans la condition idéale (S0), par rapport aux conditions où le défaut existe faiblement (S1) et fortement (S2). L'agrément explicable par la détectabilité d'un sifflement est donc plus forte dans le cas où le sifflement n'est pas présent naturellement dans la scène (S0). Ce constat souligne quantitativement ici l'importance du contexte et de la cohérence des sons dans une scène acoustique. Une interprétation faite de cette différence est en effet que c'est l'origine artificielle du sifflement, dans la condition (S0), qui augmente la valeur du facteur "détectabilité" et est très négativement perçue. Alors que dans la condition (S2), le sifflement est perçu comme naturel, ce qui contribue négativement à l'agrément de la scène, mais pas aussi fortement que s'il était artificiel. Le jugement d'agrément est donc très contextuel.

### 4.6.3 | Résumé de la section

#### Discussion

Cette section permet de discuter des résultats obtenus et des décisions prises lors de la conception de cette expérience. L'hypothèse principale ayant guidé ce test repose sur l'idée que le sifflement est un élément sonore inacceptable, rendant la scène désagréable lorsqu'il est perçu et identifié. Cette hypothèse a conduit à la conception de l'expérience, en choisissant des « *dummy players* », qui sont constitués de la scène sonore sans masque ajouté : l'une avec un sifflement et l'autre sans. Les résultats montrent cependant que le rôle des « *dummy players* » n'est pas rempli. Inclure ces « *dummy players* » lors de la conception d'une expérience de ce type semble alors délicat, cela ajoute deux stimuli a priori différents des autres, pour lesquels il est difficile de remplir le rôle attendu.

Le faible lien entre détectabilité et agréabilité peut expliquer l'échec de ces « *dummy players* », mais remet surtout en question l'hypothèse de l'expérience. En effet, la détectabilité d'un sifflement semble peu impacter le jugement d'agrément de la scène sonore dans laquelle il est présent. L'étude en laboratoire ne confirme donc pas les jugements spontanés formulés lors de la construction du lexique à propos du sifflement. Cependant, une analyse plus poussée des données de l'expérience révèle que la détection d'un sifflement est corrélée à un jugement plus sévère si celui-ci est causé par un masque (c'est-à-dire un son artificiellement ajouté) que s'il est causé par le sifflement réel déjà présent dans la scène. Cette observation souligne alors l'impact significatif du contexte d'apparition d'un sifflement « détecté » sur l'agrément perçu, et semble indiquer que le facteur « artificiel » du son, qui n'a pas été étudié ici peut être un élément clef.

## 4.7 | Perspectives : expérience in-situ

### 4.7.1 | Sélection des masques

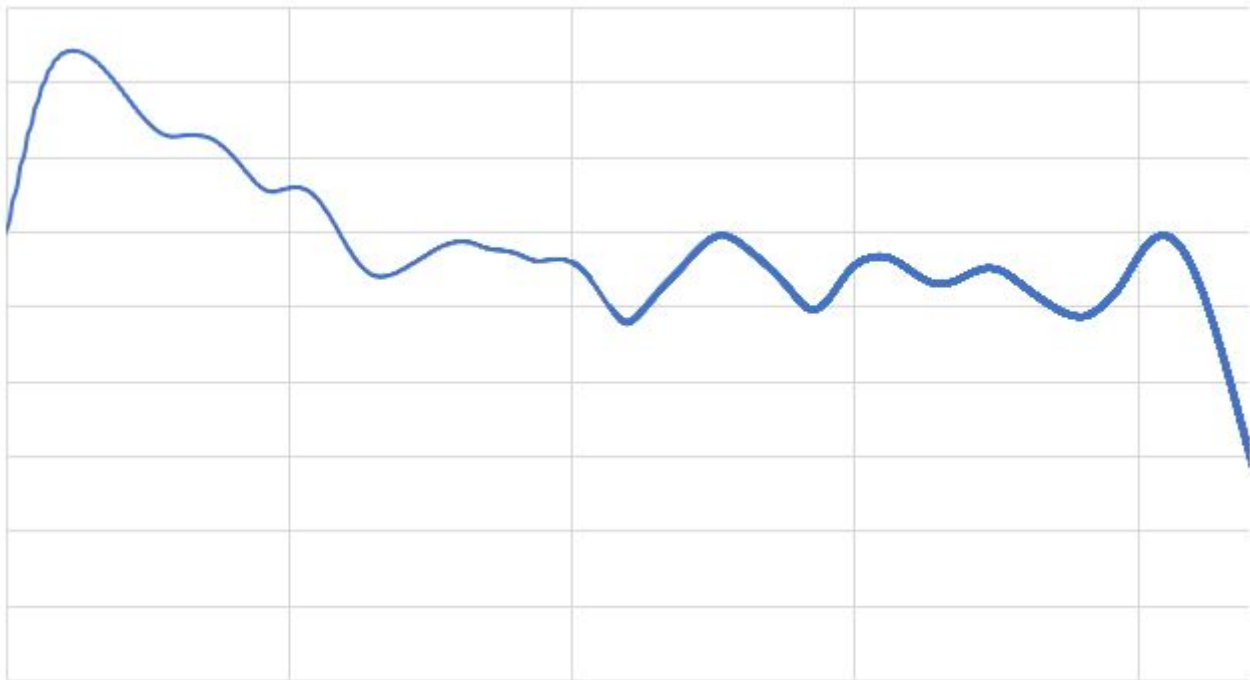
Les résultats de cette étude perceptive menée en atmosphère contrôlée se concluent, entre autres, sur la faible importance du facteur Défaut sur l'agrément, et la faible corrélation entre le jugement d'agrément et le jugement de détection du sifflement. Ces résultats, ne correspondant pas aux échanges avec les experts acousticiens du véhicule, et pourrait être lié au manque de contextualisation dans l'expérience menée en laboratoire. En revanche, ce test valide en grande partie la constitution des trois typologies de masquage. Pour compléter cette expérience et mettre en évidence ce caractère fortement contextuel, une expérience in situ est souhaitable. Ce type d'expérience permettrait de quantifier le facteur contextuel et l'acceptabilité des masques dans un cas réel de conduite et d'utilisation du véhicule.

Du fait d'une exposition requise plus longue à chacun des masques, il n'est pas envisageable de tester l'ensemble des 162 masques créés. Ainsi, seulement deux masques par typologie sont choisis ; le masque ayant le meilleur score sur le jugement d'agrément, et celui ayant le plus faible score sur le jugement de détectabilité. Pour faire cette sélection, les scores ont été moyennés pour les facteurs Défaut et Volume. Les six stimuli gardés pour l'expérience in situ sont donc reportés dans le tableau 4.5.

Comme ces masques ont été composés pour être joués en boucle, il n'est pas difficile d'étendre leurs durées pour se prêter à une expérience plus longue que celle réalisée en laboratoire. On a donc 6 masques à évaluer dans leurs contextes d'utilisation.

Typologie	Masque	Expérience	Score
Energétique	E2B	Detectabilité	0.344
Habillage	H2B	Detectabilité	0.524
Attentionnelle	A2B	Detectabilité	0.564
Energétique	E2D	Agréabilité	0.613
Habillage	H2D	Agréabilité	0.559
Attentionnelle	A2D	Agréabilité	0.521

**Table 4.5.** Classement des meilleurs masques, en moyennant sur les facteurs défauts et volume.



**Figure 4.16.** Fonction de transfert moyennée par tiers d'octaves, entre le fichier audio joué par le système de la voiture et un micro binaural intégré dans un casque, porté en position conducteur. Cette fonction de transfert prend donc en compte à la fois le système audio de la voiture et l'acoustique de l'habitacle. Les graduations des axes ont été modifiées et supprimées par soucis de confidentialité.

#### 4.7.2 | Réalisation technique

Cette expérience consiste à répliquer l'évaluation des masques en condition réelle. Pour cela, nous nous plaçons dans le véhicule dans lequel la situation sonore a été enregistrée, nous y ajoutons artificiellement le sifflement à l'aide d'une enceinte externe, dissimulée pour le participant. Les masques sont ensuite diffusés via le système audio de la voiture. Pour garantir que les masques diffusés dans l'habitacle soient restitués aussi fidèlement que ceux évalués par les participants de l'expérience en laboratoire, la réponse en fréquence du système audio de la voiture et de l'habitacle est mesurée. Cette mesure est réalisée à l'aide d'un casque binaural équipé de micros, porté par un expérimentateur qui est installé sur le siège conducteur. Le système audio de la voiture diffuse ensuite un bruit blanc. La mesure de ce bruit blanc pendant un laps de temps d'une dizaine de secondes peut ensuite être comparée au spectre joué par le système audio à l'aide de l'outil d'analyse LMS Testlab. Cette fonction de transfert, représentée 4.16 nous permet alors de savoir comment les masques joués sont modifiés. Il est donc possible de jouer les masques avec le filtre inverse, pour s'approcher autant que possible des sons tels qu'ils étaient perçus dans les casques des participants lors de l'expérience en laboratoire.

### 4.7.3 | Protocole

Une fois les masques choisis et le véhicule préparé, l'expérience peut commencer. L'étude in-situ se déroule en trois temps détaillés dans les sections suivantes.

#### 4.7.3.1 | Première étape : sélection du masque préféré

Dans la première partie de l'expérience, le participant n'a pas le contrôle sur la ventilation ou le sifflement. En revanche, il a le choix du masque qui est joué parmi les six sélectionnés. La seule question qui lui est posée ici est de choisir le masque qu'il préfère lors de l'utilisation du véhicule avec un défaut de sifflement. L'expérimentateur est sur le siège passager et laisse le participant libre dans sa manière de tester le masque. Il peut conduire sur un parcours défini et écouter chaque masque autant de temps qu'il le souhaite. L'ensemble de l'échange est enregistré, et l'expérimentateur ne pose pas de question à ce moment-là et se limite à changer le masque lorsque le participant le demande. Une fois que le participant a choisi le masque qu'il préfère, on peut passer à la seconde étape.

#### 4.7.3.2 | Deuxième étape : paramétrage du masque préféré

La deuxième étape consiste toujours à juger le masque dans l'utilisation du véhicule comportant le son de ventilation ainsi que le défaut. Cette fois-ci l'utilisateur ne peut plus modifier le masque qui est joué, mais seulement son niveau sonore. Lors de cette étape, l'expérimentateur pose des questions plus précises sur le jugement du masque. Ces questions sont guidées par la définition que l'on a du confort sonore : « *Le confort sonore est ce qui contribue à une sensation agréable dans l'utilisation d'un objet, en terme de retour sonore, cohérent avec les attentes de l'utilisateur et l'environnement sonore déjà existant.* ».

- Jugez vous que l'atmosphère sonore du véhicule est agréable ?
- Jugez vous que l'atmosphère sonore du véhicule est cohérente avec ce que vous attendez d'un véhicule ?
- Jugez vous que le son ajouté est cohérent avec son environnement sonore ?

L'expérimentateur reste à disposition du participant pour préciser les questions si besoin, ou relancer le participant pour participer à la verbalisation.

#### 4.7.3.3 | Troisième étape : évaluation de l'acoustique augmentée

Le participant a finalement la possibilité d'activer le masque ou non. La question est simplement maintenant d'évaluer la pertinence de l'ajout de son en évaluant sur une échelle de Likert à 5 point l'approbation avec l'affirmation suivante : *Je préfère lorsque le masque est allumé.*

#### 4.7.4 | Résumé de la section

##### Perspectives : expérience in-situ

Cette section détaille le protocole pour une expérience in situ permettant d'évaluer les typologies de masque et l'acoustique augmentée dans une situation écologique, en conservant un contexte réel d'utilisation, comportant un défaut : le sifflement. Trois étapes sont alors prévues. La première étape consiste à demander au participant de choisir parmi les six meilleurs masques déterminés par l'expérience en laboratoire (E2B, E2D, H2B, H2D, A2B, A2D) celui qu'il préfère dans la voiture, en condition réelle d'utilisation. Le participant dispose du temps qu'il souhaite pour effectuer son choix. La deuxième étape consiste à permettre au participant de régler le niveau sonore du masque choisi et d'évaluer plus en détail les points de notre définition du confort. La troisième étape consiste à évaluer quantitativement, sur une échelle de Likert à cinq points, la pertinence de la méthodologie de l'acoustique augmentée.

Cette expérience nécessite la présence de l'expérimentateur, pour assister le participant lors des changements de masque et de volume, et pour poser les questions.

## 4.8 | Conclusion

Ce chapitre décrit l'expérience perceptive qui nous a permis d'évaluer les masques composés au précédent chapitre sur deux dimensions : la détectabilité d'un sifflement dans la scène et l'agréabilité de la scène. Ces deux dimensions ont été choisies sur la base du travail de description verbale présenté au chapitre 2 et du travail de contextualisation réalisé lors de la création sonore au chapitre 3. Le choix des termes est primordial lors d'une expérience perceptive, c'est pourquoi nous avons retenu des termes issus des descriptions spontanées des participants, tels que reportés dans les définitions. Ces termes pouvaient alors être aisément évalués par les participants, plutôt que des notions potentiellement plus abstraites tel que le confort sonore d'une scène restituée au casque.

Le choix du type d'expérience perceptive s'est porté sur l'utilisation d'une méthodologie BWS qui permet de traiter un large corpus de sons, en proposant des jugements relatifs plutôt qu'absolus. De plus, cette méthodologie est reconnue comme plus rapide et plus facile pour les participants ce qui a donc justifié son choix dans notre expérience. La méthodologie reste assez récente dans les expériences psychophysiques et on peut noter l'avancée notable permise par l'utilisation d'algorithmes de scoring combinés à cette approche BWS. De nombreuses améliorations restent cependant possibles avec cette méthode, ce chapitre souligne en particulier les possibilités et les limites de cette méthode dans une conception expérimentale factorielle relative aux tests statistiques ANOVA habituellement utilisés. En effet, la méthode n'est pas en capacité de conclure sur la différence significative entre deux stimuli précis, ou deux groupes de stimuli trop petits.

Malgré ces limites, les résultats de cette expérience ont permis de conclure que les participants ont convenablement passé les expériences, avec des scores de cohérence et conformité corrects, et similaires entre les deux expériences, malgré la différence entre les jugements utilisés. Les résultats montrent aussi que les 3 typologies de masquages identifiées sont pertinentes puisqu'elles agissent chacune différemment sur les deux percepts choisis, alors que les sons composés au sein d'une même typologie sont plus similaires entre eux. Les meilleurs résultats sont obtenus de manière significative pour la stratégie Énergique (E) en terme d'agréabilité et de détectabilité. Alors que les résultats pour la détectabilité pouvaient

être attendus, puisque l'on pouvait théoriquement prévoir que le défaut serait physiquement masqué à certains niveaux, les résultats sur l'agréabilité sont plus surprenants. En effet les sons conçus sont assez simples, comme décrit au chapitre précédent (Chap. 3). Pour le masque Attentionnel, le détournement d'attention n'a pas fonctionné comme prévu et n'a pas pu diminuer la détectabilité qui est fait du sifflement. En revanche, cette typologie de masquage augmente la détectabilité, même lorsque la condition expérimentale supprime le sifflement (S0). Ce masque a plutôt tendance à augmenter la confusion autour de l'existence de ce sifflement plutôt que le réduire complètement. Les évaluations d'agréabilité diffèrent significativement entre les deux masques de la même typologie, ce qui traduit une variation interne à cette typologie assez grande. Enfin, le masque Habillage a peu d'impact sur la détectabilité, ce résultat est cohérent avec la stratégie adoptée telle que présentée, puisque le masque Habillage a pour fonction de : « *d'enrichir le son jugé indésirable, non pas pour le rendre imperceptible, mais en modifiant les propriétés acoustiques de celui-ci.* ». De ce fait, le sifflement est resté perceptible pour les participants.

Ce test ne nous permet donc pas de conclure sur l'amélioration du confort par acoustique augmentée directement. Nous sommes effectivement capables de diminuer la détectabilité d'un défaut dans une scène, mais cela n'a pas eu d'effet positif sur l'évaluation de son agréabilité. Le score élevé du jugement d'agrément de la scène de référence, pourtant comportant un défaut, révèle cependant les limites de ce test en laboratoire. En effet la prépondérance du facteur Volume se fait ressentir dans un test peu écologique, où les quelques secondes d'écoute au casque ne permettent probablement pas de restituer la situation d'écoute réelle, dans laquelle le sifflement est pourtant décrit comme inacceptable. Lafay et al. [47] montrent que, dans une évaluation écologique, ce facteur de niveau sonore, appliqué aux sources composant un paysage sonore complet, peut devenir secondaire, ce qui nous laisse penser qu'un test in-situ complèterait en effet les résultats présentés ici. Ce test nous permet cependant de déterminer les meilleurs candidats parmi le large corpus de sons testés, que nous pourrions évaluer dans une étude plus longue et plus écologique, mieux adaptée aux objectifs de ce travail.

## 5. Proposition d'une analyse de la variabilité inter-participants issue des données de la *Best-Worst Scaling*

### 5.1 | Introduction de chapitre

Ce chapitre présente une contribution importante de la thèse sous la forme d'un travail préliminaire, qui aurait pu avoir sa place en annexe mais que nous plaçons dans le corps du manuscrit, comme contribution additionnelle et exploratoire à la phase d'évaluation de ce travail de recherche. Ce travail porte sur une généralisation pour la méthodologie BWS, de tests statistiques qui prennent en compte la variabilité inter-participants. Ces résultats préliminaires, pourront être éventuellement valorisés dans une publication scientifique. Cette réflexion s'articule autour de la volonté d'exploiter les informations spécifiques à chaque individu dans l'analyse de la BWS. Par exemple, des analyses de classification des participants ou des stimuli pourraient être envisagées afin de vérifier, a posteriori, si un paramètre important n'a pas été négligé. À première vue, ce type d'analyse semble inadapté avec les données issues de la BWS, car la seule manière d'obtenir un score pour chaque stimulus et chaque participant repose sur le calcul d'un score « individuel ». Comme simulé au chapitre précédent (fig. 4.2), ce procédé peut au mieux atteindre un coefficient de détermination de  $r^2 = 0.75$  avec le score réel, avec 75% des inégalités respectées. Malgré cette corrélation modérée, nous souhaitons explorer ici la possibilité d'exploiter ces scores individuels sur l'ensemble des participants afin de mener des analyses pertinentes, rendant compte de la variabilité inter-participants.

La problématique guidant ce chapitre peut alors être formulée ainsi : est-il possible d'exploiter les réponses individuelles de chaque participant dans une expérience BWS en vue d'analyses statistiques plus précises, par exemple pour comparer des scores obtenus sur seulement deux stimuli, ou des sous population de participants ? Nous présenterons, dans un premier temps, le formalisme statistique qui justifie théoriquement l'utilisation de la moyenne des scores individuels pour remplir ce rôle. Dans un second temps, les tests statistiques appliqués au chapitre précédent sur les scores obtenus pour des groupes de sons seront répliqués en utilisant l'estimateur de la moyenne des scores. Enfin, des analyses sont présentées, notamment l'analyse en composantes principales.

## 5.2 | Inclure les variations inter-participants

### 5.2.1 | Formalisme statistique

Dans la suite, les définitions utilisées ainsi que les notations seront extraites de l'introduction au calcul des probabilités et à la statistique rédigée par Jean-François Delmas [162]. Nous modéliserons le problème comme cela : nous cherchons à connaître la valeur d'un jugement d'une série de  $J \in \mathbb{N}$  stimuli  $\underline{X} = (X_1, \dots, X_J)$ . Pour cela nous disposons de  $I \in \mathbb{N}$  échantillons, mesures de ces  $J$  stimuli, représentées par une série de variables aléatoires  $\underline{X} = (\underline{X}_1, \dots, \underline{X}_I)$  indépendantes les unes des autres et de même loi  $\mathcal{P}$ . Cette loi inconnue appartient à une famille paramétrée  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ .

Autrement dit, en faisant le parallèle avec la situation à l'origine de ce problème :  $\underline{X}_1, \dots, \underline{X}_I$  c'est  $I$  variables aléatoires de même lois, modélisant les jugements réalisés par  $I \in \mathbb{N}$  participants, de  $J$  stimuli. Les jugements du participant  $i$  sont donc notés :  $\underline{X}_i = (X_{i,1}, \dots, X_{i,J})$ . Chacun des choix de ces participants est modélisé par la même loi  $\mathcal{P}_\theta$ , guidée par un paramètre  $\theta \in \Theta$  intrinsèque à l'expérience. Par exemple,  $\mathcal{P}_\theta$  peut être modélisé par une loi gaussienne d'écart type 1 et de moyenne :  $\underline{\theta} = (\theta_1, \dots, \theta_J) \in \mathbb{R}^J$ . Ici le paramètre  $\underline{\theta}$  représente la valeur réelle de chaque stimuli que la statistique estimera, le participant en propose une valeur perçue, modélisée par la gaussienne, et dont l'incertitude est fixée dans cette exemple par l'écart type, pris égal à 1. Dans le cas de l'expérience menée dans le chapitre précédent,  $J = 164$ , le nombre d'échantillons évalués,  $I = 25$ , le nombre de participants.

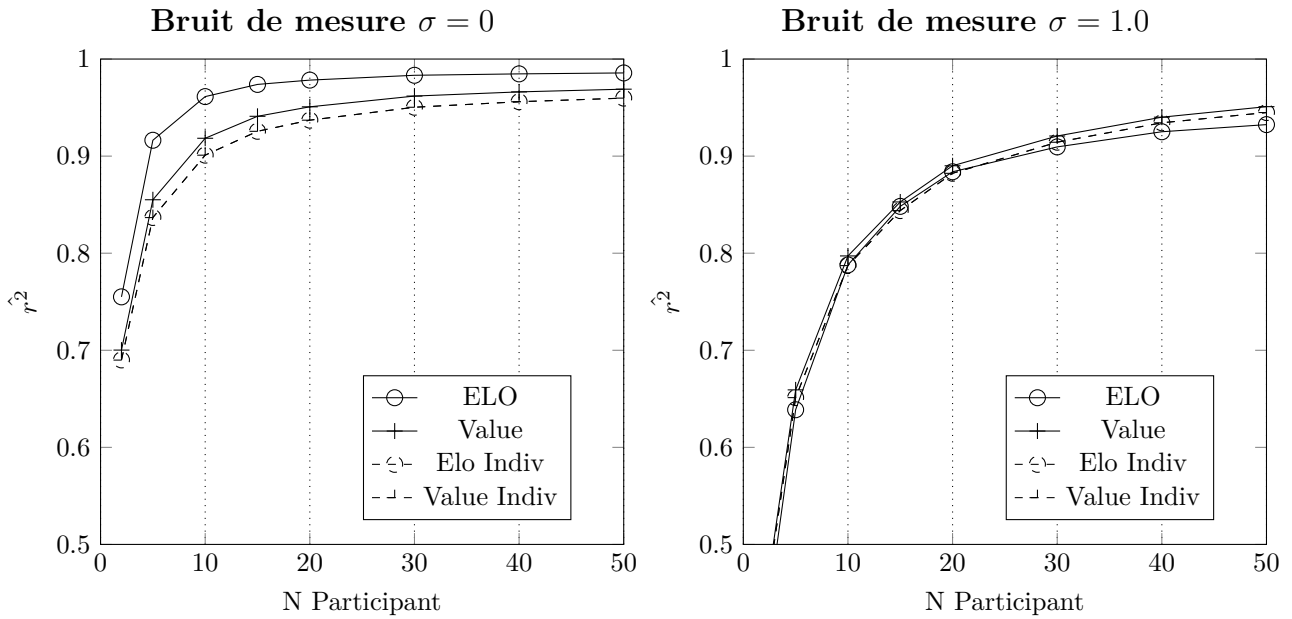
**Definition 5.2.1** (Statistique). Une statistique  $S$  est une fonction de l'échantillon  $\underline{X}$  à valeur réelle ou vectorielle, indépendante de  $\mathcal{P} \in \{\mathcal{P}_\theta, \theta \in \Theta\}$ .

**Definition 5.2.2** (Estimateur). Un estimateur  $\hat{\theta}$  de  $g(\theta)$ , où  $g$  est une fonction quelconque, est une statistique à valeur dans  $g(\Theta)$ .

La fonction  $g$  à estimer que nous utiliserons ici, couramment utilisée, est la fonction  $\mathbb{E}_\theta$ , à valeur dans  $g(\Theta) = \mathbb{R}^J$ , c'est-à-dire, la valeur moyenne de chacun des  $J$  stimuli. Ces valeurs moyennes ne sont pas connues a priori et dépendent de  $\mathcal{P} \in \{\mathcal{P}_\theta, \underline{\theta} \in \Theta\}$ . Dans le cas d'une évaluation directe, chaque stimuli  $j \in [1, J]$  a été écouté par chaque participant  $i \in [1, I]$  et évalué sur une échelle continue allant de 0 à 1, donc  $\underline{X} \in [0, 1]^{I \times J}$ . Un estimateur classique pour estimer l'espérance  $g(\underline{\theta}) = \mathbb{E}_\theta$  est la moyenne (Eq. 5.1). Ce calcul de moyenne est bien fonction de  $\underline{X}$ , il s'agit donc d'une statistique, à valeur dans  $g(\Theta)$ , c'est bien un estimateur. Cet estimateur est connu pour être non-biaisé et convergent par la loi forte des grands nombre, ce qui explique qu'il soit souvent choisi.

$$\hat{\theta} = \frac{1}{I} \sum_{i=1}^I \underline{X}_i \quad (5.1)$$

Dans le cas de la BWS, les jugements de chaque participant se présentent comme la sélection d'un meilleur et d'un pire stimuli, parmi un n-uplet,  $n \in \mathbb{N}$ . On peut l'écrire avec l'indice du meilleur stimuli  $b \in \{1, \dots, n\}$ , l'indice du pire stimuli  $w \in \{1, \dots, n\}$ , et la liste des  $n$  stimuli comparés  $l = (l_1, \dots, l_n) \in \{1, \dots, I\}^n$ , par conséquent  $\underline{X} \in \{1, \dots, n\}^2 \times \{1, \dots, I\}^n$ . Ce formalisme nous permet bien de proposer la fonction score  $s$  comme une statistique, puisque le score est fonction de  $\underline{X}$ , indépendant de  $\mathcal{P}$  et à valeur dans  $g(\Theta) = \mathbb{R}^J$ . Par conséquent, la fonction à laquelle on s'intéresse ici est la moyenne du score de chaque participant (Eq. 5.2) et est bien un estimateur selon la définition.



**Figure 5.1.**  $\hat{r}^2$  est la moyenne, sur 100 simulations, du coefficient de détermination  $r^2$  entre les scores estimés et les valeurs réelles simulées. La figure de gauche est réalisée pour une BWS à 1000 stimuli dont les valeurs sont distribuées par une loi normale centrée réduite, avec des participants simulés sans bruits de mesure. La figure de droite est réalisée avec les mêmes paramètres, mais avec des participants ayant des jugements personnels bruités, variant aléatoirement autour de la valeur réelle des stimuli (variance du bruit : 1.0). Les différentes fonctions de scoring sont calculées : ELO, Value Learning (Value) et leur moyennes sur les scores individuels.

$$\hat{\theta} = \frac{1}{I} \sum_{i=1}^I s(\underline{X}_i) \quad (5.2)$$

Une fois que la fonction d'estimateur est choisie les figures de Hollis [155], simulant l'évaluation de 1000 stimuli, dont la valeur réelle est donnée par une loi normale centrée réduite sont répliquées figure 5.1. Cette valeur est comparée en corrélation, figure de gauche 5.1, avec la valeur estimée des fonctions de scoring *Value Learning* et ELO sur tous les participants, et la moyenne des scores calculés individuellement. Cette valeur de corrélation est obtenue en moyennant sur 100 tirages aléatoires ces stimuli. La figure de droite réplique l'expérience mais en ajoutant un bruit de mesure, chaque participant a ainsi un jugement bruité, simulé par une loi normale centrée réduite. Ces figures permettent de déduire que l'estimateur proposé est convergent, puisque lorsque le nombre de participant augmente, le coefficient de détermination  $r^2$  entre les scores estimés et les valeurs réelles augmente. Ce coefficient de détermination tend vers 1, ce qui montre qu'il est alors possible de calculer une approximation de  $\mathbb{E}_{\theta}$  à un facteur multiplicatif près à partir de la moyenne des scores. La figure illustre aussi la meilleure performance de l'algorithme *Value Learning* sur l'ensemble des participants, par rapport à l'estimateur des scores individuels proposés, puisque le coefficient de détermination est systématiquement meilleur à nombre de participant égal.

$$\hat{\theta} = s(\underline{X}) \quad (5.3)$$

Une remarque pertinente est qu'il est aussi possible de choisir comme estimateur  $\hat{\theta}$  le score de l'ensemble des participants plutôt que la moyenne des scores de chaque participant. Cette fonction du score total,

vérifie aussi les propriétés d'un estimateur et converge à priori plus vite que la moyenne des scores selon la figure 5.1. Pour un nombre de stimuli et de participants fixé, la corrélation avec les valeurs réelles simulées est supérieure. En revanche, notre choix d'estimateur s'écrit comme la moyenne des scores, puisque les tests d'hypothèses statistiques classiquement utilisés, comme le test de Student et ANOVA par exemple, se basent sur un estimateur de moyenne. Plus précisément, ces tests sont construits pour des variables aléatoires suivant une distribution normale, en choisissant un estimateur qui s'écrit comme la somme de variables aléatoires indépendantes mais de même loi (Eq. 5.1 et Eq. 5.2) on s'assure par le théorème central limite que la loi de  $\hat{\theta}$  tend vers une loi normale. Cette hypothèse n'est pas nécessairement vérifiée dans le cas du score total (Eq. 5.3), ce qui rendrait inadapté l'utilisation de ces tests statistiques. Il est donc pertinent d'utiliser des tests type  $\chi^2$ , Student ou ANOVA avec l'estimateur présenté.

### 5.2.2 | Résumé de la section

#### Formalisme statistique

Cette section permet d'identifier que l'estimateur statistique construit en calculant la moyenne des scores individuels selon la formule :

$$\hat{\theta} = \frac{1}{I} \sum_{i=1}^I s(\underline{X}_i)$$

vérifie les propriétés nécessaires à son exploitation dans une analyse statistique. En particulier, il est une fonction de la variable mesurée  $\underline{X}$ , obtenue à partir des tests réalisés avec des participants, et il est indépendant de la loi  $\mathcal{P} \in \mathcal{P}_\theta, \theta \in \Theta$  suivie par les valeurs à estimer. De plus, cet estimateur présente une propriété de convergence vers les valeurs réelles, validée numériquement.

Ainsi, bien que la moyenne des scores des participants converge vers la valeur réelle des stimuli à estimer moins vite que le score total, son caractère individuel offre la possibilité, sur un plan théorique, de mener des analyses inter-participants, au coût d'une estimation du score moins précise à nombre de participant constant.

Facteur	Statistique	DDL	P-Value	d	puissance
E, A Detec	-21.6	2698	0.00	-0.87	1.0
E, H Detec	-16.8	2698	0.00	-0.67	1.0
H, A Detec	-4.3	2698	$1.9 \times 10^{-5}$	-0.17	0.99
E, A Pleasant	18.7	2482	0.00	0.81	1.0
E, H Pleasant	9.3	2482	$3.3 \times 10^{-20}$	0.40	1.0
H, A Pleasant	9.35	2482	$2.1 \times 10^{-20}$	0.41	1.0

**Table 5.1.** tests de Student pour l'expérience de détectabilité et d'agrément

## 5.3 | Application à notre expérience

### 5.3.1 | Tests de Student

#### 5.3.1.1 | Confirmation des résultats précédents

Les tests menés dans le chapitre 4 sont repris dans cette section pour permettre de comparer les résultats obtenus avec cet estimateur par rapport à ceux présentés plus tôt, fait à partir de la moyenne des scores des stimuli.

Pour l'expérience de détectabilité, aucune différence significative n'est trouvée entre les masques d'une même typologie, conformément aux résultats déjà présentés. Pour l'expérience d'agrément : les masques A1 et A2 sont perçus comme significativement différents [ $t(1132) = -4.9$ ,  $p = 1.1 \times 10^{-6}$ ,  $d = -0.30$ , puissance = 1.0] (anciennement : [ $t(52) = 2.9$ ,  $p = 5.0 \times 10^{-3}$ ,  $d = -0.79$ ]) tout comme les masques H1 et H2 [ $t(1132) = -4.2$ ,  $p = 3.1 \times 10^{-5}$ ,  $d = -0.25$ , puissance = 0.99] (anciennement [ $t(52) = -2.08$ ,  $p = 4.2 \times 10^{-2}$ ,  $d = -0.55$ ]). Ce test permet de compléter ce qui a été présenté précédemment, bien que les résultats restent les mêmes, il est maintenant possible d'affirmer avec une grande confiance ( $p = 3.1 \times 10^{-5}$ ) que les masques H1 et H2 ont eu des scores différents, alors que dans le chapitre 4, un doute existait puisque la p-value était proche du seuil de significativité ( $p = 4.2 \times 10^{-2}$ ). On complète ces résultats avec les tests entre les typologies de masquages, reportés dans un tableau 5.2 par soucis de synthèse, où les résultats ne changent pas, mais la confiance que l'on peut leur accorder augmente.

#### 5.3.1.2 | Comparaisons statistique des stimuli

De plus, il est désormais possible de tirer des conclusions sur le score d'un stimulus particulier, dont les valeurs sont reportées dans les figures de l'annexe A. Par exemple, il est possible d'affirmer que la différence d'agrément perçue entre les deux scènes de référence, celle sans sifflement ( $s = 0.748$ ) et celle avec sifflement ( $s = 0.771$ ) n'est pas significative [ $t(44) = -0.56$ ,  $p = 0.58$ ]. L'existence ou l'absence du défaut ne suffit pas à produire une différence de jugement notable pour la scène non augmentée dans cette expérience.

Certains masques présentent des scores de détectabilité surprenants. A2DV2S1 avec un sifflement intermédiaire a ainsi une détectabilité supérieure ( $s = 0.664$ ) à la scène de référence ( $s = 0.627$ ) avec un sifflement plus fort par construction. Il est possible de calculer que cette différence est bien significative [ $t(44) = -0.56$ ,  $p = 0.58$ ,  $d = -1.8$ , puissance = 1.0]. Cela indique que certains masques augmentent fortement la détectabilité perçue d'une scène, au point de la rendre supérieure à une scène présentant un défaut réellement plus élevé.

De la même manière, on trouve que la différence est non significative, qui donne un score de détectabilité plus faible pour E1FV2S1 ( $s = 0.304$ ) que pour E1FV2S0 ( $s = 0.360$ ), alors que par construction le

Facteur	Somme des carrés	DDL	P-Value	Puissance
Masque	1.36	5	0.00	1.0
Volume	$6.4 \times 10^{-3}$	2	0.24	0.13
Brillance	0.03	2	$5.0 \times 10^{-4}$	0.49
Défaut	1.19	2	0.00	1.0
Masque×Vol	0.13	10	$5.3 \times 10^{-9}$	0.52
Masque×Brill	0.07	10	$9.6 \times 10^{-4}$	0.29
Masque×Déf	0.10	10	$1.7 \times 10^{-6}$	0.42
Brill ×Déf	0.12	4	$9.8 \times 10^{-11}$	0.80
Vol× Déf	0.09	4	$8.5 \times 10^{-8}$	0.67
Résidu	8.56	3834		
<b>Facteur</b>	11.68	3887		

**Table 5.2.** Table ANOVA Pour l'expérience de détectabilité. Parmi les interactions, seuls les facteurs significatifs sont reportés dans la table.

Facteur	Somme des carrés	DDL	P-value	Puissance
Masque	1.0	5	$1.1 \times 10^{-89}$	1.0
Volume	0.89	2	$4.8 \times 10^{-78}$	1.0
Brillance	0.17	2	$5.7 \times 10^{-16}$	0.99
Défaut	0.05	2	$6.3 \times 10^{-5}$	0.62
Masque×Volume	0.06	10	$6.7 \times 10^{-3}$	0.26
Résidu	7.91	3348		
<b>Total</b>	10.21	3401		

**Table 5.3.** Reproduction de la table ANOVA pour l'expérience d'agréabilité avec l'ANOVA prenant en compte les DDL des participants. Seuls les éléments significatifs sont reportés.

sifflement est plus fort pour E1FV2S1.

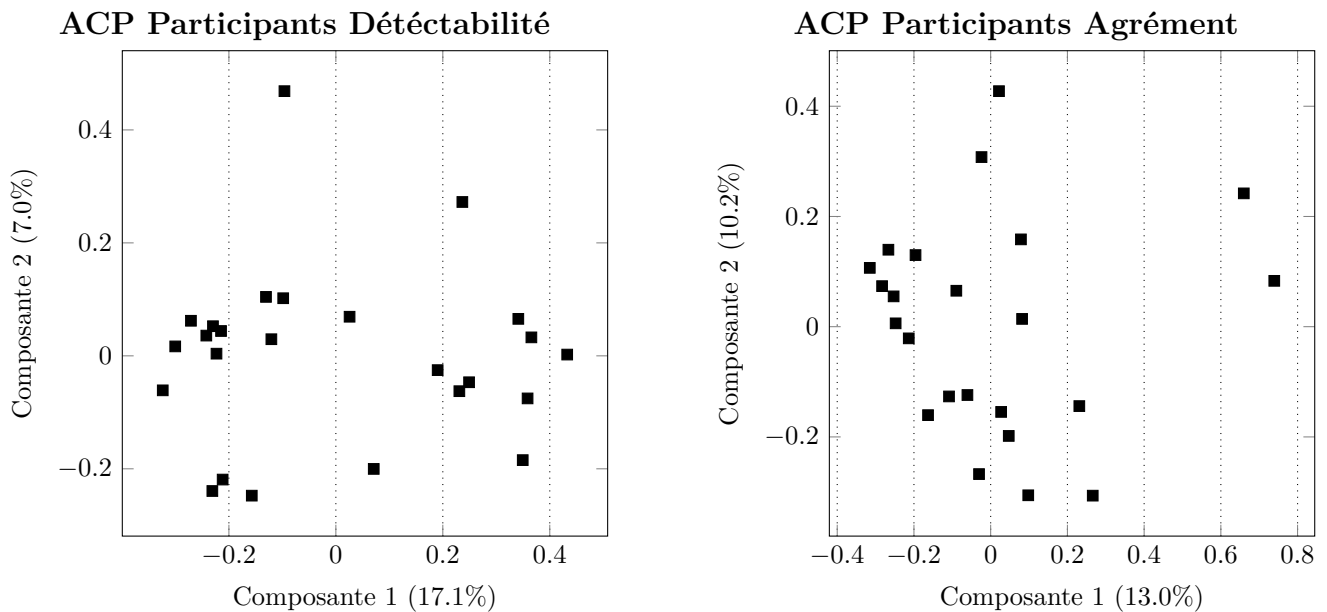
### 5.3.2 | Table ANOVA

Dans cette section, nous reproduisons les tableaux ANOVA (Tableau 5.2 pour l'expérience de détectabilité, tableau 5.3 pour l'expérience d'agréabilité) présentés au chapitre 4, en utilisant l'estimateur construit dans ce chapitre. Ces tests ANOVA possèdent alors davantage de degrés de liberté, ce qui permet de calculer l'impact des facteurs avec une p-value plus faible et une puissance statistique plus élevée. Cela ne permet pas de révéler de nouveaux facteurs ou interactions significatives qui n'auraient pas été détectés auparavant. En revanche, la puissance statistique calculée ainsi que la p-value, c'est-à-dire la confiance que l'on peut avoir dans les conclusions faites, sont nettement améliorées, ce qui renforce la confiance que l'on a dans ces dernières.

### 5.3.3 | Analyse en composantes principales

Une conséquence du choix de la moyenne des scores individuels comme estimateur, est qu'il est désormais possible de construire une matrice des scores de chaque participant et pour chaque son  $\underline{S} = s_{i,j} \ i \in [0, I], \ j \in [0, J]$ . Par construction, les valeurs de scores  $s_{i,j}$  obtenus sont comprises entre 0 et 1, centrées sur 0.5, on travaille alors sur cette matrice sans la modifier pour effectuer une Analyse en Composantes Principales (ACP).

#### 5.3.3.1 | Par participant

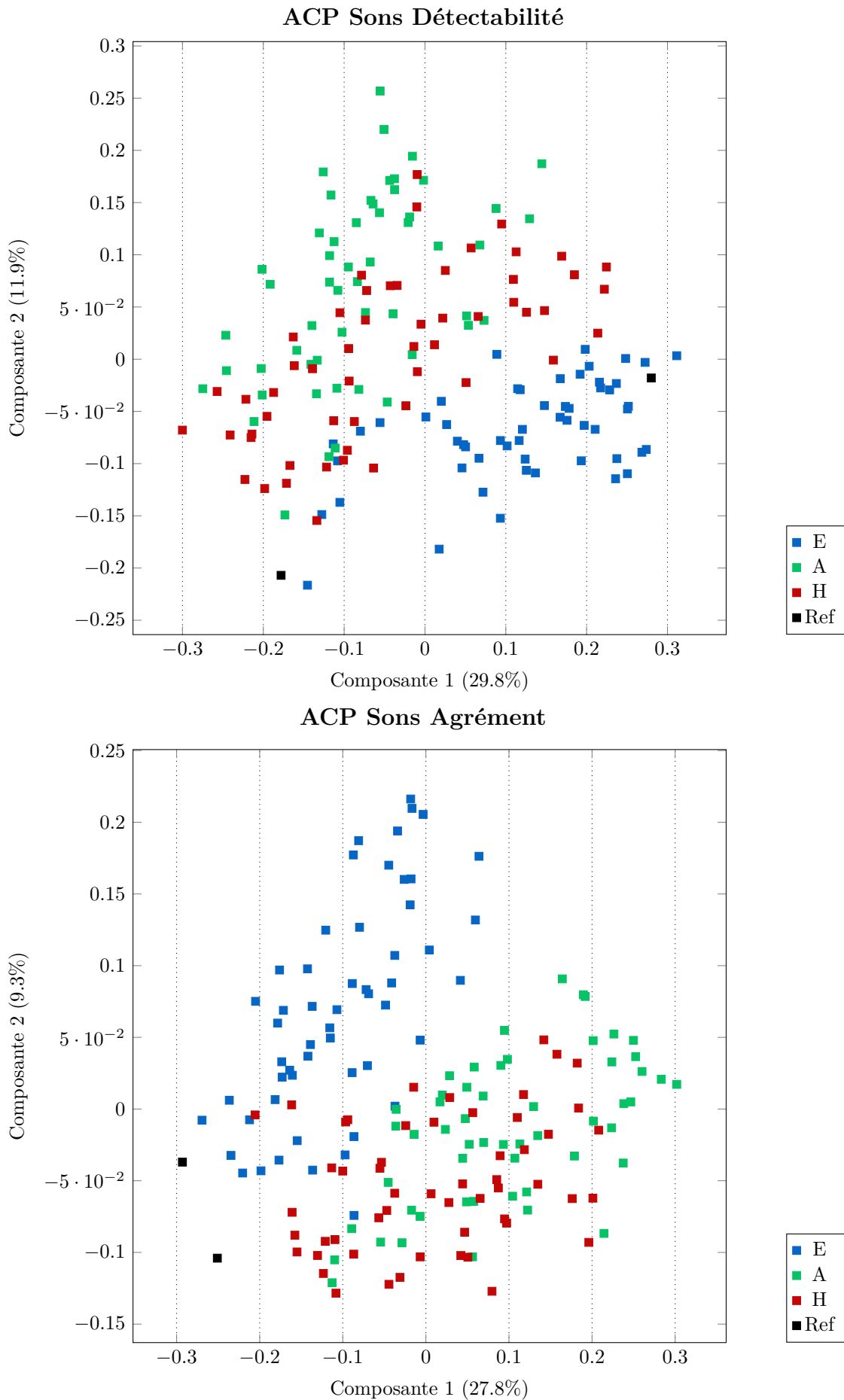


**Figure 5.2.** À gauche, ACP effectuée sur les participants pour l'expérience de jugement de détectabilité, à droite pour le jugement d'agrément. Chaque point représente un participant, les deux points à l'écart sur la figure de droite correspondent aux participants 10 et 15.

La première ACP effectuée sur les participants des expériences, permet de confirmer les résultats mis en évidence par les calculs de conformité et cohérence. À savoir que seuls deux participants ont été exclus de l'expérience de jugement d'agrément en raison de jugements trop différents de ceux des autres. Ceci est également illustré par le graphique 5.2, où deux participants, les participants 10 et 15, apparaissent très éloignés de l'ensemble des autres participants selon la dimension principale de l'ACP.

#### 5.3.3.2 | Par sons

Nous traçons maintenant, figure 5.3, l'ACP effectuée sur les 164 stimuli pour les deux expériences. Les typologies de masquages sont représentées en couleur tandis que les deux scènes de référence en noir. Pour l'expérience de détectabilité, la composante un, qui explique 29.8% de la variance, est fortement corrélée au score total de détectabilité, ce qui conforte l'idée que l'information extraite de la moyenne des scores est proche de l'information issue du score total, puisque le coefficient de détermination  $r^2$  entre cette dimension et le score total est de 0.98. De la même manière, pour l'expérience d'agrément, le coefficient de détermination entre le score total et la composante principale de l'ACP qui explique 27.8% de la variance, est de 0.99. Autrement dit, la première composante de l'ACP effectuée sur les scores individuels corrèle très fortement, dans les deux expériences, avec le score total calculé sur ces mêmes expériences. Aucune corrélation pertinente entre la deuxième dimension et des grandeurs



**Figure 5.3.** En haut, l'Analyse en Composante Principale (ACP) pour l'expérience de détectabilité. En bas, l'ACP pour l'expérience de jugement d'agrément. Les différentes typologies de masquages sont repérées en couleur, ainsi que les deux scènes de références, non augmentées, en noir.

mesurées ou calculées n'est trouvée. Ces ACP nous permettent cependant d'observer un groupement distinct entre les trois typologies de masque dans l'expérience de détectabilité. En revanche, pour l'expérience d'agrément, la visualisation met nettement en évidence la différence de jugement entre les masques énergétiques et l'ensemble des autres masques.

### 5.3.4 | Résumé de la section

#### Application à notre expérience

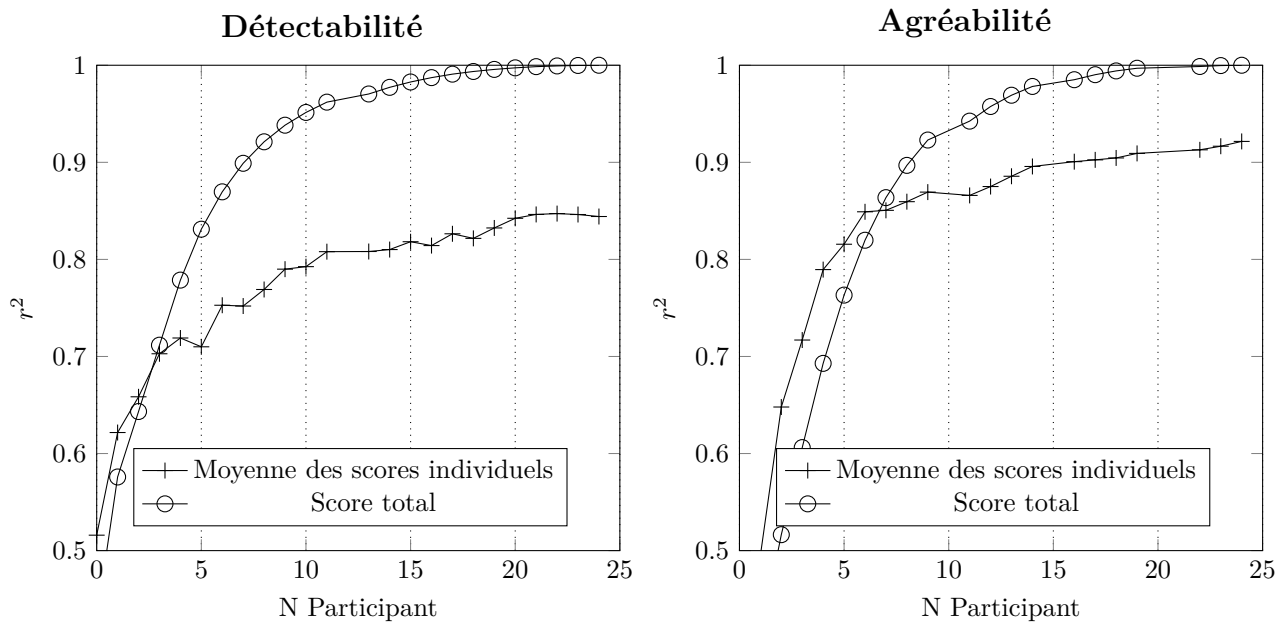
L'estimateur présenté est désormais confronté aux données réelles collectées lors du chapitre précédent (Chap. 4). Les résultats des tests statistiques réalisés à partir des scores individuels de chaque participant pour les stimuli de chaque typologie sont cohérents avec les tests effectués précédemment, qui se limitaient aux scores totaux de ces stimuli. Les différences significatives initialement observées sont également retrouvées avec un niveau de confiance accru, et aucun effet non détecté dans le chapitre précédent n'a émergé. De la même manière, les tableaux ANOVA sont correctement reproduits. Ainsi, cet indicateur permet de parvenir aux mêmes conclusions que celles du chapitre précédent, tout en offrant un degré de confiance statistique plus élevé.

Ces scores individuels permettent également de réaliser des analyses qui n'étaient pas encore proposées. Il devient notamment possible de calculer des différences statistiques pour comparer les stimuli deux à deux ou de réaliser des ACP. Bien qu'aucune valeur de référence ne soit disponible pour confirmer ou infirmer les résultats obtenus, ceux-ci ne présentent aucune incohérence qui aurait pu invalider la méthode.

## 5.4 | Discussion

Au prix d'une convergence moins rapide et de calculs numériques plus longs, une première piste encourageante a été explorée dans ce chapitre sur l'utilisation de l'estimateur construit à partir de la moyenne des scores. Celui-ci permet l'emploi d'outils statistiques classiquement utilisés pour les évaluations directes, tels que les test de Student, les tests ANOVA ou ACP sans limites. Cet estimateur permet aussi de prendre en compte le nombre de participants dans une expérience, pour quantifier la confiance dans les résultats obtenus. Ce type de résultat peut alors mener à une meilleure estimation du nombre de participants nécessaire a priori sur une expérience BWS, avec un calcul de puissance statistique, puisqu'il est désormais possible de calculer un écart-type. Cela permettrait éventuellement aussi de comparer la BWS à d'autres méthodologies, sur la qualité des résultats obtenus en fonction du type de corpus, si par exemple le corpus contient des sons extrêmement variés et différents entre eux, ou si au contraire le corpus contient des sons très similaires.

Le caractère convergent de la moyenne des scores peut alors être mis en évidence sur les données d'une expérience réelle. On trace figure 5.4 l'évolution de la moyenne des scores moyens pour un nombre croissant de participants pour les deux expériences traitées dans cette thèse. On ajoute à ce graphique la convergence observée pour le score calculé pour un nombre croissant de participants, qui converge nécessairement vers 1 ici puisque la corrélation est calculée en comparaison avec les scores pour l'ensemble des participants. Dans le cas de l'expérience de détectabilité, ce graphique nous permet de conclure à posteriori sur le choix d'un nombre suffisant de participants, puisque les deux estimateurs comparés, le score total et la moyenne des scores, semblent avoir convergé à partir du vingtième participant. Ce graphique met cependant en évidence la valeur du coefficient de détermination  $r^2$  entre la moyenne



**Figure 5.4.**  $r^2$  est le coefficient de détermination entre le score final calculé sur l'ensemble des participants, et la moyenne des scores sur un nombre croissant de participant. Cela permet de mettre en évidence la convergence du score avec le nombre de participant en question.

des scores et le score total relativement faible ( $r^2 = 0.844$ ), alors que la courbe semble avoir convergé. Pour l'expérience d'agréabilité, les deux courbes ont une tendance similaire, bien que la moyenne des scores ne semble pas avoir convergé complètement. Dans ce cas, il semblerait qu'il aurait fallu choisir un nombre de participants plus grand, pour permettre à la moyenne des score de converger un peu plus. La contrainte présentée au chapitre précédente du nombre de participant  $K < \frac{N-1}{P-1}$ , permettant de garantir une symétrie des score permet d'aller jusqu'à 54 participants avec le schéma expérimental utilisé.

Des simulations plus complètes seraient cependant nécessaires pour confirmer les premiers résultats encourageants et seront effectués dans la suite de ce travail. La convergence observée sur la figure 5.4 s'éloigne de celle simulée plus tôt figure 5.1. En particulier, une valeur seuil semble être atteinte par l'estimateur construis sur l'expérience de détectabilité, pour une valeur  $r^2 \approx 0.85$ , assez loin des  $r^2 \approx 0.95$  trouvé dans les simulations. Les simulations à venir pourrait alors porter sur l'impact de la distribution des stimuli à évaluer.

## 6. Conclusion et perspective

Ce travail de thèse offre au confort sonore une définition et des outils pour mieux le comprendre. On montre ainsi en premier lieu la proximité méthodologique avec les études menées sur les paysages sonores qui a guidé la première partie de ce travail, notamment par la manière de considérer le contexte autour des sons et pas seulement les sons eux-mêmes. Cette approche écologique met au centre du sujet du confort sonore non seulement le son, mais aussi l'utilisateur. En traitant l'habitacle comme un paysage sonore, on en extrait alors un lexique décrivant les différentes sources sonores présentes sur la base d'échanges semi-dirigés avec des experts de l'automobile maîtrisant cet environnement, encore nouveau pour une population non-experte. Ce lexique est une première brique dans le développement de la culture du son du véhicule électrique, pour les ingénieurs comme pour les utilisateurs, une culture à construire ou à reconstruire, si on la considère comme l'évolution des véhicules thermiques. Il se présente donc comme un outil pour décrire une scène sonore précise selon plusieurs niveaux de discours : causal, réduit et hédonique.

Ce travail de thèse explore aussi la possibilité d'améliorer le confort à travers l'acoustique augmentée : une interaction sonore continue, issue d'un travail de design et qui s'insère dans un environnement déjà existant. Cette amélioration ne se satisfait pas de l'absence d'inconfort seule, mais nécessite des éléments supplémentaires, faisant appel à l'agrément ressenti dans l'utilisation d'un objet, la cohérence avec les attentes de l'utilisateur et l'harmonie des sons avec leur environnement. Cette amélioration du confort est donc fortement contextuelle et nécessite une étude ad-hoc de l'environnement à améliorer. Cette étude nous permet de proposer des éléments susceptibles d'améliorer le confort, dans une situation précise, correspondant à un cas d'utilisation simple : un utilisateur à l'arrêt dans son véhicule, ventilation allumée. Plus précisément, plusieurs typologies de masques sont identifiées en fonction de l'aspect et de la fonction souhaités : une typologie énergétique, une typologie par habillage, et une typologie attentionnelle. Ce travail de création, permet alors de concrétiser cette heuristique et propose un outil de synthèse paramétrique pour les réaliser, ainsi que des masques candidats suivant chacune des typologies identifiées.

Enfin, la dernière partie de cette thèse a permis d'évaluer les stimuli précédemment construits à l'aide d'une méthodologie encore récente dans les études perceptives : une évaluation par *Best-Worst Scaling* (BWS). Cette étude évalue le jugement par des participants sur la fonction remplie par ces sons : « permettent-ils de diminuer la détectabilité perçue d'un défaut ? », et l'intention d'agrément de ces sons : « la scène sonore est-elle agréable ? ». Cette étude montre un effet différent des typologies introduites sur la fonction de ces sons : ils impactent différemment la détectabilité d'un défaut. La différence est cependant moins marquée entre les différentes typologies pour l'agrément. Outre la pertinence des typologies présentées, ce travail révèle la forte dimension contextuelle, difficile à restituer par un test en laboratoire qui ne produit alors pas les résultats attendus, tel que l'impact faible du facteur défaut sur l'agrément pourtant mentionné par les experts, ou alors qui montre des tendances surprenantes comme une sensation de défaut existante, induite par un son ajouté, qui est responsable de plus de désagrément que la perception du défaut réel. Une extension permettant une analyse numérique

plus complète est ensuite explorée. Cette extension permet de remédier aux limites de la méthodologie BWS identifiées, à savoir l'impossibilité d'étudier la variabilité inter-participants.

Cette thèse s'achève en ouvrant de nombreuses perspectives. D'abord, des perspectives scientifiques : les derniers résultats soulignent la difficulté d'évaluer une notion telle que le confort dans un cadre expérimental en laboratoire, et mettent en évidence la nécessité de développer des approches d'évaluation plus contextuelles. Un premier protocole a été présenté dans le chapitre 4, qui pourrait être exploré et affiné. L'objectif est d'évaluer le confort lui-même, ou les différents aspects de sa définition, et de le faire lors de l'utilisation d'un objet dans son contexte réel. On s'attend alors à ce que l'apparition de sons qui agissent positivement ou négativement sur le confort ait un effet plus fort dans ce cas que dans une évaluation en laboratoire. La principale difficulté réside dans la nécessité de minimiser l'interférence avec les habitudes de l'utilisateur lors de l'expérience. Une autre perspective est le développement de la notion d'acoustique augmentée à d'autres environnements sonores, chacun possédant ses propres caractéristiques spécifiques, son contexte, ses attentes.<sup>1,2</sup> Bien que l'approche soit similaire, les résultats n'ont pas de raison d'être à priori directement applicable puisque les typologies de masquages présentées sont inspirées d'exemples déjà existants dans le monde automobile. Une question à laquelle répondre, similaire à celle guidant le chapitre 3, est alors de savoir comment guider la conception de masques pour l'acoustique augmentée dans un environnement avec son propre contexte ? Enfin, une dernière perspective est offerte par l'extension proposée pour les analyses statistiques appliquées à la méthodologie BWS. L'extension de cette méthodologie permet alors de concevoir des schémas expérimentaux paramétriques avec la méthode BWS, décuplant les cas d'application de cette méthode.

Des perspectives créatives émergent également de ce travail. L'acoustique augmentée et les trois typologies ont été explorées à travers un nombre limité de masques, conçus par un seul créateur. Les possibilités de création sont cependant infinies. Les masques proposés ont été élaborés avec un outil spécifique et une vision particulière du véhicule électrique. Un travail de création plus approfondi pourrait être mené, s'appuyant sur le lexique développé dans cette thèse pour faciliter les échanges entre concepteurs sonores et ingénieurs. Contrairement à la matière physique de la carrosserie d'une voiture par exemple, la matière sonore est par nature plus flexible. Une flexibilité démontrée par les technologies de synthèse présentées, disposant de nombreux paramètres pour modifier indépendamment et en temps réel, différents aspects du son. Une perspective intéressante pour l'acoustique augmentée pourrait ainsi proposer des masques avec un fort degré de personnalisation, adaptés aux préférences des utilisateurs ou aux différents usages du véhicule, tout en gardant certaines composantes du timbre commune à chaque véhicule d'une gamme par exemple. Cette approche relève toutefois davantage de la création artistique que de la création scientifique.

---

<sup>1</sup> <https://audio.ls2n.fr/renar-reducing-noise-with-augmented-reality/>

<sup>2</sup> <https://hal.univ-lorraine.fr/AUDIOCARsNET/hal-04760379v1>





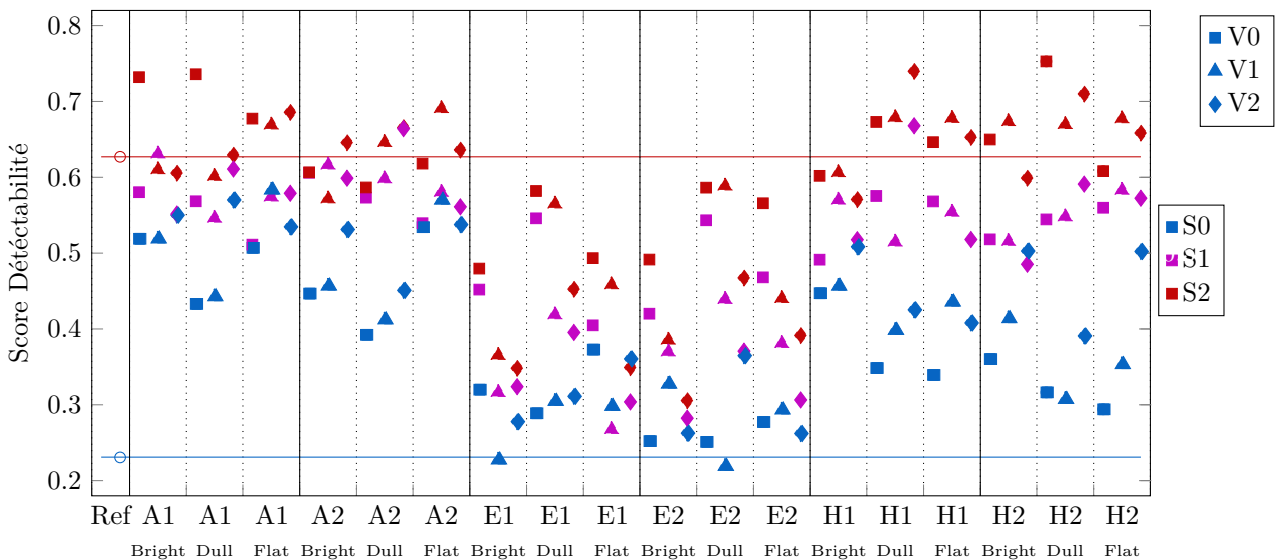
# A. Scores détaillés des participants

## A.1 | Scores détaillés

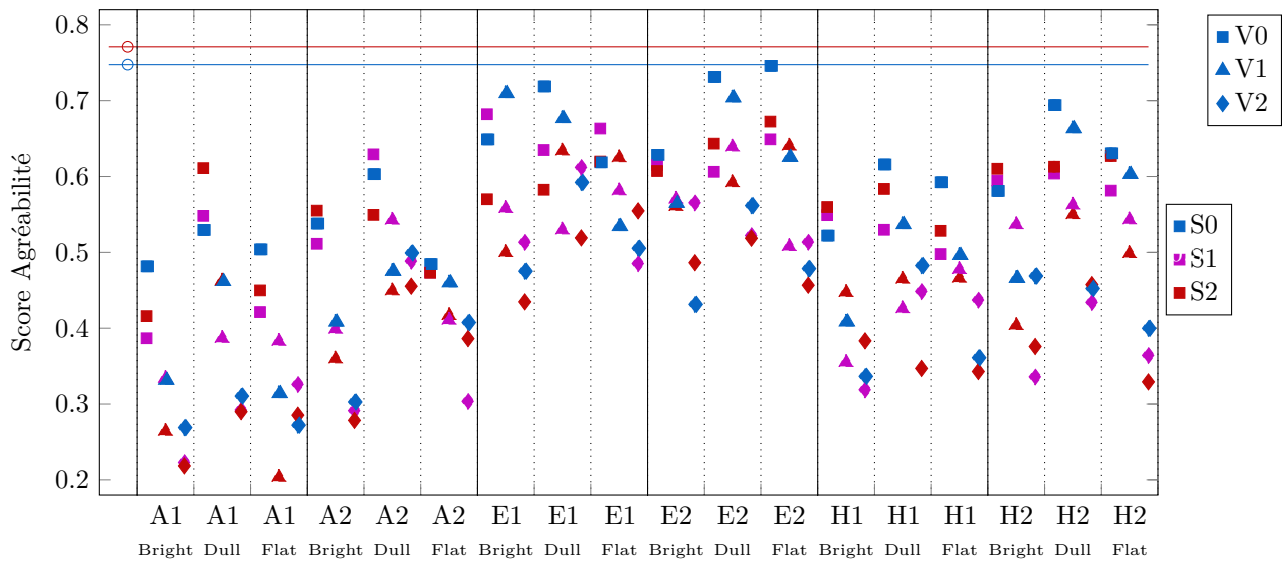
On affiche ici les scores obtenus pour chacun des 164 stimuli et pour chaque expériences. Les différentes conditions expérimentales sont aussi repérées sur la figure. Celle-ci nous permet d’apprécier plus en détail les score relatif de chaque stimulus par rapports aux scènes de références non augmentées. Les scores de ces dernières sont matérialisés par un trait plein, permettant la comparaison rapide des scores obtenus avec tous les autres stimuli.

De plus ce graphique nous permet aussi d’apprécier qualitativement le degré de résolution du score de la BWS sur l’expérience de détectabilité. Cela est possible car le score est presque systématiquement croissant en fonction des différentes conditions expérimentales S0, S1 et S2. Cela traduit qu’une faible différence d’émergence dans le défaut a bien été perçue, et bien relevée dans les valeurs de scores pour presque chaque condition.

Cette précision dans la résolution du score est moins évidente pour l’expérience d’agréabilité, notamment car le score de la scène de référence non idéale, constituée du seul son de ventilation et du sifflement a vraisemblablement obtenu le meilleur score d’agréabilité de tous les stimuli présentés.



**Figure A.1.** Scores obtenus pour l’expérience de détectabilité pour chaque stimulus. Les scores des deux scènes de références sont matérialisés par un trait plein



**Figure A.2.** Scores obtenus pour chaque stimuli dans l'expérience d'agréabilité. Les scores des deux scènes de références sont matérialisés par un trait plein

## B. Traduction du lexique

### B.1 | Traduction du lexique

Le travail de construction du lexique a été mené en français, par conséquent les termes et définitions le sont aussi. La présentation du lexique à des congrès scientifique internationaux ou dans des journaux à comité de lecture, ou simplement le caractère international du groupe Renault, pose la question de la traduction de ce lexique en anglais. Les termes relevés pour la description du son sortent parfois des normes et des nomenclatures techniques, la traduction n'est donc pas directe. De plus ce travail de traduction a aussi pour vocation de garder le côté spontané des termes en évitant des termes trop techniques. Pour mener à bien ce travail de traduction, une étude bibliographique de la littérature scientifique sur le domaine acoustique automobile a été réalisée. Le but est alors ici de présenter des articles de recherches traitant des sons décrit par le lexique. On peut s'assurer qu'il s'agit bien des mêmes sons, justement à travers la description technique offerte dans les articles, qui est confronté à la description technique et causes formulées par les experts en interviews.

#### **Roulement (*Rolling*)**

Le sujet du bruit de roulement a été largement étudié, on peut ainsi aisément trouver des références utilisant le terme *tyre/road noise* ou *road noise* [163, 164]. Ces termes ayant une origine très causale, ils ne correspondent pas parfaitement au type de noms de son souhaité ici. Le terme *rolling noise* apparait aussi fréquemment et y sera alors préféré, qui de plus est lexicalement très proche du terme "roulement" en français.

#### **Sirènement (*Whining*)**

Ce bruit est intrinsèque au véhicule électrique, et plus précisément à la motorisation électrique et a largement été identifié comme un des nouveaux défis acoustiques du véhicule électrique. On peut alors trouver [124, 165, 166] qui se réfèrent à ce son par le terme "*Whining*", certains y préfèrent le terme "*Howling*" [124, 167]. On trouve aussi, mais plus rarement les termes "*Whirring*", "*Whistling*", ou "*harmonic order noise*" [168]. Le terme "*Whining*" plus largement partagé sera plutôt utilisé ici.

#### **Sifflement Aéroacoustique (*Aeroacoustic Whistling*)**

Ce terme ne pose pas de soucis dans sa traduction, le terme "*Whistling*" est utilisé [169, 170] sans mauvaise interprétation ou confusion possible avec le son qu'il est chargé de décrire. Pour éviter toute confusion avec le sifflement électrique déjà présent dans le lexique, la traduction adoptée sera alors "*Aeroacoustic Whistling*".

#### **Sifflement Électrique (*Electric Whistling*)**

La traduction de ce terme est plus délicate, notamment à cause de la confusion possible avec le sirènement dont les causes et les propriétés sont proches. Par conséquent, la simple mention d'un son comme étant "aigu et causé par le moteur" ne permet pas de s'assurer qu'il s'agit d'un sifflement électrique ou d'un sirènement, le détail technique des causes fourni dans les articles permet alors de

s'assurer que l'on parle bien de l'un ou de l'autre. On trouve alors des études qui parlent bien de "*Whistling*" pour ce son électromagnétique causé par le moteur [171], avec la mention de la différence faite dans l'étude entre ce sifflement et le sirènement.

### **Bourdonnement (*Humming*)**

Ce terme est un cas particulier, puisque il est bien connu dans les véhicule thermiques, en particulier pour les véhicules à motorisation diesel. Dans notre cas, le terme apparaît dans le lexique à cause de nombreuse mention qu'il a été faite de ce son dans les interviews et les habitudes d'écoutes des acousticiens. Il existe ainsi de nombreuses références à ce son BF telles que "*Booming*", "*Rumbling*" or "*Humming*" [172, 173] en fonction des causes à l'origine de ce son. Doleschal [174] propose justement une définition et une description détaillée des propriétés acoustiques de chacun de ces 3 termes. Par conséquent on choisira la traduction "*Humming*" pour le bourdonnement qui y est décrit comme "*humming sounds have prominent unmodulated low-frequency components (with frequencies below about 150 Hz). The prototypical sound for humming consisted of a single 50 Hz pure tone.*"

### **Ventilation (*Ventilation*)**

Ce terme n'a pas souffert de difficulté de traduction à cause de son origine causale. On peut ainsi directement en déduite sa traduction par la simple traduction de l'objet à l'origine su son.

### **Chargement (*Load*)**

Les sons d'origine aéroacoustiques sont souvent indistinctement nommés "*Wind noise*" ou "*aeroacoustic noise*" [175]. Dans la construction de notre lexique, 3 sons différents de la famille des sons aéroacoustiques ont été identifiés, le sifflement aéroacoustique, déjà traduit, la fuite et le chargement. La traduction générique de "*Wind noise*" ou "*aeroacoustic noise*" ne nous satisfait donc pas. Le son de chargement, comme son résultant d'une pression aérodynamique sur un panneau est cependant bien mentionné par Coney [176] avec les mêmes origines, en tant que "*loading noise*". Ce terme sera donc gardé dans notre traduction du lexique.

### **Hululement (*Hooting*)**

Ce son a été étudié de nombreuse fois [177, 178] et unanimement nommé "*tire cavity noise*" or "*tire cavity resonance*". De la même manière que pour les traduction proposés pour le roulement, ces termes la sont jugés trop techniques et pas assez intuitifs pour un utilisateur potentiellement non spécialiste du lexique. Ces traductions ne seront pas gardées malgré leurs unanimités dans la littérature scientifique. On choisira ici alors la traduction littérale du terme française "Hululement", "*Hooting*", le terme français étant utilisé pour la proximité entre le son de hululement de la chouette et le son du mode de cavité du pneu, on suppose alors que la traduction du terme évoquera la même chose aux oreilles anglophones.

### **Fuite (*Leak*)**

De manière similaire au chargement, qui est souvent regroupé dans la famille des sons d'origines aérodynamiques qui ne font pas nécessairement la distinction, entre le chargement et la fuite. Des références mentionnant des sons d'origines techniques similaire, le font sous la mention de "*Wind noise*". Pengsea [175] le mentionne plutôt à travers le descripteur "*noise leakage*" or "*acoustic leakage*" qui sera gardé ici.

### **Hachage (*Switching*)**

Ce son est lié à la motorisation électrique, et il faut être prudent lors du travail de traduction de bien s'assurer que le terme ou le sujet d'étude se réfère bien au hachage, et pas au sifflement électrique ou sirènement. Dans les études traitant ce son, il est souvent mentionné par une description causale

"*PWM noise*" [179, 180]. On peut aussi trouver les termes "*Whistling*" or "*Electromagnetic interference*" [181] or "*Switching noise*" [179, 182]. Pour éviter la confusion avec le sifflement électrique ou un terme trop technique, la traduction gardée dans le lexique sera "*Switching noise*".

### **Tambourinement (*Drumming*)**

Ce bruit la était déjà un sujet d'intérêt dans le cas du véhicule thermique, il est donc déjà bien identifié et sa traduction ne pose pas problème puisqu'il est plusieurs fois mentionné avec le terme "*Drumming noise*" [183, 184].



# C. Limites de l'indicateur psychoacoustique de tonalité

## C.1 | Introduction

La situation sonore étudiée dans le chapitre 3 est caractérisée par un défaut acoustique marqué par son émergence. Ce caractère émergent est explicitement mentionné dans la définition du sifflement aéroacoustique établie lors de la création du lexique. Par ailleurs, d'autres sons identifiés au cours de ce travail lexical, représentatifs de l'habitacle d'un véhicule électrique, présentent également ce caractère émergent. Parmi eux figurent le sifflement électrique, le sirènement et le hachage. Ainsi, la question se pose de déterminer quel indicateur peut être utilisé pour quantifier cette perception. Trouver un tel indicateur permettra de quantifier ces valeurs de tonalité, et permettra par exemple de concevoir un schéma expérimental comme celui conçu au chapitre 4, avec des valeurs de tonalités spécifiques. Une revue bibliographique a permis d'identifier trois indicateurs psychoacoustiques couramment employés pour évaluer la perception d'émergence ou, plus précisément, de tonalité : l'indicateur Aures [185] ; le *PsychoAcoustic Tonality* (PAT) [186] ; et le *Prominence Ratio* (PR) [186]. Les calculs des indicateurs seront effectués à l'aide du logiciel ANSYS VRX Sound.<sup>1</sup> Une alternative open source pour les indicateurs *Tone to Noise Ratio* (TNR) et *Psychoacoustic Tonality* (PAT) est disponible via la bibliothèque Mosquito [158], accessible sur GitHub.<sup>2</sup> Certains de ces indicateurs de tonalité ont déjà été comparés dans des études antérieures. Par exemple, l'étude de Parizet et al. [187] a évalué les indicateurs Aures, TNR et *Prominence Ratio* (PR), concluant sur les meilleurs résultats d'Aures. De même, le document ECMA-74 [186] a analysé PAT, TNR et PR, avec une préférence marquée pour PAT. Étant donné la proximité, à la fois théorique et numérique, entre TNR et PR, seule la comparaison de PR avec Aures et PAT sera étudiée dans ce travail.

Les indicateurs sont testés et comparés sur trois types de tests. Le premier s'appuie sur un corpus de sons naturels issus du projet SpeaK de l'IRCAM, conçus par Thomas Rotureau pour illustrer le lexique des descripteurs réduits proposés par Maxime Carron [106]. Les sons en question, décrivant l'attribut "Tonal/Bruité", ont été spécifiquement conçus pour illustrer leur caractère tonal ou bruité, mais ne possèdent pas de valeurs de référence de tonalité perçue. Ils constituent donc un corpus idéal pour tester qualitativement des indicateurs sur des sons aussi naturels que possible, mais ne permettent pas une comparaison quantitative. Le deuxième test consiste en une reproduction d'un test perceptif réalisé dans [186], avec des signaux de type sinus pur auxquels est ajouté un bruit rose. Ces signaux, plus artificiels mais mieux contrôlés, possèdent des valeurs de tonalité perçues reportées dans l'étude, qui serviront de référence et permettent une comparaison quantitative des indicateurs. Enfin, nous testons les indicateurs sur des signaux *Iterated Rippled Noise Sample* (IRNS), présentés et utilisés dans [188] et [189]. Ces signaux sont construits en superposant un bruit blanc à lui-même de manière itérative, en ajoutant un retard et en modifiant l'amplitude du signal. Ce processus résulte en

<sup>1</sup> <https://www.ansys.com/resource-center/brochure/vrxperience-sound-datasheet>

<sup>2</sup> <https://github.com/Eomys/MoSQITo/tree/master/mosquito>

un signal ayant un caractère tonal, pour lequel un indicateur spécifique est proposé et testé dans la littérature [188, 189], ce qui permet une comparaison quantitative sur des signaux sortant de l'ordinaire.

## C.2 | Echantillons SpeaK

### C.2.1 | Présentation

Les échantillons SpeaK illustrant le descripteur de tonalité<sup>3</sup> se présentent sous la forme d'une série de signaux, majoritairement stationnaires, regroupés en cinq catégories (Instruments de musique, Voix femme/homme, sons environnementaux, sons élémentaires, sons complexes). Chaque catégorie contient trois sons de tonalité croissante, appelés "*Bruité*", "*TonalBruité*" et "*Tonal*". Parmi ces catégories, seuls quatre sont conservés pour leur caractère stationnaire : "Voix femme/homme", "Instrument", ainsi que deux échantillons synthétiques, "Son élémentaire" et "Son complexe".

### C.2.2 | Comparaison

Les échantillons sont soumis aux indicateurs précédemment présentés. Ces échantillons n'ont pas de valeurs de tonalités perçue connue, il n'est donc pas possible de comparer les valeurs numériques obtenues à des valeurs de références. En revanche les échantillons ont été composés pour être de tonalité croissante, il est alors possible d'observer si ce caractère croissant est conservé. Pour pouvoir comparer les indicateurs entre eux il a été choisi de tracer l'évolution (Fig. C.1) de l'indicateur entre deux échantillons de tonalité croissante d'une même catégorie de son. Ce choix est motivé par le fait que la plage de valeurs prises par les indicateurs est très différentes, comparer leurs valeur absolues est donc peu pertinent. Les résultats sont alors présentés sous forme d'histogramme figure C.1. Cet histogramme représente le ratio entre la tonalité calculée pour le signal conçu comme le plus tonal, et le moins tonal (Eq. C.1).

$$r_i = \frac{t_i}{b_i} \quad (\text{C.1})$$

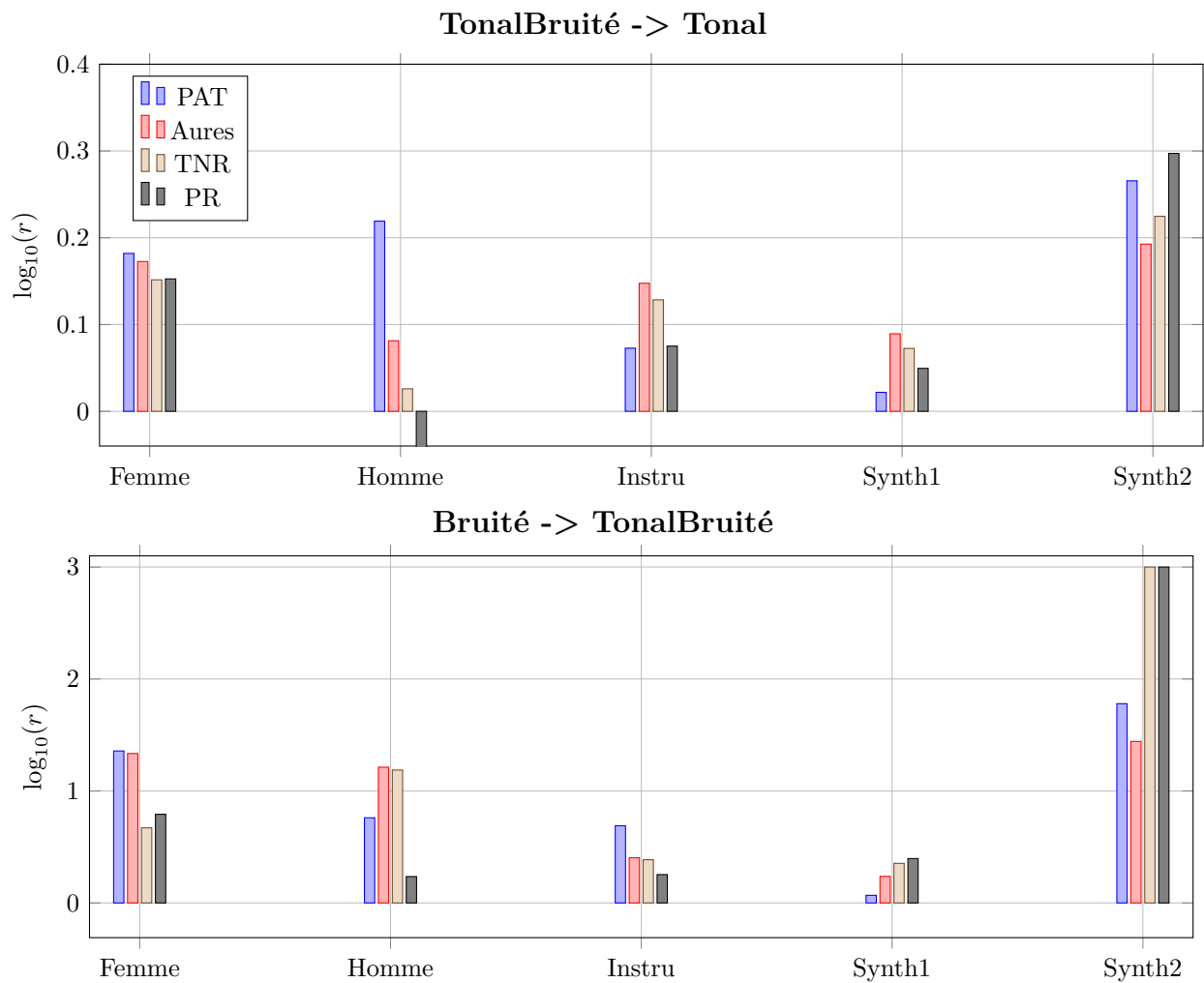
Où :

- $b_i$  est la valeur de l'indicateur du signal le plus bruité
- $t_i$  est la valeur de l'indicateur du signal le plus tonal

Une valeur  $\log_{10}(r) > 0$  indique un comportement de l'indicateur cohérent avec la conception des sons : la valeur augmente lorsque la tonalité augmente. À l'inverse, une valeur négative traduit immédiatement un comportement incohérent. Une valeur de  $\log_{10}(r) > 2$  indique que la valeur de tonalité perçue estimée augmente d'un facteur supérieur à 100 entre deux échantillons. On remarque alors que l'indicateur PR présente une valeur  $r$  négative pour certains échantillons, un résultat incohérent avec le choix des échantillons, conçus pour avoir une tonalité croissante. De la même manière, on relève des valeurs très élevées de  $r > 1000$  pour les indicateurs PR et TNR entre les signaux Tonal et TonalBruité de la catégorie complexe.

Cette première analyse qualitative nous permet de confronter les indicateurs à des signaux variés, spécialement conçus pour mettre en évidence le caractère tonal de certains sons. Ces résultats ne nous

<sup>3</sup> <https://speak.ircam.fr/lexique/lexique-ircam/tonal-buite/>



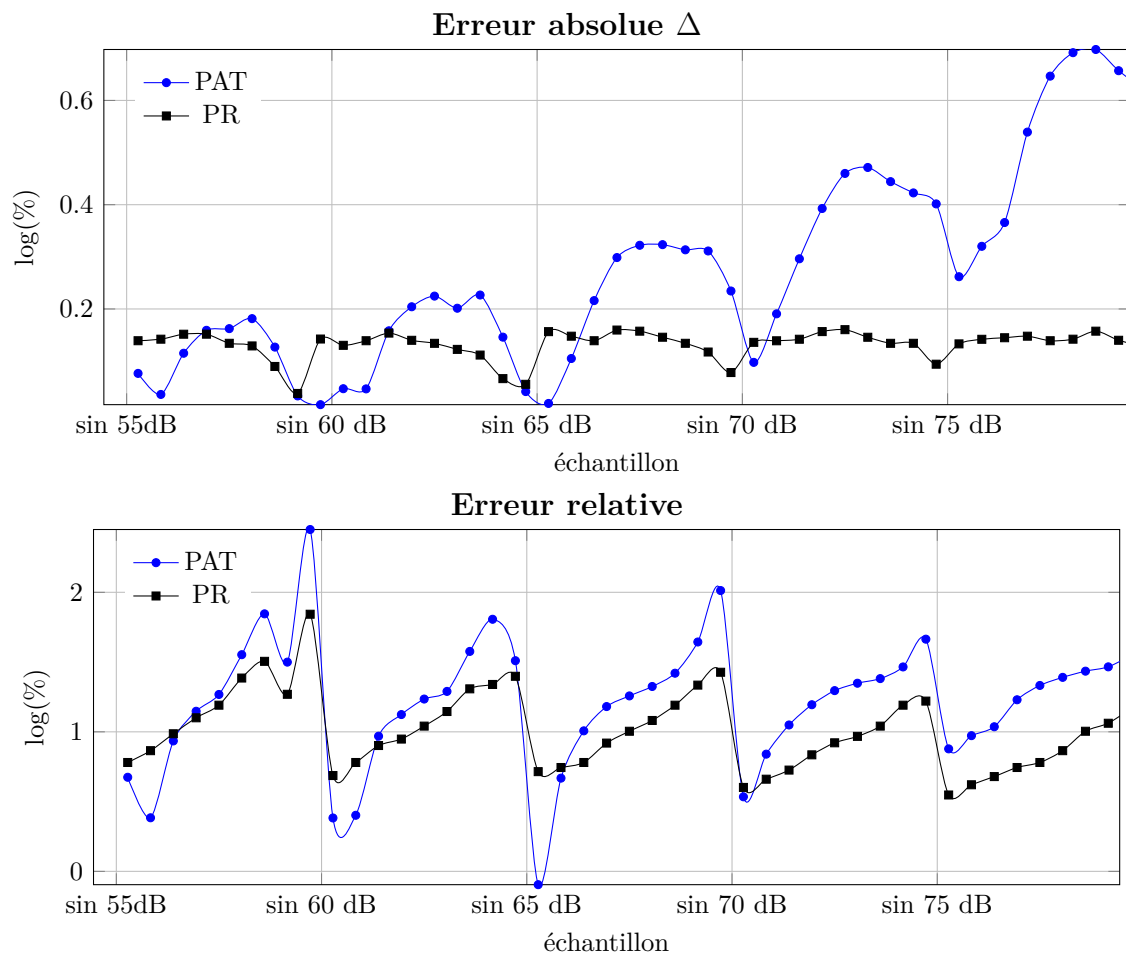
**Figure C.1.** Ratio des indicateurs entre deux signaux de même catégorie issue du SpeaK Words4Sounds de tonalité croissante. Le ratio st représenté en valeur log pour mettre en évidence la différence de sensibilité de chacun des indicateurs.

permettent pas de quantifier la qualité de chacun de ces indicateurs, mais ils mettent déjà en évidence des comportements incohérents pour certains, notamment PR, dont les valeurs semblent incohérentes au vu des stimuli choisis.

### C.3 | Comparaison ECMA-74

#### C.3.1 | Présentation

Dans cette section, les indicateurs sont comparés de manière plus quantitative sur une série de signaux artificiels simples, constitués d'une composante périodique comportant un sinus pur et d'une composante bruitée. Une combinaison de 45 signaux est ainsi construite, combinant la composante pure à 5 niveaux différents (55 dB, 60 dB, 65 dB, 70 dB, 75 dB) et la composante bruitée à neuf niveaux différents (40 dB, 45 dB, 50 dB, 55 dB, 60 dB, 65 dB, 70 dB, 75 dB, 80 dB). Ces signaux spécifiques ont en particulier servi dans la norme ECMA-74 [186], dans laquelle les valeurs des indicateurs psychoacoustiques sont comparées à un test perceptif sur ces signaux précis. Les résultats sont alors reproduits ici, en incluant les indicateurs psychoacoustiques présents dans la norme (PR, TNR, PAT), ainsi que celui de notre choix, absent du document (Aures). On crée ensuite une série de signaux correspondant à ceux présentés dans le document, à savoir un sinus pur à 1000 Hz additionné à un bruit rose. Les 45 déclinaisons des signaux sont alors recomposées en respectant les écarts en dB SPL indiqués.



**Figure C.2.** Évolution de l'erreur absolue et relative entre les valeurs des indicateurs relevé sur l'étude et celles calculés à l'aide du logiciel VR ANSYS

Une fois cette étape effectuée, la première tâche consiste à vérifier que les valeurs numériques données

pour les indicateurs utilisés et les valeurs reportées dans la norme sont proches. Cela permet de s'assurer que les signaux reconstruits sont bien similaires à ceux utilisés dans l'étude de référence. On trace l'erreur entre les indicateurs PR et PAT calculés et les valeurs annoncées dans l'étude, sur les 45 stimuli construits (fig. C.3). Sur le graphique est représenté deux types d'erreur, l'erreur absolue  $\Delta = |TR_{ref} - TR_{calc}|$ , ainsi que l'erreur relative  $\epsilon = \frac{|TR_{ref} - TR_{calc}|}{TR_{ref}}$ . Pour l'indicateur PR, l'erreur absolue est similaire pour les différents niveaux de sinus pur, et l'erreur relative reste également constante. En revanche, pour un niveau de sinus constant, l'erreur relative augmente avec l'augmentation du bruit. Cela s'explique par le fait que l'erreur absolue reste constante, mais que la valeur de tonalité estimée par PR diminue, car le niveau de bruit augmente et la tonalité diminue en conséquence. Le calcul de l'erreur relative donne donc des valeurs plus élevées. Pour PAT, l'erreur absolue augmente avec les différents niveaux de sinus pur, tout comme l'erreur relative lorsque le bruit augmente. Cela suggère que la version implémentée par ANSYS n'est pas exactement la même que celle utilisée dans la norme ECMA-74. Cette figure illustre un réel écart entre le calcul effectué de PAT et les valeurs présentées dans la norme. Cela s'explique par la difficulté technique d'implémenter un tel indicateur, ce qui confirme les difficultés rapportées par [190].

### C.3.2 | Comparaison

L'étude présentée dans ECMA-74 est ici reproduite, avec l'ajout de l'indicateur Aures ainsi que de l'indicateur PAT recalculé. Les valeurs de PR et de TNR ne seront pas tracées, ces indicateurs ayant déjà été mis en défaut plus haut. De plus, les valeurs de PR calculées sont très proches de celles rapportées dans ECMA-74, il n'est donc pas nécessaire de les recalculer ni de les afficher car aucune nouvelle information en est tirée. On conservera uniquement les points correspondant aux valeurs de référence de l'indicateur PAT, qui sera nommé PATref dans la figure C.3.

Pour comparer les indicateurs, ils sont représentés sur un même graphique. Étant donné que les unités des indicateurs diffèrent, un coefficient multiplicatif de conversion  $k$  est appliqué à chacun d'eux. Ce coefficient est déterminé en minimisant la distance *Root Mean Square* (RMS) entre l'indicateur en question et PATref. Ainsi, tous les indicateurs sont représentés dans l'unité associée à PATref :  $tu_{HMS}$ . Par exemple, pour l'indicateur Aures, on obtient un coefficient  $k = 4.63$  ( $r^2 = 0.85$ ) avec les données calculées de PAT, ou  $k = 4.16$  ( $r^2 = 0.90$ ) avec les données relevées de PATref. Cette conversion d'unité n'est pas exacte par construction, car l'indicateur Aures possède une plage de valeurs bornée entre 0 et 1, tandis que PAT n'est pas borné. Il n'existe donc pas d'application linéaire permettant de convertir les résultats d'Aures en ceux de PAT. Ce coefficient multiplicatif de  $k \approx 4$  sera utilisé uniquement pour cette figure, de la même manière que ce qui a été proposé dans ECMA-74.

Pour éviter d'utiliser la conversion basée sur un coefficient  $k$ , une comparaison des indicateurs peut être réalisée en calculant la corrélation entre ces derniers et la tonalité perçue reportée dans le document. La figure C.4 présente le coefficient de détermination  $r^2$  des trois indicateurs comparés : Aures, PAT, et PATref. Pour mettre en évidence les différences entre ces indicateurs, le logarithme de la différence à 1 est tracé. Comme un coefficient de détermination de 1 représente une corrélation parfaite, ce graphique illustre le logarithme de l'erreur pour permettre de mettre finement en évidence les indicateurs minimisant cette erreur. Ce graphique C.4 permet de conclure sur la supériorité de PATref par rapport à Aures, avec un critère quantifiable : un coefficient  $r^2$  systématiquement meilleur pour le test reproduit. En revanche, il est également mis en évidence que l'implémentation de PAT accessible dans ANSYS ne permet pas d'atteindre les résultats annoncés puisque les résultats sont moins bons pour PAT que pour PATref, bien que ces résultats restent meilleurs que Aures.

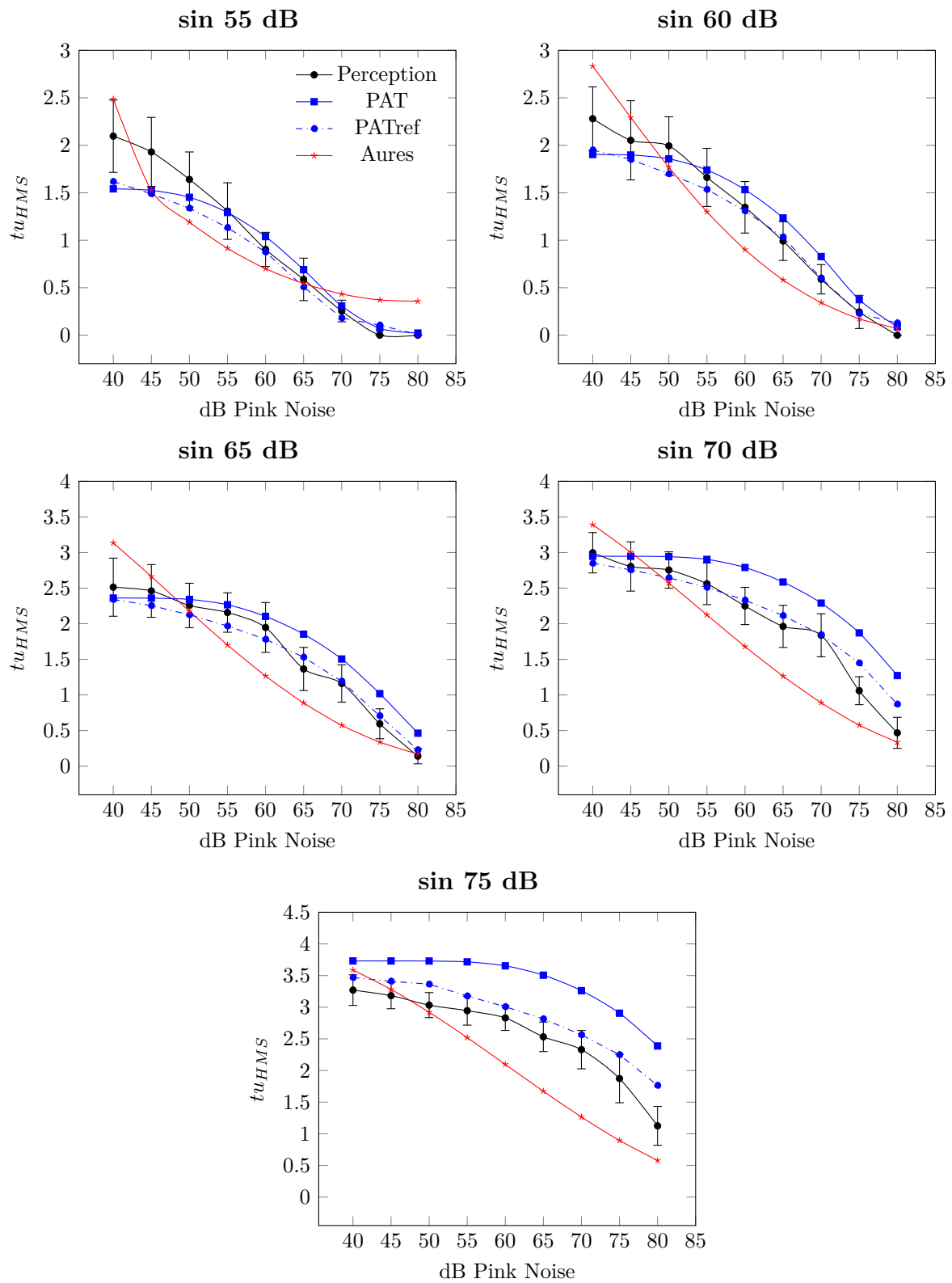
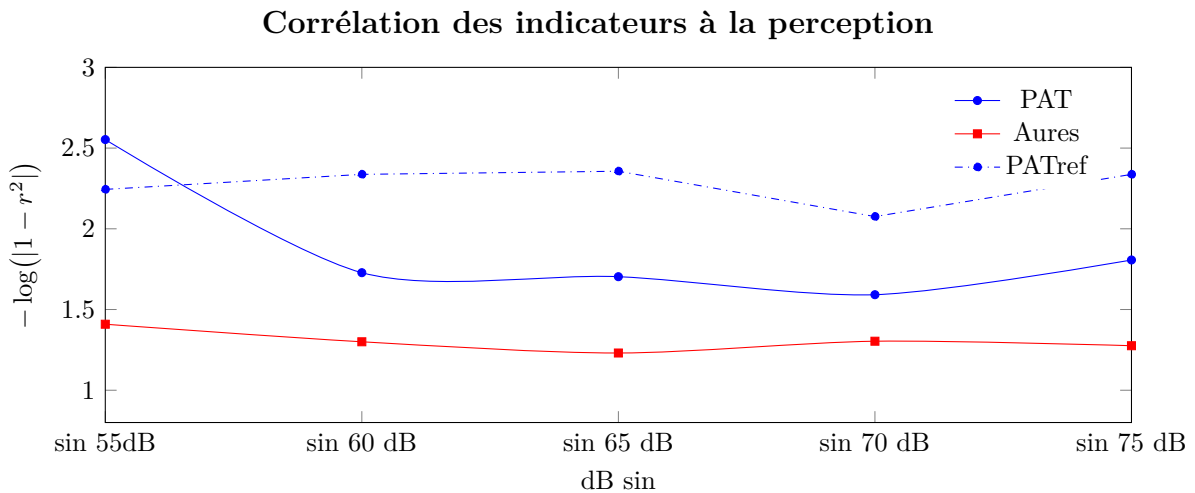
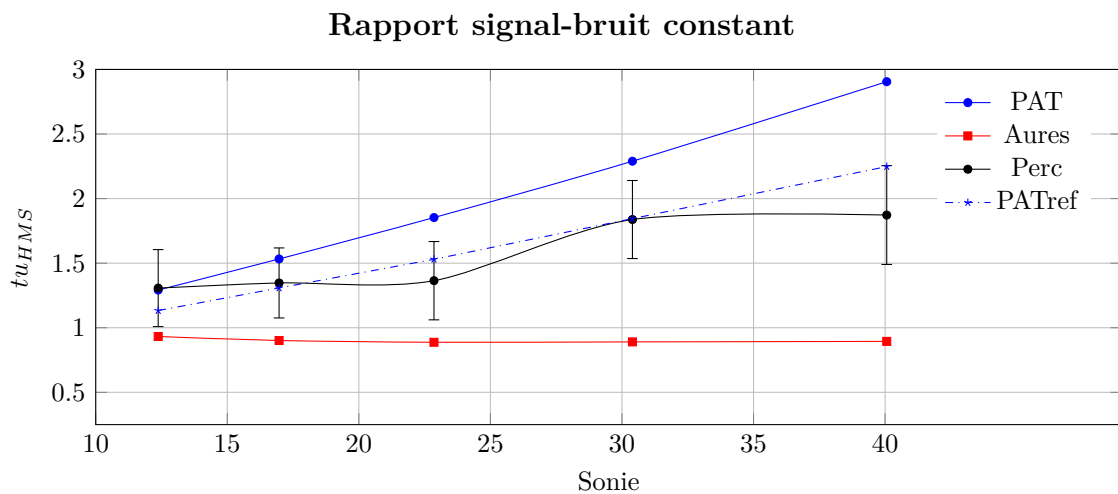


Figure C.3. Reproduction de l'étude perceptive issue de [186] auquel on ajoute l'indicateur 'Aures'

Une différence majeure entre ces deux indicateurs réside dans la sensibilité de PAT au niveau sonore du signal évalué. La figure C.5 présente les résultats précédents, mais pour un rapport signal-bruit constant. Elle est tracée pour cinq signaux artificiels dont les niveaux en dB du sinus pur et du bruit rose sont identiques : 55 dB, 60 dB, 65 dB, 70 dB, et 75 dB. La sonie totale de ces signaux est calculée à l'aide du modèle ANSI S3.4-2007. Ce graphique illustre l'évolution de la tonalité perçue en fonction de la sonie pour un rapport signal-bruit constant, une caractéristique reproduite par PAT et PATref, mais absente chez Aures.



**Figure C.4.** Évolution de la corrélation entre 'Aures' puis 'PAT' avec la perception en fonction des différents niveaux des sinus pur



**Figure C.5.** Évolution des indicateurs pour des rapport signal sur bruit similaire mais une intensité sonore différente

## C.4 | IRNS

### C.4.1 | Présentation

Un troisième test de comparaison est mené pour mettre à l'épreuve les différents indicateurs sélectionnés. Pour ce dernier test, les signaux IRNS seront utilisés [188] ayant la particularité de proposer un processus de création de signaux de tonalité variable à partir d'un bruit blanc, sans rajouter de composantes pur artificielles. L'article propose alors un indicateur de tonalité adapté à ces signaux,  $AC1$  le pic numéro 1 de l'autocorrélation qui comporte l'intégralité de l'information tonal du signal, il présente aussi une fonction de la seule variable  $AC1$  nommé  $\text{FitYost} = 10 + 10^{2AC1}$ , indicateur de la perception de tonalité. L'intérêt de cet indicateur et qu'il a été validé à de multiple reprise sur les signaux IRNS dans la littérature et servira donc de référence pour servir de comparaison aux autres indicateurs.

### C.4.2 | Comparaison

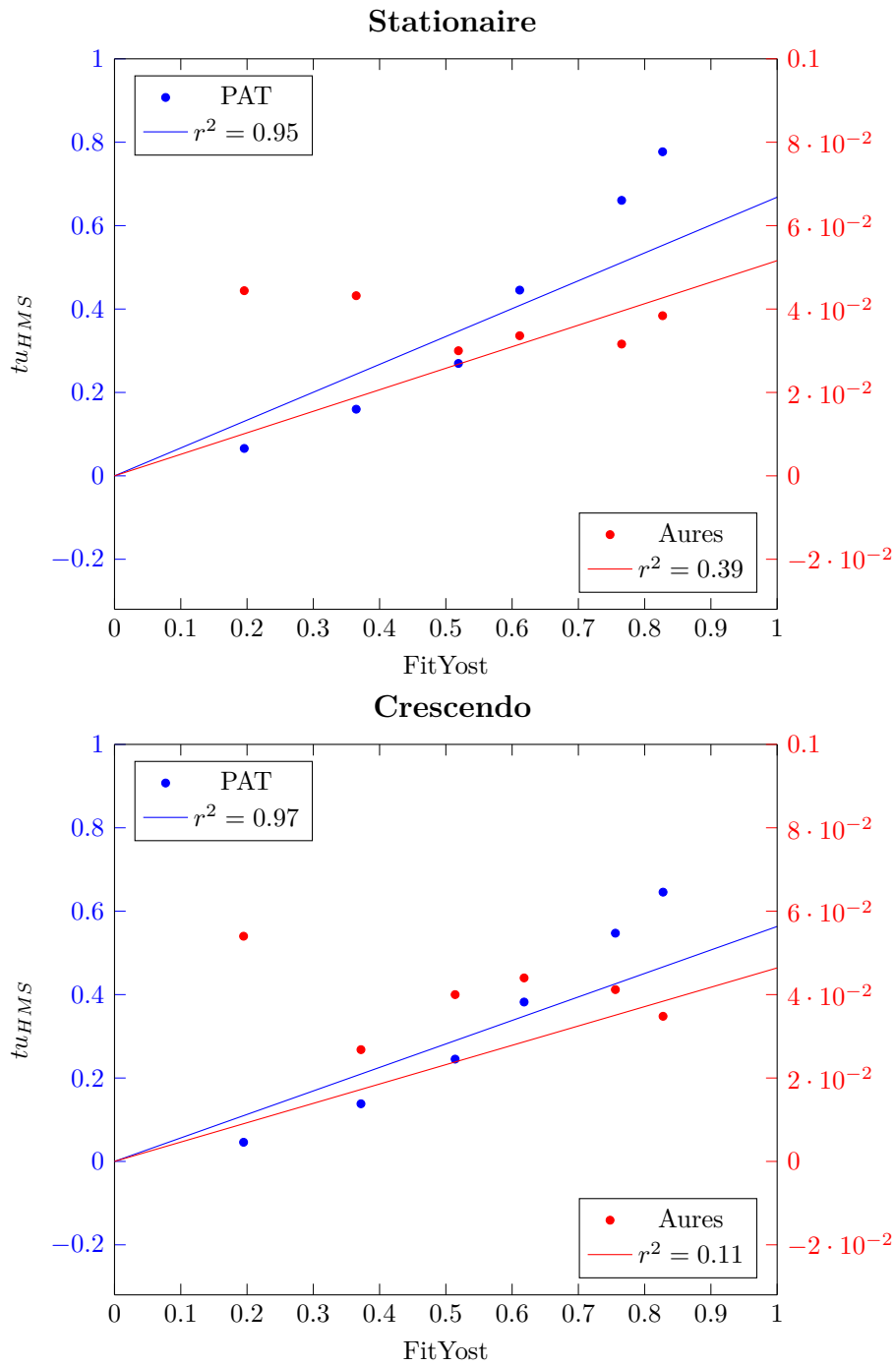
Une série de six signaux stationnaires IRNS de tonalité croissante est générée, pour lesquels les valeurs de tonalité sont calculées et comparées à l'indicateur de référence  $\text{FitYost}$  (figure C.6). Sur la même figure, les valeurs calculées pour six signaux non stationnaires, caractérisés par une augmentation du niveau sonore et une tonalité croissante, sont également représentées.

Un obstacle récurrent ici réside dans la comparaison des différentes unités utilisées par les indicateurs. La comparaison est réalisée en mesurant la corrélation entre chaque indicateur et l'indicateur  $\text{FitYost}$ . Les valeurs de cet indicateur sont considérées comme des références, car elles ont été validées par la littérature scientifique sur ce type précis de signaux. Ces figures mettent en évidence l'incapacité d'Aures à estimer correctement la tonalité des signaux. En effet, Aures attribue des valeurs proches de zéro à tous les signaux, bien que sa plage de valeurs s'étende normalement de 0 pour un signal bruité à 1 pour un signal purement tonal. Ainsi, tous les signaux IRNS sont évalués comme très bruités par Aures. En revanche, PAT montre de meilleures performances, avec un coefficient de détermination très élevé, y compris pour les signaux non stationnaires présentant un niveau sonore croissant. Ces résultats s'expliquent en partie par la méthode commune utilisée par PAT et  $\text{FitYost}$  pour estimer la tonalité : les deux indicateurs se basent sur la fonction d'autocorrélation du signal, bien que les détails de leur implémentation diffèrent.

## C.5 | Discussion

Cette étude annexe sur les indicateurs psychoacoustiques de tonalité vise plusieurs objectifs. Premièrement, elle cherche à mieux comprendre le fonctionnement de ces indicateurs, qui sont de plus en plus utilisés, notamment dans des normes telles que l'ECMA-74, qui a servi de référence ici. Deuxièmement, elle propose une analyse comparative des indicateurs selon trois approches distinctes. La première est une comparaison qualitative, consistant à examiner les valeurs obtenues pour des sons issus du *SpeaK Word4Sounds*,<sup>4</sup> conçus spécifiquement pour illustrer des caractéristiques tonales. La deuxième est une comparaison quantitative, qui reproduit des résultats présentés dans la norme ECMA-74 à partir de signaux artificiels, tels que des combinaisons de sinus purs et de bruit rose. Enfin, une troisième comparaison, également quantitative, utilise des signaux IRNS issus de la littérature scientifique, reconnus pour leur caractère tonal particulier. Cette analyse met en évidence la difficulté de comparer plusieurs indicateurs dont les unités diffèrent.

<sup>4</sup> <https://speak.ircam.fr/lexique/lexique-ircam/tonal-buite/>



**Figure C.6.** Corrélation de 'FitYost'= $10 + 10^{2AC1}$  et 'PAT' puis 'FitYost' et 'Aures' pour différent signaux IRNS

Parmi les trois indicateurs examinés (PR, PAT et Aures), PAT obtient les meilleurs résultats dans les trois tests proposés. Cependant, cet indicateur présente un inconvénient majeur : sa complexité technique, qui entraîne des implémentations variables selon le logiciel utilisé, conduisant à des résultats différents. Cette complexité donne à PAT une grande robustesse face aux différents types de signaux testés ici, présentant des tonalités très variées, mais au prix d'un coût de calcul élevé. Cet indicateur exige des capacités de calcul significatives et se révèle nettement plus lent pour l'analyse de signaux longs. De plus, conformément au modèle de perception [191], cet indicateur est sensible au niveau sonore. Cette sensibilité au niveau sonore, ainsi que la variabilité dans l'implémentation rend l'utilisation de cet indicateur délicate, notamment pour présenter ou comparer des valeurs de tonalités sans avoir accès aux signaux.

# Références

- [1] M. Duroyon, N. Misdariis, P. Susini, N. Dauchez, L.-F. Pardo, and E. Vialatte, “How do automotive acousticians talk about sounds and comfort in electric vehicle?” in *10th Convention of the European Acoustics Association Forum Acusticum 2023*. European Acoustics Association, 2023, pp. 6335–6342. (Cited on pages iii and iv.)
- [2] M. Duroyon, P. Susini, N. Misdariis, N. Dauchez, L.-F. Pardo, and E. Vialatte, “A lexicon to describe specific sounds of the electric car cabin : Verbal approach toward comfort improvement,” *Available at SSRN 4867929*, 2024. (Cited on pages iii and iv.)
- [3] M. Duroyon, P. Susini, N. Misdariis, N. Dauchez, and L.-F. Pardo, “How to deal with unsuspected source of annoyance in electric vehicle cabin using the" sonic augmented" concept ?” in *internoise 2024*, 2024. (Cited on pages iii and iv.)
- [4] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala, “Recent advances in active noise control inside automobile cabins : Toward quieter cars,” *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 61–73, 2016. (Cited on pages 1 and 62.)
- [5] J.-F. Petiot, B. G. Kristensen, and A. M. Maier, “How should an electric vehicle sound? user and expert perception,” in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 55928. American Society of Mechanical Engineers, 2013, p. V005T06A028. (Cited on page 7.)
- [6] L. Humphreys, S. Giudice, P. Jennings, R. Cain, W. Song, and G. Dunne, “The influence of company identity on the perception of vehicle sounds,” *Ergonomics*, vol. 54, no. 4, pp. 338–346, 2011. (Cited on page 7.)
- [7] M. S. Wogalter, R. N. Ornan, R. W. Lim, and M. R. Chipley, “On the risk of quiet vehicles to pedestrians and drivers,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 23. SAGE Publications Sage CA : Los Angeles, CA, 2001, pp. 1685–1688. (Cited on page 7.)
- [8] M. Wessels, C. Zähme, and D. Oberfeld, “Auditory information improves time-to-collision estimation for accelerating vehicles,” *Current Psychology*, vol. 42, no. 27, pp. 23 195–23 205, 2023. (Cited on page 7.)
- [9] N. Misdariis, A. Cera, E. Levallois, and C. Locqueteau, “Do electric cars have to make noise? an emblematic opportunity for designing sounds and soundscapes,” in *Acoustics 2012*, 2012. (Cited on page 7.)
- [10] N. Misdariis and A. Cera, “Knowledge in Sound Design – The Silent Electric Vehicle : a Relevant Case Study,” in *DeSForM - Sense and Sensitivity*, Delft, Netherlands, Oct. 2017. (Cited on page 7.)
- [11] H. Konet, M. Sato, T. Schiller, A. Christensen, T. Tabata, and T. Kanuma, “Development of approaching vehicle sound for pedestrians (vsp) for quiet electric vehicles,” *SAE International Journal of Engines*, vol. 4, no. 1, pp. 1217–1224, 2011. (Cited on page 7.)
- [12] M. Fabra-Rodriguez, R. Peral-Orts, H. Campello-Vicente, and N. Campillo-Davo, “Gear sound model for an approach of a mechanical acoustic vehicle alerting system (mavas) to increase ev’s detectability,” *Applied Acoustics*, vol. 184, p. 108345, 2021. (Cited on page 7.)
- [13] M. Eriksson and R. Bresin, “Improving running mechanics by use of interactive sonification,” *Proceedings of ISON*, pp. 95–98, 2010. (Cited on page 8.)
- [14] T. Takeuchi, “Auditory information in playing tennis,” *Perceptual and motor skills*, vol. 76, no. 3\_suppl, pp. 1323–1328, 1993. (Cited on page 8.)
- [15] N. Schaffert, B. Oldag, and P. Cesari, “Sound matters : The impact of auditory deprivation on movement precision in rowing,” *European journal of sport science*, vol. 20, no. 10, pp. 1299–1306, 2020. (Cited on page 8.)
- [16] L. Bringoux, J. Monnoyer, P. Besson, C. Bourdin, S. Denjean, E. Dousset, C. Goulon, R. Kronland-Martinet, P. Mallet, T. Marqueste *et al.*, “Influence of speed-related auditory feedback on braking in a 3d-driving simulator,” *Transportation research part F : traffic psychology and behaviour*, vol. 44, pp. 76–89, 2017. (Cited on page 8.)
- [17] E. Hellier, A. Naweed, G. Walker, P. Husband, and J. Edworthy, “The influence of auditory feedback on speed choice, violations and comfort in a driving simulation game,” *Transportation research part F : traffic psychology and behaviour*, vol. 14, no. 6, pp. 591–599, 2011. (Cited on page 8.)

- [18] M. S. Horswill and A. M. Plooy, "Auditory feedback influences perceived driving speeds," *Perception*, vol. 37, no. 7, pp. 1037–1043, 2008. (Cited on page 8.)
- [19] T. Melman, P. Visser, X. Mouton, and J. de Winter, "Creating the illusion of sportiness : Evaluating modified throttle mapping and artificial engine sound for electric vehicles," *Journal of Advanced Transportation*, vol. 2021, no. 1, p. 4396401, 2021. (Cited on page 8.)
- [20] G. Goetchius, "Leading the charge—the future of electric vehicle noise control," *Sound & Vibration*, vol. 45, no. 4, pp. 5–8, 2011. (Cited on pages 8 and 29.)
- [21] G. Eisele, P. Genender, K. Wolff, and G. Schürmann, "Electric vehicle sound design—just wishful thinking," in *Proceedings of Aachen acoustics colloquium, Aachen (D)*, 2010, pp. 22–24. (Cited on page 8.)
- [22] M. Munder and C.-C. Carbon, "A literature review [2000–2022] on vehicle acoustics : Investigations on perceptual parameters of interior soundscapes in electrified vehicles," in *Frontiers in Mechanical Engineering*, vol. 8, Aug. 2022, pp. –. <http://doi.org/10.3389/fmech.2022.974464> (Cited on pages 8 and 19.)
- [23] N. Merat and H. Jamson, "A driving simulator study to examine the role of vehicle acoustics on drivers' speed perception," in *Driving assessment conference*, vol. 6, no. 2011. University of Iowa, 2011. (Cited on page 8.)
- [24] O. Le Goff, *L'invention du confort : naissance d'une forme sociale*. Presses universitaires de Lyon, 2021. (Cited on pages 9 and 11.)
- [25] G. Delepaut, "Contribution de la linguistique cognitive à l'identification du confort : analyse des discours des passagers sur le confort en train," Ph.D. dissertation, Université de la Sorbonne nouvelle-Paris III, 2007. (Cited on pages 9, 10, and 12.)
- [26] H. T. E. Hertzberg, "The human buttocks in sitting : pressures, patterns, and palliatives," *SAE Transactions*, pp. 39–47, 1972. (Cited on page 10.)
- [27] L. Zhang, M. G. Helander, and C. G. Drury, "Identifying factors of comfort and discomfort in sitting," *Human factors*, vol. 38, no. 3, pp. 377–389, 1996, publisher : SAGE Publications Sage CA : Los Angeles, CA. (Cited on page 10.)
- [28] M. P. De Looze, L. F. Kuijt-Evers, and J. Van Dieen, "Sitting comfort and discomfort and the relationships with objective measures," *Ergonomics*, vol. 46, no. 10, Aug. 2003. <http://doi.org/10.1080/0014013031000121977> (Cited on pages 10 and 11.)
- [29] P. Vink and S. Hallbeck, "Editorial : Comfort and discomfort studies demonstrate the need for a new model," *Applied Ergonomics*, vol. 43, no. 2, pp. 271–276, Mar. 2012. <http://doi.org/10.1016/j.apergo.2011.06.001> (Cited on page 10.)
- [30] N. Ahmadpour, G. Lindgaard, J.-M. Robert, and B. Pownall, "The thematic structure of passenger comfort experience and its relationship to the context features in the aircraft cabin," *Ergonomics*, vol. 57, no. 6, Jun. 2014. <http://doi.org/10.1080/00140139.2014.899632> (Cited on pages 11, 12, and 88.)
- [31] K. Slater, "Human comfort," 1985. (Cited on page 11.)
- [32] N. Mansfield, A. Naddeo, S. Frohriep, and P. Vink, "Integrating and applying models of comfort," *Applied Ergonomics*, vol. 82, p. 102917, Jan. 2020. <http://doi.org/10.1016/j.apergo.2019.102917> (Cited on page 11.)
- [33] G. Chelkoff, "Du confort acoustique au confort sonore : évolutions des pratiques et de l'architecture du logement," 1993. (Cited on pages 11 and 12.)
- [34] G. Chelkoff, P. Amphoux *et al.*, "Bien être sonore à domicile," *Architectures du logement et potentiel de confort sonore*. Grenoble : CRESSON, 1991. (Cited on page 11.)
- [35] K. Bijsterveld, "Acoustic Cocooning," *The Senses and Society*, vol. 5, Jul. 2010. <http://doi.org/10.2752/174589210X12668381452809> (Cited on page 12.)
- [36] D. Marchand and K. Weiss, "Représentations sociales du confort dans le train : vers une conceptualisation de la notion de confort social," *Les cahiers internationaux de psychologie sociale*, no. 4, pp. 107–124, 2009. (Cited on page 12.)
- [37] M. Mzali, "Perception de l'ambiance sonore et évaluation du confort acoustique dans les trains," Ph.D. dissertation, Paris 6, 2002. (Cited on pages 12 and 20.)
- [38] M. S. Khan, "Evaluation of acoustical comfort in passenger trains," *Acta Acustica united with Acustica*, vol. 88, no. 2, pp. 270–277, 2002. (Cited on page 12.)
- [39] L. Langeveld, R. van Egmond, R. Jansen, and E. Özcan, "Product sound design : Intentional and consequential sounds," *Advances in industrial design engineering*, vol. 47, no. 3, pp. 47–73, 2013. (Cited on pages 13, 15, and 23.)
- [40] T. Hermann, "Taxonomy and definitions for sonification and auditory display," 2008. (Cited on page 13.)

- [41] É. O. Boyer, F. Bevilacqua, P. Susini, and S. Hanneton, “Investigating three types of continuous auditory feedback in visuo-manual tracking,” *Experimental brain research*, vol. 235, pp. 691–701, 2017. (Cited on page 13.)
- [42] J. Tardieu, N. Misdariis, S. Langlois, P. Gaillard, and C. Lemerrier, “Sonification of in-vehicle interface reduces gaze movements under dual-task condition,” *Applied ergonomics*, vol. 50, pp. 41–49, 2015. (Cited on page 13.)
- [43] L. Brocolini, E. Parizet, and P. Chevret, “Effect of masking noise on cognitive performance and annoyance in open plan offices,” *Applied acoustics*, vol. 114, pp. 44–55, 2016. (Cited on pages 14 and 65.)
- [44] C. P. Beaman, “Auditory distraction from low-intensity noise : A review of the consequences for learning and workplace environments,” *Applied Cognitive Psychology : The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 19, no. 8, pp. 1041–1064, 2005. (Cited on page 14.)
- [45] J. Cai, J. Liu, N. Yu, and B. Liu, “Effect of water sound masking on perception of the industrial noise,” *Applied Acoustics*, vol. 150, pp. 307–312, 2019. (Cited on page 14.)
- [46] B. D. Coensel, S. Vanwetswinkel, and D. Botteldooren, “Effects of natural sounds on the perception of road traffic noise,” *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. EL148–EL153, 2011. (Cited on page 14.)
- [47] G. Lafay, M. Rossignol, N. Misdariis, M. Lagrange, and J.-F. Petiot, “Investigating the perception of soundscapes through acoustic scene simulation,” *Behavior Research Methods*, vol. 51, pp. 532–555, 2019. (Cited on pages 14 and 116.)
- [48] K. Bijsterveld and S. Krebs, “Listening to the sounding objects of the past : The case of the car,” *Sonic interaction design*, pp. 3–38, 2013. (Cited on page 14.)
- [49] D. Rocchesso, P. Polotti, and S. Delle Monache, “Designing continuous sonic interaction,” *International Journal of Design*, vol. 3, no. 3, 2009. (Cited on pages 14 and 69.)
- [50] J. A. Russell, “A circumplex model of affect.” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980. (Cited on page 14.)
- [51] Ö. Axelsson, M. E. Nilsson, and B. Berglund, “A principal components model of soundscape perception,” *The Journal of the Acoustical Society of America*, vol. 128, no. 5, pp. 2836–2846, 2010. (Cited on pages 14 and 15.)
- [52] W. F. Thompson, E. G. Schellenberg, and G. Husain, “Arousal, mood, and the mozart effect,” *Psychological science*, vol. 12, no. 3, pp. 248–251, 2001. (Cited on page 14.)
- [53] M. Zentner, D. Grandjean, and K. R. Scherer, “Emotions evoked by the sound of music : characterization, classification, and measurement.” *Emotion*, vol. 8, no. 4, p. 494, 2008. (Cited on page 14.)
- [54] J. Navarro, V. Gaujoux, M. C. Ouimet, L. Ferreri, and E. Reynaud, “How does background music affect drivers’ behaviours, emotions and mood behind the wheel?” *Transportation research part F : traffic psychology and behaviour*, vol. 98, pp. 47–60, 2023. (Cited on pages 14 and 15.)
- [55] S. Cayrou, P. Dickès, and S. Dolbeault, “Version française du profile of mood states (poms-f).” *Journal de thérapie comportementale et cognitive*, 2003. (Cited on page 14.)
- [56] S. Shacham, “A shortened version of the profile of mood states,” *Journal of personality assessment*, vol. 47, no. 3, pp. 305–306, 1983. (Cited on page 14.)
- [57] J. Posner, J. A. Russell, and B. S. Peterson, “The circumplex model of affect : An integrative approach to affective neuroscience, cognitive development, and psychopathology,” *Development and psychopathology*, vol. 17, no. 3, pp. 715–734, 2005. (Cited on page 14.)
- [58] M. D. Van Der Zwaag, C. Dijksterhuis, D. De Waard, B. L. Mulder, J. H. Westerink, and K. A. Brookhuis, “The influence of music on mood and performance while driving,” *Ergonomics*, vol. 55, no. 1, pp. 12–22, 2012. (Cited on page 15.)
- [59] B. Millet, S. Ahn, and J. Chattah, “The impact of music on vehicular performance : A meta-analysis,” *Transportation research part F : traffic psychology and behaviour*, vol. 60, pp. 743–760, 2019. (Cited on page 15.)
- [60] M. L. Turner, J. E. Fernandez, and K. Nelson, “The effect of music amplitude on the reaction to unexpected visual events,” *The Journal of General Psychology*, vol. 123, no. 1, pp. 51–62, 1996. (Cited on page 15.)
- [61] G. Sikander and S. Anwar, “Driver fatigue detection systems : A review,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2018. (Cited on page 15.)
- [62] S. K. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biological psychology*, vol. 55, no. 3, pp. 173–194, 2001. (Cited on page 15.)
- [63] A. Anund, G. Kecklund, B. Peters, and T. Åkerstedt, “Driver sleepiness and individual differences in preferences for countermeasures,” *Journal of sleep research*, vol. 17, no. 1, pp. 16–22, 2008. (Cited on page 15.)

- [64] A. Anund, E. Lahti, C. Fors, and A. Genell, “The effect of low-frequency road noise on driver sleepiness and performance,” *Plos one*, vol. 10, no. 4, p. e0123835, 2015. (Cited on page 15.)
- [65] A. Moradi, S. S. H. Nazari, and K. Rahmani, “Sleepiness and the risk of road traffic accidents : A systematic review and meta-analysis of previous studies,” *Transportation research part F : traffic psychology and behaviour*, vol. 65, pp. 620–629, 2019. (Cited on page 15.)
- [66] S. Bioulac, J.-A. Micoulaud-Franchi, M. Arnaud, P. Sagaspe, N. Moore, F. Salvo, and P. Philip, “Risk of motor vehicle accidents related to sleepiness at the wheel : a systematic review and meta-analysis,” *Sleep*, vol. 40, no. 10, p. zsx134, 2017. (Cited on page 15.)
- [67] E. Smets, B. Garssen, B. d. Bonke, and J. De Haes, “The multidimensional fatigue inventory (mfi) psychometric qualities of an instrument to assess fatigue,” *Journal of psychosomatic research*, vol. 39, no. 3, pp. 315–325, 1995. (Cited on page 15.)
- [68] A. Minard and A. CERA, “Experimenting innovative active sound design to reduce driver weariness in electric vehicles,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 270, no. 4. Institute of Noise Control Engineering, 2024, pp. 7856–7866. (Cited on page 15.)
- [69] M. Bodden, “Principles of active sound design for electric vehicles,” in *INTER-NOISE and NOISE-CON congress and conference proceedings*, vol. 253, no. 1. Institute of Noise Control Engineering, 2016, pp. 7700–7704. (Cited on page 15.)
- [70] S. Moon, S. Park, D. Park, M. Yun, K. Chang, and D. Park, “Active sound design development based on the harmonics of main order from engine sound,” *Journal of the Audio Engineering Society*, vol. 68, no. 7/8, pp. 532–544, 2020. (Cited on page 15.)
- [71] L.-F. Pardo, T. Antoine, A. Minard, and C. Dendievel, “Automotive audio augmented reality,” in *Forum Acusticum*, 2020, pp. 2367–2374. (Cited on page 16.)
- [72] R. G. Boboc, F. Gîrbacia, and E. V. Butilă, “The application of augmented reality in the automotive industry : A systematic literature review,” *Applied Sciences*, vol. 10, no. 12, p. 4259, 2020. (Cited on page 16.)
- [73] P. Susini and O. Houix, “Foreword : Could a functional sound be beautiful?” *Frontiers of Sound in Design A Guide for the Development of Product Identity Through Sounds*, 2018. (Cited on page 16.)
- [74] V. T. Ament, *The Foley grail : The art of performing sound for film, games, and animation*. Routledge, 2014. (Cited on page 16.)
- [75] N. Misdariis, “Sciences du design sonore-approche intégrée du design sonore au sein de la recherche en design.” (Cited on page 16.)
- [76] P. Susini, O. Houix, and N. Misdariis, “Sound design : an applied, experimental framework to study the perception of everyday sounds,” *The New Soundtrack*, vol. 4, no. 2, pp. 103–121, Sep. 2014, publisher : Edinburgh University Press. <http://doi.org/10.3366/sound.2014.0057> (Cited on page 17.)
- [77] G. Lecointre, *Les sciences face aux créationnismes : Ré-expliciter le contrat méthodologique des chercheurs*. Editions Quae, google-Books-ID : vCG9Jo1MbK4C. (Cited on page 17.)
- [78] S. Vial, “Qu’est-ce que la recherche en design ? introduction aux sciences du design,” vol. 1, no. 1, pp. 22–36, place : Paris cedex 14 Publisher : Presses Universitaires de France. <http://doi.org/10.3917/sdd.001.0022>. [Online]. Available : <https://www.cairn.info/revue-sciences-du-design-2015-1-page-22.htm> (Cited on page 17.)
- [79] M. Schaffer, “Le paysage sonore,” *Paris : Ed. Jean-Claude Lattés*, 1979. (Cited on pages 19 and 29.)
- [80] ISO 12913-1 :2014. [Online]. Available : <https://www.iso.org/fr/standard/52161.html> (Cited on page 19.)
- [81] A. Mitchell, F. Aletta, T. Oberman, and J. Kang, “How do we define soundscape?” in *Proceedings of Forum Acusticum*. European Acoustics Association Forum Acusticum, 2023. (Cited on page 19.)
- [82] J. Tardieu, P. Susini, F. Poisson, P. Lazareff, and S. McAdams, “Perceptual study of soundscapes in train stations,” *Applied Acoustics*, vol. 69, no. 12, 2008. <http://doi.org/10.1016/j.apacoust.2007.10.001> (Cited on pages 19, 20, and 31.)
- [83] C. Guastavino, “The Ideal Urban Soundscape : Investigating the Sound Quality of French Cities,” *Acta Acustica united with Acustica*, vol. 92, pp. 945–951, Nov. 2006. (Cited on pages 19 and 20.)
- [84] V. Maffiolo, “De la caractérisation sémantique et acoustique de la qualité sonore de l’environnement urbain : structuration des représentations mentales et influence sur l’appréciation qualitative : application aux ambiances sonores de paris| theses. fr,” Ph.D. dissertation, Le Mans, 1999. (Cited on page 19.)
- [85] F. Guyot, M. Castellengo, and B. Fabre, “Chapitre 2. étude de la catégorisation d’un corpus de bruits domestiques,” in *Catégorisation et cognition : de la perception au discours*. Éditions Kimé, 1997, pp. 41–58. (Cited on page 19.)

- [86] M. Kossachka, M. Mzali, F. Quélin, D. Dubois, and M.-A. Morel, “Une approche psycholinguistique et cognitive du sentiment de «confort» : de l’analyse linguistique à la physique76,” *Dénomination, désignation et catégories*, p. 65, 2006. (Cited on page 20.)
- [87] A. Paté, J.-L. Le Carrou, B. Navarret, D. Dubois, and B. Fabre, “Influence of the Electric Guitar’s Fingerboard Wood on Guitarists’ Perception,” *Acta Acustica united with Acustica*, vol. 101, no. 2, 2015. <http://doi.org/10.3813/AAA.918831> (Cited on pages 20, 23, and 31.)
- [88] G. Delepaut, “Formes linguistiques adjectivales et identification des représentations cognitives,” *Autour des langues et du langage*, p. 193, 2007. (Cited on page 20.)
- [89] J. Tardieu, I. Legriffon, C. Magnen, I. Richard, I. Boulet, and C. Lavandier, “Qualifying and quantifying the different ways in which people are annoyed by aircraft noise,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 270, no. 10. Institute of Noise Control Engineering, 2024, pp. 1285–1293. (Cited on page 20.)
- [90] M. Solomos, “Bruits «entonnés» et sons «convenables» : Russolo et schaeffer ou la domestication des bruits,” *Filigrane. Musique, Sons, Esthétique, Société*, no. 7, 2008. (Cited on page 21.)
- [91] P. Schaeffer, “Traité des objets musicaux [Treatise on musical objects],” *Paris : Editions du Seuil*, 1966. (Cited on page 21.)
- [92] W. W. Gaver, “What in the World Do We Hear? : An Ecological Approach to Auditory Event Perception,” *Ecological Psychology*, vol. 5, no. 1, Mar. 1993. (Cited on pages 22 and 23.)
- [93] M. Chion, *Audio-vision : sound on screen*. Columbia University Press, 1994. (Cited on pages 22 and 23.)
- [94] N. J. VanDerveer, *Ecological acoustics : Human perception of environmental sounds*. Cornell University, 1979. (Cited on page 22.)
- [95] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, “Listener expertise and sound identification influence the categorization of environmental sounds.” *Journal of Experimental Psychology : Applied*, vol. 16, no. 1, p. 16, 2010. (Cited on page 22.)
- [96] M. Carron, “Méthodes et outils pour définir et véhiculer une identité sonore : application au design sonore identitaire de la marque sncf,” Ph.D. dissertation, Université Pierre et Marie Curie-Paris VI, 2016. (Cited on pages 22 and 24.)
- [97] P. Cheminée, C. Gherghinoiu, and C. Besnainou, “Analyses des verbalisations libres sur le son du piano versus analyses acoustiques,” *Colloque interdisciplinaire de musicologie (CIM05), Montreal (Québec), Canada.*, 2005. (Cited on page 23.)
- [98] D. Dubois, “Categories as acts of meaning : The case of categories in Olfaction and audition,” *Cognitive Science Quarterly*, vol. 1, pp. 35–68, Jan. 2000. (Cited on pages 23 and 33.)
- [99] L. N. Solomon, “Semantic approach to the perception of complex sounds,” *The journal of the Acoustical Society of America*, vol. 30, no. 5, pp. 421–425, 1958. (Cited on pages 23 and 88.)
- [100] C. Saitis, C. Fritz, G. P. Scavone, C. Guastavino, and D. Dubois, “Perceptual evaluation of violins : A psycholinguistic analysis of preference verbal descriptions by experienced musicians,” *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2746–2757, 2017. (Cited on page 23.)
- [101] J. Stepánek, “Musical sound timbre : Verbal description and dimensions,” in *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*. Citeseer, 2006, pp. 121–126. (Cited on page 23.)
- [102] A. Faure, “Des sons aux mots, comment parle-t-on du timbre musical?” phdthesis, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Dec. 2000. (Cited on page 23.)
- [103] T. Porcello, “Speaking of Sound : Language and the Professionalization of Sound-Recording Engineers,” *Social Studies of Science*, vol. 34, no. 5, pp. 733–758, 2004. <http://doi.org/10.1177/0306312704047328> (Cited on page 23.)
- [104] V. Rosi, O. Houix, N. Misdariis, and P. Susini, “Investigating the Shared Meaning of Metaphorical Sound AttributesBright, Warm, Round, and Rough,” *Music Perception*, vol. 39, no. 5, pp. 468–483, Jun. 2022, publisher : University of California Press. <http://doi.org/10.1525/mp.2022.39.5.468> (Cited on pages 23 and 31.)
- [105] P. Susini, O. Houix, N. Misdariis, P. Faramaz, and G. Pellerin, “The speak project : A collaborative platform for presenting and sharing sound lexicons,” in *16e Congrès Français d’Acoustique de la SFA*, 2022. (Cited on pages 23 and 47.)
- [106] M. Carron, N. Misdariis, T. Rotureau, F. Dubois, and P. Susini, “Speaking about sounds : a tool for communication on sound features,” *J. of Design Research*, vol. 15, p. 85, 2017. <http://doi.org/10.1504/JDR.2017.10007799> (Cited on pages 23, 25, 36, 47, 79, and 137.)
- [107] G. V. Civille and H. T. Lawless, “The Importance of Language in Describing Perceptions,” *Journal of Sensory Studies*, vol. 1, no. 3-4, 1986. (Cited on page 23.)

- [108] L. J. Lawless and G. V. Civile, “Developing Lexicons : A Review,” *Journal of Sensory Studies*, vol. 28, no. 4, 2013. <http://doi.org/10.1111/joss.12050> (Cited on page 23.)
- [109] P. Vermersch, *L’entretien d’explicitation*. Esf, 2014. (Cited on page 31.)
- [110] S. Delle Monache, N. Misdariis, and E. Özcan, “Semantic models of sound-driven design : Designing with listening in mind,” *Design Studies*, vol. 83, p. 101134, 2022. (Cited on pages 31, 37, and 39.)
- [111] L. Van Audenhove, “Expert interviews and interview techniques for policy analysis,” *Vrije University, Brussel Retrieved May*, vol. 5, p. 2009, 2007. (Cited on page 31.)
- [112] A. Pearce, T. Brookes, and R. Mason, “Timbral attributes for sound effect library searching,” in *Audio Engineering Society Conference : 2017 AES International Conference on Semantic Audio*. Audio Engineering Society, 2017. (Cited on pages 36 and 39.)
- [113] P. Sotiriadou, J. Brouwers, and T.-A. Le, “Choosing a qualitative data analysis tool : A comparison of nvivo and leximancer,” *Annals of leisure research*, vol. 17, no. 2, pp. 218–234, 2014. (Cited on page 37.)
- [114] A. Smith, “Automatic extraction of semantic networks from text using leximancer,” in *Companion Volume of the Proceedings of HLT-NAACL 2003-Demonstrations*, 2003, pp. 23–24. (Cited on page 37.)
- [115] G. A. Miller, “Wordnet : a lexical database for english,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995. (Cited on page 39.)
- [116] J. Le Besnerais, V. Lanfranchi, M. Hecquet, and P. Brochet, “Bruit audible d’origine magnétique dans les machines asynchrones,” *Techniques de l’ingénieur. Génie électrique*, vol. 6, no. D3580, 2008. (Cited on page 48.)
- [117] H. Rittel, “Some principles for the design of an educational system for design,” publisher : Taylor & Francis. [Online]. Available : <https://www.tandfonline.com/doi/abs/10.1080/10464883.1971.11102482> (Cited on page 59.)
- [118] N. Zafeiropoulos, J. Zollner, and V. Kandade Rajan, “Active road noise cancellation for the improvement of sound quality in the vehicle,” *ATZ worldwide*, vol. 120, no. 3, pp. 38–43, 2018. (Cited on page 62.)
- [119] E. Zwicker and H. Fastl, *Psychoacoustics : Facts and models*. Springer Science and Business Media, 2013, vol. 22. (Cited on pages 65 and 66.)
- [120] B. Moore, “Mechanisms of masking,” *The Journal of the Acoustical Society of America*, vol. 57, no. 2, pp. 391–399, 1975. (Cited on pages 65, 66, and 95.)
- [121] D. S. Brungart, “Informational and energetic masking effects in the perception of two simultaneous talkers,” vol. 109, no. 3, pp. 1101–1109, publisher : Acoustical Society of America. <http://doi.org/10.1121/1.1345696>. [Online]. Available : <https://asa.scitation.org/doi/abs/10.1121/1.1345696> (Cited on page 65.)
- [122] J. F. Culling and M. A. Stone, *Energetic Masking and Masking Release*. Springer International Publishing, 2017. (Cited on page 65.)
- [123] S. Brandl, B. Graf, and A. Rust, “NVH Challenges and Solutions for Vehicles with Low CO2 Emission,” *SAE International Journal of Passenger Cars - Mechanical Systems*, vol. 5, no. 3, Jun. 2012. <http://doi.org/10.4271/2012-01-1532> (Cited on pages 65 and 82.)
- [124] M. Münder and C.-C. Carbon, “Howl, whirr, and whistle The perception of electric powertrain noise and its importance for perceived quality in electrified vehicles,” *Applied Acoustics*, vol. 185, Jan. 2022. <http://doi.org/10.1016/j.apacoust.2021.108412> (Cited on pages 65 and 133.)
- [125] I. Pollack, “Auditory informational masking,” vol. 57, p. S5. <http://doi.org/10.1121/1.1995329>. [Online]. Available : <https://doi.org/10.1121/1.1995329> (Cited on page 65.)
- [126] V. Fraisse, C. Tarlao, and C. Guastavino, “Shaping city soundscapes : In-situ comparison of four sound installations in an urban public space,” *Landscape and Urban Planning*, vol. 251, p. 105173, 2024. (Cited on page 65.)
- [127] T. Dupré, S. Denjean, M. Aramaki, and R. Kronland-Martinet, “Analysis by synthesis of engine sounds for the design of dynamic auditory feedback of electric vehicles,” *Acta Acustica*, vol. 7, 2023. <http://doi.org/10.1051/aacus/2023031> (Cited on page 67.)
- [128] B. Bouvier, P. Susini, C. Marquis-Favre, and N. Misdariis, “Étude de la saillance auditive : analyse de l’influence de la brillance par la méthode du singleton additionnel.” [Online]. Available : <https://hal.science/hal-03848382> (Cited on page 68.)
- [129] C. Suied, P. Susini, and S. McAdams, “Evaluating warning sound urgency with reaction times.” *Journal of experimental psychology : applied*, vol. 14, no. 3, p. 201, 2008. (Cited on page 68.)
- [130] C. Lavandier and B. Defréville, “The contribution of sound source characteristics in the assessment of urban soundscapes,” vol. 92, no. 6, pp. 912–921. (Cited on page 68.)
- [131] G. Lafay, M. Rossignol, N. Misdariis, M. Lagrange, and J.-F. Petiot, “Investigating the perception of soundscapes

- through acoustic scene simulation,” vol. 51, no. 2, pp. 532–555. <http://doi.org/10.3758/s13428-018-1138-0>. [Online]. Available : <https://doi.org/10.3758/s13428-018-1138-0> (Cited on page 68.)
- [132] J.-F. Petiot, K. Legeay, and M. Lagrange, “Optimization of the sound of electric vehicles according to unpleasantness and detectability,” in *Proceedings of the Design Society : International Conference on Engineering Design*, vol. 1, no. 1. Cambridge University Press, 2019, pp. 3949–3958. (Cited on page 69.)
- [133] D. Gabor, “Acoustical quanta and the theory of hearing,” *Nature*, vol. 159, pp. 591–594, 1947. (Cited on pages 72 and 83.)
- [134] C. Roads, “Introduction to granular synthesis,” *Computer Music Journal*, vol. 12, no. 2, pp. 11–13, 1988. (Cited on page 72.)
- [135] S. Siddiq, “Morphing of granular sounds,” in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)*, 2015, pp. 4–11. (Cited on page 72.)
- [136] S. Denjean, R. Kronland-Martinet, V. Roussarie, and S. Ystad, “Zero-emission vehicles sonification strategy based on shepard-risset glissando,” in *Perception, Representations, Image, Sound, Music : 14th International Symposium, CMMR 2019, Marseille, France, October 14–18, 2019, Revised Selected Papers 14*, 2021, pp. 709–724. (Cited on page 77.)
- [137] G. Fechner, “Elements of psychophysics. vol. i. henry holt editions in psychology,” 1966. (Cited on page 88.)
- [138] G. B. Henning, “Frequency discrimination of random-amplitude tones,” *The Journal of the Acoustical Society of America*, vol. 39, no. 2, pp. 336–339, 1966. (Cited on page 88.)
- [139] E. J. Allen and A. J. Oxenham, “Symmetric interactions and interference between pitch and timbre,” *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1371–1379, 2014. (Cited on page 88.)
- [140] P. Susini, G. Lemaitre, and S. McAdams, “Chap11-psychological measurement for sound description and evaluation,” in *Measurement With Persons*. Psychology Press, num Pages : 28. (Cited on page 88.)
- [141] R. A. Kendall and E. C. Carterette, “Verbal attributes of simultaneous wind instrument timbres : I. von bismarck’s adjectives,” *Music Perception*, vol. 10, no. 4, pp. 445–467, 1993. (Cited on page 88.)
- [142] G. Von Bismarck, “Timbre of steady sounds : A factorial investigation of its verbal attributes,” *Acta Acustica united with Acustica*, vol. 30, no. 3, pp. 146–159, 1974. (Cited on page 88.)
- [143] A. Zacharakis, K. Pasiadis, and J. D. Reiss, “An interlanguage study of musical timbre semantic dimensions and their acoustic correlates,” *Music Perception : An Interdisciplinary Journal*, vol. 31, no. 4, pp. 339–358, 2012. (Cited on page 88.)
- [144] C. Tarlao, P. Aumond, C. Lavandier, and C. Guastavino, “Converging towards a french translation of soundscape attributes : Insights from quebec and france,” *Applied Acoustics*, vol. 211, p. 109572, 2023. (Cited on page 88.)
- [145] S. Namba, S. Kuwano, T. Hashimoto, B. Berglund, Z. Da Rui, A. Schick, H. Hoege, and M. Florentine, “Verbal expression of emotional impression of sound a cross-cultural study,” *Journal of the Acoustical Society of Japan (E)*, vol. 12, no. 1, pp. 19–29, 1991. (Cited on page 88.)
- [146] D. Dal Palù, E. Buiatti, G. E. Puglisi, O. Houix, P. Susini, C. De Giorgi, and A. Astolfi, “The use of semantic differential scales in listening tests : A comparison between context and laboratory test conditions for the rolling sounds of office chairs,” *Applied Acoustics*, vol. 127, pp. 270–283, 2017. (Cited on page 88.)
- [147] M. Schütte, U. Müller, S. Sandrock, B. Griefahn, C. Lavandier, and B. Barbot, “Perceived quality features of aircraft sounds : An analysis of the measurement characteristics of a newly created semantic differential,” *Applied Acoustics*, vol. 70, no. 7, pp. 903–914, 2009. (Cited on page 88.)
- [148] P. Susini, O. Houix, N. Misdariis, B. Smith, and S. Langlois, “Instruction’s effect on semantic scale ratings of interior car sounds,” *Applied Acoustics*, vol. 70, no. 3, pp. 389–403, 2009. (Cited on page 88.)
- [149] M. G. Kendall and B. B. Smith, “On the method of paired comparisons,” *Biometrika*, vol. 31, no. 3/4, pp. 324–345, 1940. (Cited on page 89.)
- [150] P. Susini, S. McAdams, and S. Winsberg, “A multidimensional technique for sound quality assessment,” *Acta acustica united with Acustica*, vol. 85, no. 5, pp. 650–656, 1999. (Cited on page 89.)
- [151] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres : Common dimensions, specificities, and latent subject classes,” *Psychological research*, vol. 58, pp. 177–192, 1995. (Cited on page 89.)
- [152] A. Caclin, S. McAdams, B. K. Smith, and S. Winsberg, “Acoustic correlates of timbre space dimensions : A confirmatory study using synthetic tones,” *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 471–482, 2005. (Cited on page 89.)

- [153] S. Winsberg and G. De Soete, “A latent class approach to fitting the weighted euclidean model, clascal,” *Psychometrika*, vol. 58, no. 2, pp. 315–330, 1993. (Cited on page 89.)
- [154] J. J. Louviere, T. N. Flynn, and A. A. J. Marley, *Best-worst scaling : Theory, methods and applications*. Cambridge University Press, 2015. (Cited on page 89.)
- [155] G. Hollis, “Scoring best-worst data in unbalanced many-item designs, with applications to crowdsourcing semantic judgments,” *Behavior research methods*, vol. 50, no. 2, pp. 711–729, 2018. (Cited on pages 89, 90, 91, 94, 101, 102, 104, and 119.)
- [156] A. E. Elo, “The international chess federation rating system.” Chess, 1973. (Cited on page 90.)
- [157] V. Rosi, A. Ravillion, O. Houix, and P. Susini, “Best-worst scaling, an alternative method to assess perceptual sound qualities,” *JASA Express Letters*, vol. 2, no. 6, 2022. (Cited on pages 91, 102, and 103.)
- [158] R. San Millán-Castillo, E. Latorre-Iglesias, M. Glesser, S. Wanty, D. Jiménez-Caminero, and J. M. Álvarez-Jimeno, “Mosquito : An open-source and free toolbox for sound quality metrics in the industry and education,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 263. Institute of Noise Control Engineering, 2021, pp. 1164–1175. (Cited on pages 95 and 137.)
- [159] “Psychoacoustics : Facts and Models - Eberhard Zwicker, Hugo Fastl - Google Livres.” (Cited on pages 95 and 97.)
- [160] J. R. De Leeuw, “jspsych : A javascript library for creating behavioral experiments in a web browser,” *Behavior research methods*, vol. 47, pp. 1–12, 2015. (Cited on page 98.)
- [161] P. W. Donhauser and D. Klein, “Audio-tokens : a toolbox for rating, sorting and comparing audio samples in the browser,” *Behavior research methods*, vol. 55, no. 2, pp. 508–515, 2023. (Cited on page 98.)
- [162] J.-F. Delmas, *Introduction au calcul des probabilités et à la statistique*. Les Presses de l’ENSTA, 2013. (Cited on page 118.)
- [163] U. Sandberg, *Tyre/road noise : myths and realities*. Statens väg- och transportforskningsinstitut, 2001. (Cited on page 133.)
- [164] A. Kuijpers and G. Van Blokland, “Tyre/road noise models in the last two decades : a critical evaluation,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 2001, no. 2. Institute of Noise Control Engineering, 2001, pp. 2494–2499. (Cited on page 133.)
- [165] V. Prakash, O. Sauvage, J. Antoni, and L. Gagliardini, “Assessment of electric powertrain whining noise contribution under early-stage design uncertainties,” 2023. (Cited on page 133.)
- [166] O. Harris, P. Lanlois, and A. Gale, “Electric Vehicle Whine Noise - Gear Blank Tuning as an Optimization Option,” 2019. (Cited on page 133.)
- [167] R. Sottek, “Progress in calculating tonality of technical sounds,” p. 9, 2014. (Cited on page 133.)
- [168] S. Wang, J.-L. Jouvray, and T. Kalos, “NVH Technologies and Challenges on Electric Powertrain,” SAE International, Warrendale, PA, SAE Technical Paper 2018-01-1551, Jun. 2018. <http://doi.org/10.4271/2018-01-1551> (Cited on page 133.)
- [169] M. Helfer, *General Aspects of Vehicle Aeroacoustics*, Jun. 2005. (Cited on page 133.)
- [170] J. Park and H. Lee, “Numerical Simulations of Whistle Noise from Air Ventilation Duct,” SAE International, Warrendale, PA, SAE Technical Paper 2005-01-2496, May 2005, iISSN : 0148-7191, 2688-3627. <http://doi.org/10.4271/2005-01-2496> (Cited on page 133.)
- [171] K. Qian, J. Wang, Y. Gao, Q. Sun, and J. Liang, “Interior noise and vibration prediction of permanent magnet synchronous motor,” *Journal of Vibroengineering*, vol. 20, no. 5, pp. 2225–2236, Aug. 2018, number : 5 Publisher : JVE International Ltd. <http://doi.org/10.21595/jve.2018.18605> (Cited on page 134.)
- [172] K. Genuit, “Analysis of annoyance of low-frequency beat inside a vehicle,” *Journal of The Acoustical Society of America - J ACOUST SOC AMER*, vol. 103, pp. 2849–2850, May 1998. <http://doi.org/10.1121/1.421976> (Cited on page 134.)
- [173] S. Hatano and T. Hashimoto, “Booming index as a measure for evaluatin booming sensation,” 2000. (Cited on page 134.)
- [174] F. Doleschal, G.-T. Badel, and J. L. Verhey, “Rumbling, humming, booming – Perception of vehicle interior noise at low engine speeds,” *Applied Acoustics*, vol. 197, p. 108915, Aug. 2022. <http://doi.org/10.1016/j.apacoust.2022.108915> (Cited on page 134.)
- [175] G. C. Peng, “SEA Modeling of Vehicle Wind Noise and Load Case Representation,” SAE International, Warrendale, PA, SAE Technical Paper 2007-01-2304, May 2007, iISSN : 0148-7191, 2688-3627. <http://doi.org/10.4271/2007-01-2304> (Cited on page 134.)

- [176] W. B. Coney, J. Y. Her, and J. A. Moore, "Characterization of the Wind Noise Loading of Production Automobile Greenhouse Surfaces." American Society of Mechanical Engineers Digital Collection, Mar. 2022, pp. 411–418. <http://doi.org/10.1115/IMECE1997-0084> (Cited on page 134.)
- [177] L. R. Molisani, R. A. Burdisso, and D. Tsihlias, "A coupled tire structure/acoustic cavity model," *International Journal of Solids and Structures*, vol. 40, no. 19, pp. 5125–5138, Sep. 2003. [http://doi.org/10.1016/S0020-7683\(03\)00259-2](http://doi.org/10.1016/S0020-7683(03)00259-2) (Cited on page 134.)
- [178] J. Yi, X. Liu, Y. Shan, and H. Dong, "Characteristics of sound pressure in the tire cavity arising from acoustic cavity resonance excited by road roughness," *Applied Acoustics*, vol. 146, pp. 218–226, Mar. 2019. <http://doi.org/10.1016/j.apacoust.2018.11.025> (Cited on page 134.)
- [179] Y. Fang and T. Zhang, "Sound Quality of the Acoustic Noise Radiated by PWM-Fed Electric Powertrain," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4534–4541, Jun. 2018, conference Name : IEEE Transactions on Industrial Electronics. <http://doi.org/10.1109/TIE.2017.2767558> (Cited on page 135.)
- [180] M. Sarrazin, K. Janssens, and H. Van der Auweraer, "Influence of inverter PWM control schemes on noise signature of electric powertrains," vol. 4, pp. 2899–2906, Jan. 2013. (Cited on page 135.)
- [181] A. Peyghambari, A. Dastfan, and A. Ahmadifard, "Noise spectrum shaping in power electronic converters by New RPWM technique," in *The 5th Annual International Power Electronics, Drive Systems and Technologies Conference (PEDSTC 2014)*, Feb. 2014, pp. 233–237. <http://doi.org/10.1109/PEDSTC.2014.6799377> (Cited on page 135.)
- [182] S.-H. Na, Y.-G. Jung, Y.-C. Lim, and S.-H. Yang, "Reduction of audible switching noise in induction motor drives using random position space vector PWM," *IEE Proceedings - Electric Power Applications*, vol. 149, no. 3, pp. 195–200, May 2002, publisher : IET Digital Library. (Cited on page 135.)
- [183] H. Yamaguchi, M. Shibaoka, and A. Kudo, "Study and Application of Prediction Method for Low Frequency Road Noise," SAE International, Warrendale, PA, SAE Technical Paper 2010-01-0507, Apr. 2010, iSSN : 0148-7191, 2688-3627. <http://doi.org/10.4271/2010-01-0507> (Cited on page 135.)
- [184] S. Gong, B. Peng, D. Chen, Y. Tang, and P. Wang, "Diagnosis of Road-Induced Drumming Noise of Passenger Car Based on Multiple Coherence Method," in *Proceedings of China SAE Congress 2018 : Selected Papers*, ser. Lecture Notes in Electrical Engineering, C. S. o. A. E. (China SAE), Ed. Singapore : Springer, 2020, pp. 737–748. (Cited on page 135.)
- [185] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *The Journal of the Acoustical Society of America*, vol. 71, no. 3, pp. 679–688, Mar. 1982. <http://doi.org/10.1121/1.387544>. [Online]. Available : <http://asa.scitation.org/doi/10.1121/1.387544> (Cited on page 137.)
- [186] P. Charollais, "ECMA-74, 17th edition, December 2019," -, p. 172, 2017. (Cited on pages 137, 140, and 142.)
- [187] E. Parizet, A. Bolmont, and S. Fingerhuth, "Subjective evaluation of tonalness and relation between tonalness and unpleasantness," -, vol. 3, Aug. 2009. (Cited on page 137.)
- [188] Yost, "Pitch strength of iterated rippled noise : The Journal of the Acoustical Society of America : Vol 100, No 5," 1996. [Online]. Available : <https://asa.scitation.org/doi/abs/10.1121/1.416973> (Cited on pages 137, 138, and 144.)
- [189] W. A. Yost, "Pitch of iterated rippled noise," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 511–518, Jul. 1996, publisher : Acoustical Society of America. <http://doi.org/10.1121/1.415873>. [Online]. Available : <https://asa.scitation.org/doi/abs/10.1121/1.415873> (Cited on pages 137 and 138.)
- [190] S. Wanty, M. Glessner, and F. CASAGRANDE HIRONO, "Ecma-418-2 roughness, a challenging implementation," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 270, no. 6. Institute of Noise Control Engineering, 2024, pp. 5205–5215. (Cited on page 141.)
- [191] R. Sottek, F. Kamp, and A. Fiebig, "A new hearing model approach to tonality," *Proc. Internoise, Innsbruck*, 2013. (Cited on page 146.)



# De l'étude du confort sonore de l'habitacle du véhicule électrique à son amélioration par l'acoustique augmentée

Matthieu Duroyon

Thèse de Doctorat en Acoustique

---

Les progrès récents de l'insonorisation des habitacles dans l'automobile ainsi que l'avènement de la motorisation électrique, silencieuse, ont durablement transformé l'expérience sonore de l'utilisateur et remettent en question les acquis du confort, et notamment du confort sonore. Cette thèse a pour objectif d'explorer le confort sonore dans cet environnement. Pour cela un travail de littérature scientifique permet de clarifier et définir cette notion de confort sonore, et de mettre en évidence les outils adaptés pour traiter et identifier l'émergence de sources sonores, auparavant masquées ou inexistantes tout en prenant en compte la place particulière du son dans un milieu comme l'automobile.

Cette thèse se sépare en trois parties distinctes et complémentaires. La première aborde le point de la description de l'environnement sonore à travers un travail d'interviews avec des spécialistes automobiles du son, sur les descripteurs numérique et sémantique du son. Ce travail aboutit sur la construction d'un lexique en suivant un paradigme élaboré au sein de l'équipe PDS (Perception et Design Sonores), le paradigme SpeaK. La deuxième partie de la thèse s'attelle à améliorer le confort sonore d'une situation sonore précise, comportant un son décrit comme désagréable ou gênant. La solution explorée consiste à ajouter des sons conçus suivant différentes typologies de masquage, appelées masquage énergétique, habillage et masquage attentionnel. Parmi ces différentes typologies, des masques sont alors créés grâce à un synthétiseur paramétrique développé spécifiquement pour cette tâche. Une étude perceptive en atmosphère contrôlée, utilisant une méthodologie Best-Worst Scaling est alors menée. Cette étude, menée avec 50 participants, permet de quantifier les différences entre ces méthodologies sur deux dimensions évaluées : la détectabilité du défaut et l'agréabilité de la scène. À l'issue de cette étude, une perspective d'expérience in-situ pour confronter les résultats, ainsi que la mise en lumière de limites et d'améliorations possibles de la méthode expérimentale sont proposées.