



HAL
open science

Conception d'un outil de diagnostic de la gêne sonore en milieu urbain

Raphaël Leiba

► **To cite this version:**

Raphaël Leiba. Conception d'un outil de diagnostic de la gêne sonore en milieu urbain. Acoustique [physics.class-ph]. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066602 . tel-01688847v2

HAL Id: tel-01688847

<https://hal.sorbonne-universite.fr/tel-01688847v2>

Submitted on 18 Feb 2019

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

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

École doctorale n° 391 : Sciences mécaniques, acoustique,
électronique et robotique de Paris (SMAER)

THÈSE

pour obtenir le grade de docteur de

l'Université Pierre et Marie Curie

Spécialité "Acoustique"

**Conception d'un outil de diagnostic
de la gêne sonore en milieu urbain**

présentée et soutenue publiquement par

Raphaël LEIBA

le 19 décembre 2017 devant un jury composé de :

Rapporteur,	Quentin LECLERE	Maître de Conférences, HDR, INSA Lyon
Rapporteuse,	Catherine MARQUIS-FAVRE	Directrice de Recherche, ENTPE
Examineur,	Quentin GRIMAL	Professeur, UPMC
Examinatrice,	Catherine LAVANDIER	Professeur, Univ. Cergy-Pontoise
Examineur,	Judicaël PICAUT	Directeur de Recherche, IFSTTAR
Examineur,	Vincent ROUSSARIE	Ingénieur de Recherche, PSA Peugeot-Citroën
Directeur de thèse, Invité,	François OLLIVIER	Maître de Conférence, HDR, UPMC
Co-directeur de thèse, Invité,	Régis MARCHIANO	Professeur, UPMC
Co-encadrant de thèse, Invité,	Nicolas MISDARIIS	Chargé de Recherche, IRCAM
Co-encadrant de thèse, Invité,	Jacques MARCHAL	Maître de Conférence, UPMC

Institut Jean le Rond d'Alembert
UMR 7190
4, Place Jussieu
75252 Paris Cedex 05, France

STMS Ircam-CNRS-UPMC
UMR 9912
1, place Igor-Stravinsky
75004 Paris, France

*“Le vacarme de l’humanité est intolérable, et la
confusion est telle qu’on ne peut plus dormir.”*

L’Épopée de Gilgamesh (environ 3 000 ans av. J.-C.)

Résumé

Le bruit, en particulier celui dû au trafic routier, est cité par de nombreuses études comme une source de préoccupation sociétale majeure. Jusqu'à présent les réponses des pouvoirs publics ne se basent que sur une quantification énergétique de l'exposition sonore, souvent via la mesure ou l'estimation du L_A ou du L_{den} , et des prises de décisions relatives à la diminution du niveau sonore. Or des études psychoacoustiques ont montré que le niveau sonore n'expliquait qu'une faible part de la gêne sonore ressentie. Il est donc intéressant d'avoir plus d'information sur la source de bruit et de ne pas la réduire à un simple niveau sonore.

Dans cette thèse, nous proposons de concevoir un outil permettant d'estimer la gêne sonore associée à chaque véhicule du trafic routier via l'utilisation de son signal audio et de modèles de gêne sonore.

Pour ce faire, le signal audio du véhicule est isolé de l'ensemble du trafic routier urbain grâce à l'utilisation de méthodes inverses et de grand réseaux de microphones ainsi que du traitement d'images pour obtenir sa trajectoire. Grâce à la connaissance de la trajectoire ainsi que du signal, le véhicule est classifié par une méthode de *machine learning* suivant la taxonomie de Morel *et al.*. Une fois sa catégorie obtenue, la gêne spécifique du véhicule est estimée grâce à un modèle de gêne sonore utilisant des indices psychoacoustiques et énergétiques. Cela permet l'estimation des gênes sonores spécifiques à chaque véhicule au sein du trafic routier. L'application de cette méthode est faite lors d'une journée de mesure sur une grande artère parisienne.

Mots-clé : Milieu urbain, gêne sonore, séparation de source, grand réseau de microphone, apprentissage machine, perception sonore, indices psycho-acoustiques

Noise annoyance diagnostic tool conception in urban areas

Abstract

Noise, especially road traffic noise, is cited by many studies as a source of major societal concern. So far, public responses are based only on energy quantification of sound exposure, often by measuring or estimating L_A or L_{den} , and sound-level reduction related decision are taken. Nevertheless, psychoacoustic studies have shown that the sound level explains only a small part of the perceived noise annoyance. It is interesting to have more information about the source of noise and not to reduce the information to its sound level.

In this thesis a tool is proposed for estimating the noise annoyance induced by each road vehicle using its audio signal and noise annoyance models.

To do so, the audio signal of the vehicle is isolated by using inverse methods, large microphone arrays and image processing to obtain its trajectory. The knowledge of the trajectory and of the signal allows the vehicle to be classified by a machine learning method according to Morel *et al.* taxonomy. Once its category obtained, the specific annoyance of the vehicle is estimated thanks to a noise annoyance model using psychoacoustic and energetic indices. This allows the estimation of specific noise annoyance for each vehicle within the road traffic. The application of this method is made during a measurement day on a large Parisian artery.

Keywords : Urban environment, noise annoyance, source separation, large microphone array, machine learning, sound perception, psychoacoustic indices

Remerciements

Cette thèse s'est déroulée entre l'équipe Perception et Design Sonores de l'UMR STMS de l'IRCAM et l'équipe Modélisation Propagation et Imagerie Acoustique de l'Institut Jean le Rond d'Alembert. Elle s'est donc déroulée sur trois sites : l'IRCAM, le Campus Jussieu et celui de Saint-Cyr-l'École. Sur chacun de ces sites l'accueil fut très bon et l'ambiance propice à la réalisation de cette thèse. Je voudrais ainsi remercier les différents services administratifs et financiers, avec un remerciement particulier à Anne MARCHAL et Simona OTARASANU pour leur réactivité et leur aide précieuse à chaque instant.

De plus, cette thèse se basant sur des expériences, elle n'aurait pas pu être réalisée par un doctorant isolé. Et c'est pour cela que je souhaite remercier très chaleureusement la "*Saint-Cyr Tech Team*" qui a toujours répondu présent et su se rendre disponible pour aider à monter et participer aux expériences. Je pense particulièrement à Dominique BUSQUET, Hélène MOINGEON et Christian OLLIVON qui ont toujours su anticiper les éventuels problèmes inhérents à des expériences de grande échelle, d'autant plus en plein air.

J'aimerais aussi remercier Pascal CHALANDE car il a toujours su se rendre disponible pour apporter des ajouts au module d'acquisition *Megamicros* pour les différents besoins des expériences. Je tiens aussi à remercier Vincent ROUSSARIE et Jean-Christophe CHAMARD de PSA Peugeot-Citroën pour nous avoir permis de réaliser l'expérience avec véhicules isolés sur la piste de La Ferté-Vidame et nous avoir aidé à l'organiser.

J'en viens à remercier mes encadrants. Cette thèse n'aurait pas été la même sans leur aide. Chacun ayant apporté son point de vue, tantôt novice, tantôt expert en fonction des sujets, ils ont permis de structurer ce travail de thèse. À force de persévérance, document après document, leurs avis et corrections orthographiques, m'ont permis de peut-être, enfin, avoir compris les règles de base de la grammaire française. Ils ont aussi été d'une grande aide dans la mise en place et la réalisation des expériences. Pour toute cette aide, ces nombreux échanges et toutes ces bonnes années passées ensemble, sincèrement, merci!

D'autre part, un laboratoire ne serait rien sans les doctorants et les post-doctorants. Cela est vrai pour le travail de recherche qu'ils réalisent, mais aussi pour l'ambiance qui règne au sein du laboratoire qui les accueille. J'ai pu avoir la chance d'être accueilli dans deux très bons laboratoires sur ces deux aspects. J'aimerais donc remercier tous les doctorants et post-doctorants avec qui j'ai pu interagir autour d'un verre ou d'un café, parlant de science, de "petits trucs", de conseils, de politique ou que sais-je encore. Ces remerciements se focalisent naturellement sur les doctorants avec qui j'ai partagé mes bureaux et plus particulièrement sur Charles VANWYNSBERGHE avec qui j'ai énormément échangé scientifiquement depuis mon arrivée à d'Alembert en stage de master 2 jusqu'à sa soutenance de thèse. Merci aussi pour ton aide lors de mes expériences ou en amont où tu as pris de ton temps pour m'aider et pour tous les bons moments partagés! Au plaisir de travailler de nouveau avec toi par la suite.

J'aimerais remercier chaleureusement les examinateurs de ce travail de thèse pour leur regard critique sur ce travail et sur leurs questions et remarques intéressantes qui apportent des pistes nombreuses d'améliorations de ce projet, voir de collaborations futures.

Cette thèse s'est déroulée dans le cadre de la Chaire d'excellence "Mobilité et Qualité de Vie en Milieu Urbain" financée par la Fondation UPMC et dont les mécènes étaient PSA Peugeot-Citroën et Renault. Je veux donc remercier tous les acteurs qui ont travaillé à la construction et à la vie de ce projet de recherche pluridisciplinaire.

Merci à mes parents pour leur soutien pendant ces trois années et particulièrement à ma mère qui a acceptée de relire l'intégralité du manuscrit pour chasser les fautes d'orthographe et d'accords.

En conclusion, cette thèse est le fruit d'une collaboration, d'un travail d'équipe. C'est à mon sens la seule forme de d'organisation de la recherche scientifique étique et efficace; la concurrence inter-individus ou inter-laboratoires n'étant que trop souvent destructrice d'espoir, broyeuse d'initiative et créatrice d'ego surdimensionnés. Ainsi, merci à vous toutes et tous qui y avez participé que ce soit scientifiquement, techniquement ou humainement!

Table des matières

Introduction	1
Objectifs et contributions originales de la thèse	10
Plan de la thèse	12
1 Modélisation de la gêne sonore	13
Introduction	13
1.1 Indices énergétiques et psychoacoustiques	14
1.1.1 Le niveau équivalent	14
1.1.2 Pondération fréquentielle	15
1.1.3 Pondération temporelle	17
1.1.4 Niveaux statistiques	18
1.1.5 Indices événementiels	18
1.1.6 La sonie	20
1.1.7 L'acuité	21
1.1.8 La force de fluctuation	22
1.1.9 La rugosité	23
1.2 Gêne sonore et modélisation	24
1.2.1 Facteurs influençant l'estimation de la gêne sonore	25
1.2.2 Modèles de gêne	27
Conclusion	33
2 Imagerie acoustique et extraction du signal audio de sources mobiles	35
Introduction	35
2.1 Méthodes inverses en antennerie acoustique	38
2.1.1 Modélisation du processus direct	39
2.1.2 La formation de voies, ou <i>beamforming</i>	40
2.2 Mégamicros, un système d'acquisition modulaire	44
2.2.1 Le système <i>Megamicros</i>	45
2.2.2 Une géométrie d'antenne pour chaque application	46
2.3 Localisation et quantification des sources de bruit	50
2.3.1 Véhicules légers	52
2.3.2 Autres véhicules routiers	58
2.4 L'antennerie acoustique, outil de séparation de sources	60
2.4.1 Extraction du signal audio d'une source fixe	60
2.4.2 Extraction du signal audio d'un véhicule en mouvement	61

Conclusion	67
3 Classification du trafic routier par apprentissage supervisé	69
Introduction	69
3.1 Classification de signaux audio	70
3.1.1 Modèles et paramètres pertinents	70
3.2 Classification des véhicules par <i>Support Vector Machines</i>	72
3.2.1 Principe des SVM	72
3.2.2 Application des SVM à la base de donnée de véhicules isolés	74
Conclusion	79
4 Implémentation du modèle de gêne sonore	81
Introduction	81
4.1 Modèles de calcul d'indices psychoacoustiques	82
4.1.1 Calcul de la sonie	82
4.1.2 Calcul de la rugosité	84
4.1.3 Calcul de la force de fluctuation	88
4.2 Implémentation du modèle de Morel <i>et al.</i> et application à une mesure de référence	92
Conclusion	95
5 Estimation <i>in situ</i> de la gêne spécifique de chaque véhicule du trafic routier	97
Introduction	97
5.1 Extraction du signal audio d'un véhicule dans la ville	98
5.1.1 Dispositif expérimental	98
5.1.2 Du pixel à la position 3D	99
5.1.3 Séparation de sources en milieu urbain	100
5.2 Classification <i>in situ</i>	103
5.2.1 Classification avec apprentissage sur véhicules isolés	104
5.2.2 Classification avec apprentissage sur données acquises dans la ville	105
5.2.3 Influence du noyau du classifieur	105
5.3 Estimation de la gêne sonore <i>in situ</i>	110
Conclusion	112
Conclusion	115
Perspectives	116
Annexe : Localisation de sources sur deux-roues et véhicules lourds	129
Véhicules lourds	129
Vitesses constantes	129
Accélération et décélération	132
Deux-roues motorisés	134
Vitesses constantes	134
Accélération et décélération	137
Deux-roues électrique	140

Liste des symboles

Paramètres acoustiques

- $L_{eq,T}$ Niveau sonore équivalent intégré sur un période T, exprimé en dB
- $L_{A,T}$ Niveau sonore équivalent pondéré A intégré sur un période T, exprimé en dB(A)
- L_{den} Niveau sonore jour, soir, nuit (pondéré fréquentiellement et temporellement), exprimé en dB(A)
- L_n Niveau sonore nocturne (pondéré A et intégré entre 22h et 6h), exprimé en dB(A)
- L_α Niveau statistique : Niveau sonore dépassé $\alpha\%$ du temps, généralement calculé à partir du $L_{A,1s}$
- c_0 Célérité du son dans le milieu
- M Nombre de microphones de l'antenne
- I Nombre de sources de la scène sonore
- $s_i(t)$ Pression rayonnée par la source i en fonction du temps t
- $G_{im}(t)$ Fonction de Green entre la source i et le microphone m
- r_{im} Distance euclidienne entre la source i et le microphone m
- τ_{im} Temps de propagation entre la source i et le microphone m
- $p_m(t)$ Pression mesurée par le microphone m
- $\hat{s}_i(f)$ Transformée de Fourier de la pression rayonnée par la source i
- $\hat{p}_m(f)$ Transformée de Fourier de la pression mesurée par le microphone m
- A** Matrice de propagation de dimension $M \times I$
- f Fréquence
- ω Pulsation, $\omega = 2\pi f$
- k Nombre d'onde $k = \frac{\omega}{c_0} = \frac{2\pi f}{c_0}$
- λ Longueur d'onde $\lambda = \frac{c_0}{f}$
- d Espace inter-microphone
- θ_i Angle d'arrivée de l'onde de la source i sur une antenne linéaire

- $2\theta_{3dB}$ Ouverture angulaire à -3 dB du maximum d'une antenne
- R** Matrice d'inter-correlation, de dimension $M \times M$ et calculée pour la fréquence f
- $\tilde{s}_i(t)$ Estimation du signal de pression de la source i
- \tilde{s}_i^κ Estimation du signal de pression de la source i pour la trame temporelle κ
- T_f Durée d'une trame temporelle d'analyse et de calcul de la formation de voies
- T_s Temps entre deux trames temporelles d'analyse
- $w_s(t)$ Fenêtre de pondération d'une trame temporelle d'analyse

Paramètres psychoacoustiques

- N** Sonie du son, exprimée en sonnes
- N'** Sonie partielle dans une bande critique
- L_N** Niveau de sonie, exprimé en phones
- S** Acuité du son
- F** Force de fluctuation du son
- R** Rugosité du son
- f_c Fréquence centrale du signal
- f_{mod} Fréquence de modulation de l'enveloppe du signal
- m Degré de modulation de l'enveloppe du signal (entre 0 et 1)
- ΔL Profondeur de masquage temporel perçue
- A_i, A_{tot} Gêne sonore spécifique à la source ou à la catégorie de sources i et gêne sonore totale de la scène sonore
- m_i^* Degré de modulation généralisé dans le modèle d'Aures
- k_i, k_{i-2} Coefficient de corrélation entre les degrés de modulation m_i^* des bandes critiques supérieur (m_{i+2}^*) et inférieurs (m_{i-2}^*)
- $g(z_i)$ Coefficient de pondération fréquentielle pour chaque demi-bande critique i , z_i étant le nombre de Bark
- c_R, c_{FS} Coefficient de normalisation des modèles de rugosité et de force de fluctuation

Paramètres d'apprentissage machine

- X** Ensemble des n vecteurs \mathbf{x}_i , avec n le nombre d'échantillons
- \mathbf{x}_i Vecteurs de caractéristique $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})$, avec p le nombre de caractéristiques du signal
- w_i** Vecteur normal de l'hyperplan de séparation des catégories
- y_i Catégorie de l'échantillon i

- $K(\mathbf{x}_i, \mathbf{x}_j)$ Noyau de calcul de la machine à vecteurs de support (SVM), pour deux échantillons \mathbf{x}_i et \mathbf{x}_j
- C Paramètre de contrôle du compromis entre nombre d'erreurs de classement, et largeur de la marge inter-catégorie
- d Paramètre du noyau polynomial, déterminant le degré du polynôme
- γ Paramètre du noyau gaussien, déterminant la largeur de la gaussienne
- \mathbf{W} Vecteur de pondération des données d'apprentissage à partir de données normalisées

Autres symboles

- F_e Fréquence d'échantillonnage d'un signal
- \mathcal{F} Transformée de Fourier
- V Vitesse d'avancement du véhicule

Introduction

Contrairement aux autres modalités sensorielles, il n'est pas possible d'inhiber l'ouïe. L'environnement sonore d'un être humain – qui peut varier en fonction de ses activités, de son lieu de vie ou encore de son voisinage – s'impose donc à lui. Depuis la révolution industrielle, le nombre de textes rapportant des environnements sonores trop bruyants a largement augmenté. En effet, jusqu'alors, seules les villes concentraient les problèmes de bruits, principalement dus à la densité de population. Dès lors, de nouveaux environnements sonores s'imposent dans les banlieues – qui accueillent la population grandissante issue de l'exode rural – et provinces qui deviennent industrialisées. Depuis, de nombreuses études auprès des populations montrent toujours l'importance du bruit dans les préoccupations populaires.

Selon un sondage commandé en 2014 par le Ministère de l'Écologie, du Développement Durable et de l'Énergie¹, 82% des Français indiquent se préoccuper des nuisances sonores, et plus d'un tiers (35%) affirme même s'en préoccuper "tout à fait". De plus, d'après cette étude, cette préoccupation est partagée par toute la population car l'écart n'est que de 8 points entre les habitants de communes rurales (78%) et les habitants des grandes agglomérations, celle de Paris étant en tête avec 86%.

Et cela n'est pas uniquement dû aux industries et modes de transport. En effet, un sondage réalisé pour la Journée nationale de l'audition (JNA) en 2008² montrait que 51% des français sont au moins un peu gênés par le bruit au travail, dont 31% "souvent" et "très fréquemment".

Un sondage réalisé pour la JNA 2016³ confirme ces chiffres et ajoute que la quasi-totalité des personnes sondées (94%) pense que le bruit a des effets directs sur la santé et près de la moitié en est intimement convaincue (48%).

1. Sondage IFOP de septembre 2014 sur un échantillon de 1001 personnes, représentatif de la population française âgée de 18 ans et plus. La représentativité de l'échantillon a été assurée par la méthode des quotas (sexe, âge, catégorie socio-professionnelle) après stratification par région et catégorie d'agglomération. Questionnaire auto-administré en ligne du 1er septembre au 3 septembre 2014.

2. Étude réalisée par IPSOS auprès de 1613 individus âgés de 18 ans et plus exerçant une activité professionnelle entre le 30 novembre et le 16 décembre 2007. Quotas représentatifs de la population française en termes de sexe, âge, CSP du chef de famille, région et catégorie de commune. Les résultats ont été redressés selon la structure de la population active en France.

3. Enquête IFOP auprès d'un échantillon de 1003 personnes, représentatif de la population française âgée de 15 ans et plus. Méthode des quotas (sexe, âge, profession de la personne interrogée) après stratification par région et catégorie d'agglomération.

Conséquences sanitaires du bruit. Les effets sur le système auditif sont souvent pointés lors de ces enquêtes d'opinion. Ils sont effectivement connus et avérés : l'exposition prolongée à des niveaux sonores trop importants induit la destruction des cellules ciliées dans l'oreille interne et une perte d'audition. Mais le bruit peut aussi avoir des répercussions sur la santé via le stress accru dont il est responsable.

Selon l'étude CE Delft de 2007[37] – qui s'intéresse notamment aux effets sur la santé des trafics routier et ferroviaire – environ 50 000 personnes dans l'Union Européenne⁴ meurent prématurément chaque année de crises cardiaques imputables au bruit dû aux trafics routier et ferroviaire et plus de 200 000 personnes souffrent de différents types de maladies cardiovasculaires liées au bruit de la circulation.

En 2011, l'OMS indiquait[42] qu'entre 1 million et 1,6 million d'années en bonne santé seraient perdues (en anglais : *disability-adjusted life-years*, ou DALYs) chaque année en Europe occidentale sous l'effet du bruit causé par les modes de transport, et principalement le bruit du trafic routier. D'après l'étude, parmi les facteurs de risque environnemental en Europe, la sur-exposition aux bruits serait ainsi la seconde cause de morbidité⁵ derrière la pollution atmosphérique. L'OMS indique qu'une personne sur trois est gênée en journée et une sur cinq voit son sommeil perturbé par le bruit routier, ferroviaire ou aérien, ce qui accroît le risque de maladies cardiovasculaires et d'hypertension. D'autres études épidémiologiques (comme [13, 113, 14]) mettent en avant le lien entre trafic routier et maladies cardiovasculaires .

Contexte réglementaire. Dans l'enquête citée précédemment (IFOP, 2016), l'action publique apparaît comme un levier important pour les Français. Par exemple, 89% des personnes interrogées indiquent qu'une meilleure gestion des nuisances sonores par les pouvoirs publics permettrait d'améliorer la qualité de vie et 85% indiquent que cela permettrait de bénéficier d'une meilleure santé.

Dans la figure 1, extraite du sondage IFOP pour la Journée Nationale de l'Audition 2014, le trafic routier est la première ou deuxième source de bruit principale pour 67% des personnes interrogées. L'autre grande préoccupation est le bruit de voisinage. En regardant dans le détail, cette répartition est globalement assez constante dans les différentes catégories d'agglomérations, sauf pour les habitants d'agglomération de 20 000 à 100 000 habitants qui placent en premier les bruits de voisinages devant les bruits routiers.

Ce constat n'est pas récent. En effet, la Communauté Économique Européenne s'est dotée en 1970 de la directive 70/157/CEE[1] instaurant le rapprochement des législations des États membres quant au niveau sonore admissible et au dispositif d'échappement des véhicules à moteur. Cette législation constitue le point de départ de la réglementation européenne sur le niveau de bruit. Elle concerne tous les véhicules à moteur capables de se déplacer à une vitesse supérieure à 25 km/h sur les routes. Pour les automobiles, la limite imposée est alors de 82 dB(A)⁶. Cette législation s'attaque alors à réduire l'influence d'une source du bruit : le trafic routier.

4. sauf Chypre et Malte

5. A ne pas confondre avec la mortalité qui mesure le nombre de décès, la morbidité quantifie le nombre de personnes atteintes par une pathologie

6. et de 91 dB(A) pour les véhicules utilitaires de grande puissance. Niveaux sonores pondérés A mesurés à 7,5 m des véhicules - Méthode similaire à celle de la norme ISO 362.

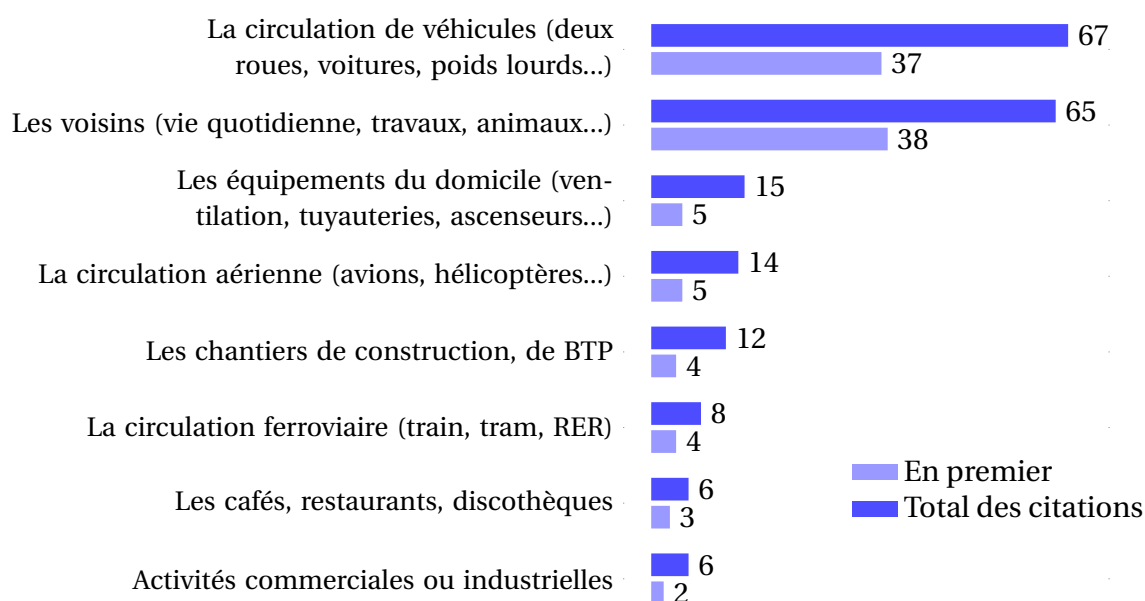


FIGURE 1 – Pourcentage de réponse à la question : A votre domicile, quelles sont les deux sources de bruit et nuisances sonores que vous considérez comme les plus gênantes? En premier? Et en second? - Question posée uniquement aux personnes déclarant être gênées par le bruit et les nuisances sonores à leur domicile, soit 86% de l'échantillon – Sondage IFOP, 2014.

En 1992, le parlement français a adopté une loi relative à la lutte contre le bruit. Elle se donne notamment comme objectif, dans son article premier, de "*prévenir, supprimer ou limiter l'émission ou la propagation sans nécessité ou par manque de précautions, des bruits ou des vibrations de nature à présenter des dangers, à causer un trouble excessif aux personnes, à nuire à leur santé ou à porter atteinte à l'environnement*". Toutes les notions qui seront reprises dans les dispositions suivantes sont déjà mentionnées dans cette loi. Elle est axée sur la protection de l'environnement au sens large en traitant aussi bien les transports terrestres qu'aériens. La réduction de l'impact de ces trafics sur les riverains est le cœur de la loi en prévoyant des travaux sur les infrastructures routières (article 15) et en mettant en place des taxes dont la recette est affectée à l'ADEME⁷ pour permettre aux riverains d'aéroport d'investir dans des travaux d'isolation phonique.

Il faut cependant attendre 2002 et la directive européenne 2002/49/CE [2] pour que l'Europe se dote d'outils objectifs de quantification et de lutte contre le bruit dans l'environnement. Traduite en droit français par l'ordonnance n°2004-1199 du 12 novembre 2004, cette directive impose la réalisation de cartes stratégiques du bruit. Elles sont suivies de plans d'action, intitulés plan de prévention du bruit dans l'environnement (PPBE) dans le code de l'environnement. L'ordonnance a ainsi imposé aux agglomérations de plus de 100 000 habitants de faire réaliser ces cartes avant le 30 juin 2012 (30 juin 2007 pour les agglomérations de plus de 250 000 habitants). Les cartes stratégiques présentent le niveau de bruit en fonction de sources diverses (trafic routier, ferroviaire, aérien et bruits dus aux industries) mais

7. Agence de l'Environnement et de la Maîtrise de l'Énergie

évaluent également le nombre de personnes exposées par tranche de niveau de bruit. Ces cartes doivent être fournies suivant deux indices :

- le niveau "*Day, Evening, Night*" (L_{den}) qui pondère les niveaux relevés en soirée et la nuit par rapport à ceux relevés en journée pour prendre en compte l'accroissement des nuisances dans ces périodes (détail du calcul en section 1.1.3),
- le niveau de nuit (L_n) mesuré de 22h à 6h.

Un exemple de carte normative est disponible en figure 2. On y voit le niveau L_{den} induit par le trafic routier pour la métropole du Grand Paris. Ce type de carte fournit une information facilement interprétable à destination des riverains des infrastructures de transport. Malheureusement, d'après un sondage cité précédemment (IFOP, 2014), seuls 19% des sondés connaissent l'existence de ces cartes de bruit.

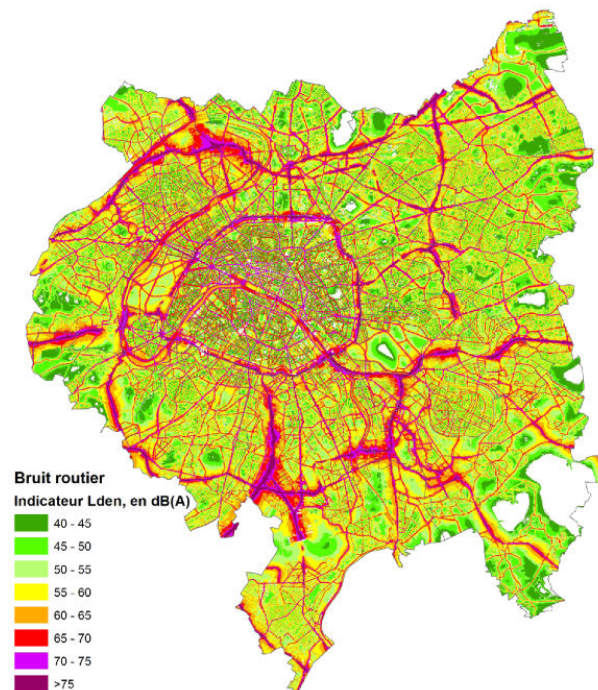


FIGURE 2 – Carte du bruit routier à l'échelle de la Métropole du Grand Paris - Mars 2017 - Source : Bruitparif.

Cette réglementation apporte de réelles avancées mais contient malgré tout des limitations. La première est la non-obligation de valider ces cartes par des mesures, elles sont donc le plus souvent uniquement issues de modèles de propagation, utilisant des données de flux de trafic, de revêtement de chaussée et de cadastre quand elles existent. De plus, jusqu'à présent les deux-roues motorisés étaient considérés comme des voitures dans les modèles générant les cartes, ce qui induisait une sous-estimation du niveau sonore. Cela est d'autant plus vrai que la part des deux-roues dans le trafic routier peut monter à 38% le matin sur certains tronçons proches de Paris[74], ce pourcentage diminuant en s'éloignant de la capitale. Une autre critique qui peut être faite est que cet outil de communication ne propose qu'une vision statique de l'environnement sonore urbain. Une carte heure par heure, avec

l'information du nombre et du type de véhicules pourrait être intéressante. Par ailleurs, cette réglementation ne prend pas en compte toutes les sources acoustiques comme les bruits résultants des activités domestiques, les bruits de voisinage ou autres, alors que ces bruits prennent une place importante dans les considérations des Français (voir figure 1).

Une autre réponse des pouvoirs publics est la création d'observatoires du bruit. Ayant pour but d'installer des stations de mesure fixes ou temporaires, ces observatoires permettent de connaître l'évolution du niveau sonore en fonction du temps en un lieu donné. C'est le cas, par exemple, avec la plateforme Rumeur de Bruitparif (rumeur.bruitparif.fr) qui propose d'avoir accès aux données de plus de 60 points de mesure en région parisienne⁸ ou encore [Acoucité](#), l'observatoire de l'environnement sonore de la Métropole de Lyon.

Ces approches réglementaires (diffusion de cartes de bruit ou développement de d'observatoires du bruit) sont une première réponse aux inquiétudes des citoyens mais ne permettent pas encore d'avoir une approche globale de l'environnement sonore.

Ainsi, des projets de recherche collaboratifs se mettent aussi en place depuis quelques années. On peut noter, par exemple, le Projet Cart_ASUR qui a proposé l'utilisation de smartphones pour la caractérisation physique et perceptive par des citoyens (noisetube.net/cartasur) de l'environnement sonore. L'un des objectifs a été de trouver des indicateurs perceptifs permettant de modéliser la qualité perçue de l'environnement sonore sur une échelle de 1 à 11 à partir de questionnaires remplis par les participants [90].

A mi-chemin de l'observatoire de bruit et des mesures participatives, des études permettant la caractérisation de l'environnement sonore urbain grâce à des mesures sonométriques mobiles ont commencé à émerger. Ce fut par exemple le cas dans le cadre des projets Eureka (Can *et al.* [29]) et GRAFIC (Aumond *et al.*[10]). Can *et al.* ont réalisé des mesures en continu ainsi que neuf arrêts d'une minute avec un sonomètre embarqué. Ils ont ainsi proposé une représentation suivant six groupes : quatre pour les périodes diurnes et deux pour les périodes nocturnes. Chacun de ces groupes se caractérise par différentes valeurs de trois indices qui représentent le niveau sonore médian, la variation du niveau sonore et la balance spectrale.

La figure 3 montre un exemple de cette représentation pour les périodes diurnes. On peut noter la présence du groupe 2 (croix violettes) proche de la voie rapide. Celui-ci est défini pour un niveau sonore médian élevé et variant faiblement. A l'inverse, dans de plus petites rues, on peut retrouver le groupe 3 (triangle orange) qui se caractérise par un niveau médian faible mais ayant une grande plage de variation, ce qui est caractéristique des petites rues avec peu de passage de véhicules routiers.

Ces deux dernières approches apportent une vision plus détaillée et plus complète de l'environnement sonore que ce soit avec des métriques physiques ou en associant mesures physiques et le ressenti des riverains.

8. nombre de points mesure permanents estimé sur <http://rumeur.bruitparif.fr> au 04/05/2017



FIGURE 3 – Distribution des environnements sonores en fonction des saisons basée sur les quatre groupes diurnes - Les groupes se caractérisent par différentes valeurs du niveau sonore médian, de la variation du niveau sonore et de la balance spectrale - De Can *et al.* [29]

Le concept de paysage sonore ou *Soundscape*. Raymond Murray Schafer [94] introduit le concept de paysage sonore dans les années 1970. Par analogie au paysage qui est “tout ce qui peut être vu” il décrit le paysage sonore comme “tout ce qui [peut] être entendu”. Il décrit l’étude du paysage sonore comme étant le lien entre tous les travaux en acoustique, psychoacoustique ou encore la limitation du bruit qui “chacun à leur manière [se posent] la même question : quelle est la relation entre l’homme et son environnement acoustique [...]?”. L’humain est donc au centre de ce concept de paysage sonore.

Voulant un programme positif de l’acoustique environnementale s’opposant à une recherche permanente de diminution de l’intensité sonore dans nos politiques publiques, il propose de répondre d’abord à la question de quelles sources voulons-nous “conserver, encourager ou multiplier”? Cela permettra, dit-il, de mieux cerner les bruits gênants ou nocifs et de savoir pourquoi il nous faut les éliminer.

Schafer, note que depuis le début du XX^{ème} siècle la musique a évolué et que “tout son est aujourd’hui, en permanence, susceptible d’entrer dans le domaine de la musique”. Il voit alors notre environnement sonore comme une composition musicale macroscopique. Pour appuyer son propos il va définir deux types de paysages sonores (par analogie à l’électroacoustique) : les environnements “hi-fi” et “lo-fi”. Il définit un environnement “hi-fi” en ces termes :

Le paysage sonore hi-fi est celui dans lequel chaque son est clairement perçu, en raison du faible niveau sonore ambiant. La campagne est généralement plus hi-fi que la ville, la nuit l’est plus que le jour,[...]. Dans un paysage sonore hi-fi, les sons se chevauchent moins fréquemment.

L’environnement “lo-fi” est donc, par opposition, un paysage sonore dans lequel les sources de bruit ne sont pas discernables les unes des autres, en d’autres termes, le rapport signal/bruit est faible. Il écrit que l’équilibre esthétique aurait été perdu dans nos villes au cours des siècles passant d’un environnement “hi-fi” à “lo-fi”.

Dans la lignée des réflexions de R.M. Schafer, Brigitte Schulte-Fortkamp [97] insiste sur le fait que réfléchir en termes de paysage sonore fait voir le bruit comme une ressource plutôt qu’un déchet ou une pollution⁹. Elle complète en ajoutant qu’une ressource est un facteur

9. *Soundscape understand noise rather as a “resource” than “waste”*

nécessaire pour accomplir une activité tout comme la terre, le travail ou le capital est une ressource pour l'économie.

Kozo Hiramatsu [53] ajoute que la réflexion en termes de paysage sonore représente une méthode comparable au contrôle du bruit en y ajoutant une amélioration des qualités psychologiques, esthétiques et sociales de l'environnement sonore. Schulte-Fortkamp [97] s'inscrit dans ses propos et rajoute que ce concept permet d'avoir une approche globale sur notre environnement sonore. Elle indique même "*It is important to distinguish the totality of Soundscape from the limited idea of quiet zone*"¹⁰.

Hiramatsu [53] confirme l'intérêt de cette approche par rapport aux cartes de bruit en L_{den} en insistant sur le fait que l'identification des sources de bruit et des événements sonores donne une information plus complète et plus sophistiquée.

Un projet de norme internationale est en cours (ISO 12913). La phase 1, finalisée en 2014 (ISO 12913-1:2014), a eu pour but de définir le concept de *soundscape* et de lui donner un cadre conceptuel. La norme définit alors le paysage sonore comme l'"environnement sonore tel que perçu, vécu et/ou compris, dans son contexte, par un (des) individu(s)".¹¹

Mesurer, représenter et diffuser le paysage sonore. La phase 2 de ce projet de norme ISO 12913, encore en cours de rédaction, va chercher à uniformiser les approches et les procédures de mesurage, les procédures d'analyse, de modélisation et de communication vers le grand public. En effet, la question de la cartographie du paysage sonore est encore ouverte.

R.M. Schafer [94] propose d'évaluer le paysage sonore d'un parc, d'un quartier ou d'une ville via des marches sonores. Il les définit en ces termes :

La marche sonore est une exploration du paysage sonore dans un lieu donné, guidée par une partition. La partition consiste en une carte indiquant le climat sonore et les sons inhabituels qui se présenteront tout au long du trajet.

Un exemple de ces partitions est présenté en figure 4. Cette carte propose une représentation possible des faits sonores telle que R.M. Schafer la pense comme support des marches sonores.

Hong et Jeon [54] proposent une méthode basée sur des prises de son de 5 minutes à partir d'un maillage d'une ville. Des sujets écoutent ensuite ces enregistrements et il leur est demandé de décrire la qualité de l'environnement sonore en distinguant quatre types de sons différents ainsi que l'intensité sonore perçue via un questionnaire. Ces enregistrements sont binauraux, c'est-à-dire que les signaux ont été enregistrés à l'entrée du canal auditif, soit d'un sujet, soit d'une tête artificielle. Pour évaluer la présence de chaque type de son¹² (bruits de trafic, sons d'origine humaine et sons naturels), ils proposent une échelle graduée

10. Il est important de distinguer l'ensemble du paysage sonore de l'idée limitée de zone tranquille

11. *Acoustic environment as perceived or experienced and/or understood by a person or people, in context*

12. Dans leur article, Hong et Jeon utilisent les termes *noise* et *sound* (respectivement bruit et son) sans expliciter la différence sémantique. Ils sont ici rapportés en l'état. Dans le reste du document, si le terme "son" est utilisé, ce sera pour parler de n'importe quelle onde acoustique entendue ou enregistrée par un microphone alors que le terme "bruit" est utilisé pour parler d'un son qui a été interprété par l'auditeur et potentiellement perçu comme inharmonique ou comme non-désiré. Les bruits sont donc un sous-ensemble des sons.

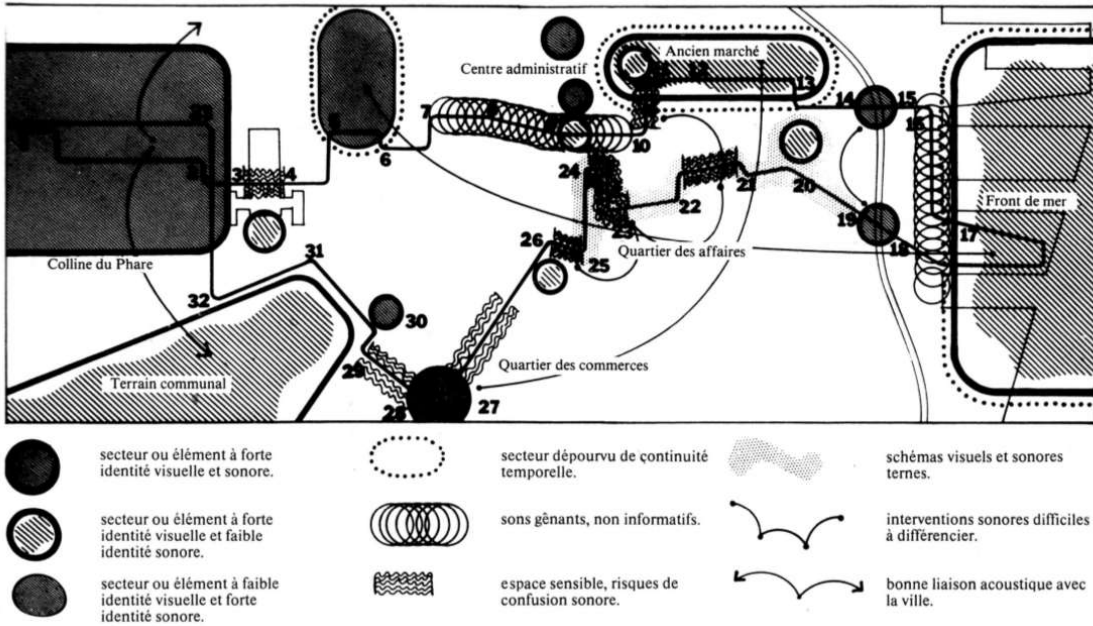


FIGURE 4 – Carte des faits sonores dans la ville de Boston (USA) par Michael Southworth, d’après [94].

à cinq niveaux allant de “n’entend pas du tout” à “domine complètement”. Cette méthodologie a été implémentée dans la ville de Séoul. Des exemples de leurs résultats sont présentés en figure 5. Ils sont issus de l’interpolation, par un logiciel de Système d’Information Géographique (SIG), des données obtenues pour 118 positions de mesures.

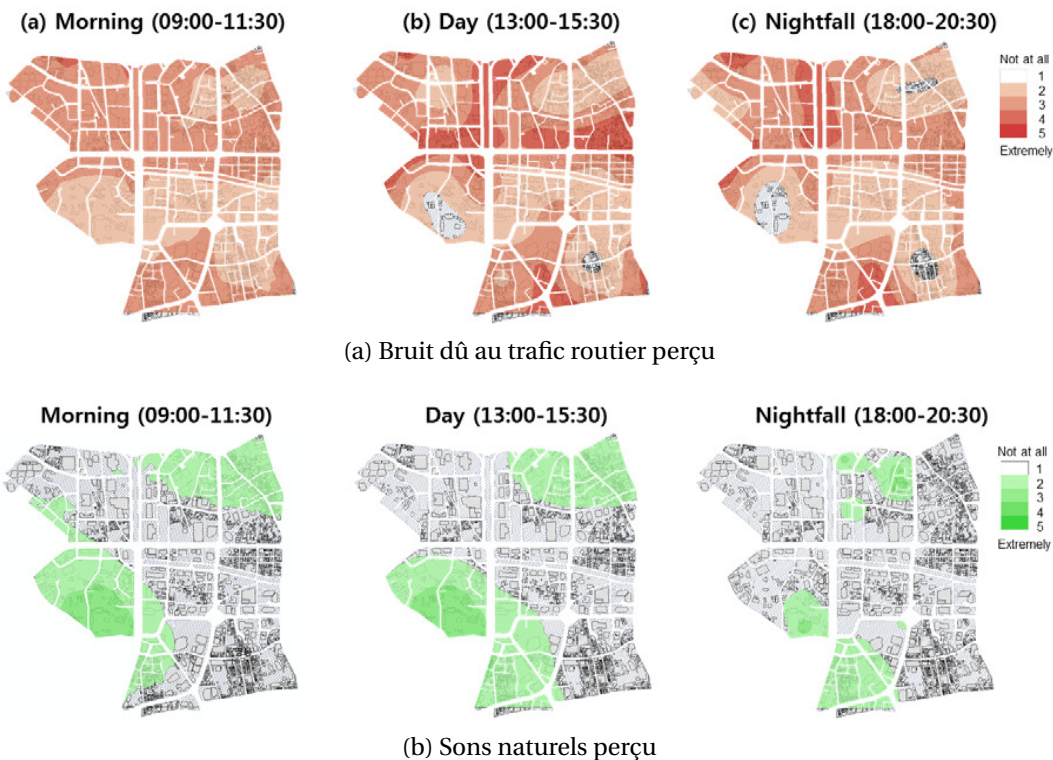


FIGURE 5 – Répartition spatiale de la perception de la présence de différents types de bruits à trois périodes de la journée - D’après Hong et Jeon [54].

S’inspirant du concept de paysage sonore et pour apporter une information globale sur les bruits présents dans la ville, une initiative anglaise et italienne appelée [Chatty Maps](#) propose une carte de douze villes dont Londres, New-York et Rome sur laquelle, pour chaque rue, les pourcentages de présence de chaque type de bruits sont représentés et les émotions qui y sont ressenties sont indiquées. Cette initiative, couplée à des enregistrements audio (ou audiovisuels) géolocalisés, permettrait une bonne immersion dans le paysage sonore.

A la suite des travaux sur le répertoire des effets sonores d’Augoyard *et al.*[9], le Centre de Recherche sur l’Espace Sonore et l’environnement urbain (CRESSON) a mis en ligne en 2012 un outil de découverte des paysages sonores via des enregistrements géolocalisés sur des vues satellites. Le projet s’appelle Cartophonies ([cartophonies.fr](#)) et permet grâce aux enregistrements, aux photos et aux descriptions de s’immerger efficacement dans l’ambiance sonore captée. D’autres bases de données de captations sonores géolocalisées comparables sont accessibles, comme par exemple à Montréal ([montrealsoundmap.com](#)). Toutes ces initiatives permettent de s’immerger dans l’environnement sonore, mais peut-on pour autant parler de quantification ou de mesure du paysage sonore ?

Avec le concept de paysage sonore, R.M. Schafer a mis au centre de sa réflexion - et de sa vision de l’environnement sonore - la source du bruit et insiste sur l’importance de connaître les bruits qui nous gênent et ceux qu’il faut encourager. Il semble donc important de mettre en place des méthodologies de mesure adaptées à cette vision et ainsi des métriques cohérentes avec le ressenti humain permettant d’exprimer la gêne sonore perçue.

Chaire “Mobilité et qualité de vie en milieu urbain”

Cette thèse a été effectuée dans le cadre de la Chaire d’excellence “Mobilité et qualité de vie en milieu urbain” (MoUVie) de la Fondation UPMC. Cette chaire bénéficie du mécénat des sociétés PSA Peugeot-Citroën et Renault. La titulaire de la chaire est M^{me} L. Eymard (OSU Ecce Terra). Ce projet de recherche vise à évaluer l’influence des modes de transports urbains sur la qualité de vie et la santé des citoyens en mettant les sources de pollutions chimiques (gaz et particules fines) et sonores au centre de la réflexion. Cette chaire d’excellence pluridisciplinaire associe des chercheurs appartenant à plusieurs laboratoires de l’UPMC.

Le volet acoustique de la chaire MoUVie a trois axes principaux : (i) apporter une meilleure caractérisation des sources sonores en ville (en particulier les sources mobiles). (ii) évaluer l’environnement sonore expérimenté individuellement grâce à l’utilisation de capteurs mobiles. (iii) améliorer les modèles de sonie multi-sources.

Ce projet de thèse s’insère dans le premier axe du projet de recherche du volet acoustique de MoUVie. Il est basé sur une approche pluridisciplinaire via la collaboration de deux Unités Mixtes de Recherche : l’Institut Jean le Rond d’Alembert (équipe Modélisation, Propagation et Imagerie Acoustique) et STMS (équipe Perception et Design Sonore).

Objectifs et contributions originales de la thèse

Nous l'avons vu par les actions prioritaires des législations et dans les enquêtes d'opinion, l'une des sources principales de bruit pour les citoyens est le trafic routier. L'Union Européenne a proposé de baser son action de lutte contre les nuisances sonores via l'utilisation du L_{DEN} . Cet indice a l'avantage d'être facile à calculer à partir du niveau équivalent pondéré A (L_{Aeq}). Cependant, pour qu'une législation soit bonne, elle doit s'attacher à réduire l'impact des sources de bruit les plus gênantes d'après Schafer[94]. Or, Berglund a montré [19] que seul 30% de la gêne sonore ressentie est explicable par le niveau sonore L_{Aeq} . Ce chiffre pourrait même être inférieur à 20% d'après Lercher [62]. Il est donc nécessaire de trouver de nouveaux indices qui peuvent expliquer une plus grande part de la variance de la gêne sonore et développer de nouveaux moyens pour caractériser l'exposition sonore à laquelle sont exposés les citoyens.

Cherchant à décrire finement les sources de bruit du trafic routier, Morel *et al.* [77] ont réalisé une typologie perceptive de bruit routier en supposant une dépendance à différents facteurs comme le type de véhicules, son allure et la géométrie de la rue. Alors que Polack *et al.* [86] montrent une influence de la morphologie de la rue (canyon urbain ou rue ouverte) sur le niveau sonore, l'étude de Morel *et al.*[77] montre qu'en égalisant les niveaux des 57 bruits de passage de véhicules utilisés ce facteur n'est plus discriminant. En effet, l'étude a montré, suite à une tâche de catégorisation libre, que le trafic routier pouvait se résumer en sept catégories perceptives distinctes qui dépendent uniquement du type de véhicules (deux-roues motorisés, véhicules légers et bus/véhicules lourds) et de l'allure (accélération, décélération et vitesse constante).

L'utilisation d'une telle typologie n'est pas directement possible par les pouvoirs publics car les outils de mesure de trafic disponibles (comme les tuyaux pneumatiques ou boucle électromagnétiques) ne peuvent distinguer que les poids lourds et les véhicules légers. Des projets comme METRA MOTO (MEsure du TRAfic des deux roues MOTOrisés - 2010–2013) ont cherché à proposer de nouveaux outils de mesure pour quantifier la part de deux-roues motorisés dans le trafic routier de façon automatique. Mais les données ne sont pas encore disponibles.

Il faut donc de nouveaux outils globaux, capables de détecter le type de véhicule, son allure et s'il est motorisé ou non. De plus, la contribution de chaque source à l'environnement sonore doit pouvoir être connue pour extraire le signal acoustique de chaque véhicule et ainsi estimer la gêne sonore associée de façon automatique.

Objectifs de la thèse. Dans cette thèse, nous proposons de concevoir un outil permettant d'estimer la gêne sonore associée à chaque véhicule du trafic routier via l'utilisation de son signal audio et de modèles de gêne sonore.

Pour ce faire, le signal audio du véhicule sera isolé de l'ensemble du trafic routier urbain grâce à l'utilisation de méthodes inverses et de grands réseaux de microphones ainsi que du traitement d'image pour obtenir sa trajectoire. Grâce à la connaissance de la trajectoire ainsi

que du signal, le véhicule sera classifié grâce à un algorithme de classification supervisée suivant la taxonomie de Morel *et al.* citée précédemment. Une fois sa catégorie obtenue, la gêne spécifique du véhicule sera estimée grâce à un modèle de gêne sonore utilisant des indices psychoacoustiques et énergétiques. Cette méthodologie est synthétisée schématiquement en figure 6.

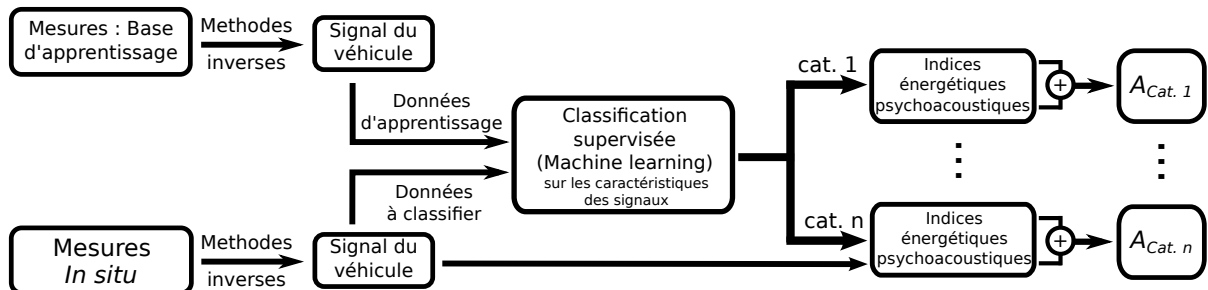


FIGURE 6 – Schéma de principe de la méthode d’estimation de la gêne sonore pour chaque véhicule (noté $A_{Cat. n}$ pour un véhicule appartenant à la catégorie n) présentée dans cette thèse.

Contributions originales de la thèse. Durant cette thèse nous associons différents concepts et méthodes pour répondre à cet objectif de la mesure de la gêne sonore dans le trafic routier urbain. Pour cela, plusieurs contributions originales sont proposées :

- Réalisation d’antennes de géométries adaptées aux problématiques de la mesure en milieu urbain nécessitant grand nombre de microphones,
- Localisation de sources de bruit sur des véhicules de différentes catégories, pour un grand nombre d’allures et un large spectre fréquentiel,
- Adaptation de la formation de voies fréquentielle pour l’extraction du signal audio d’un véhicule en mouvement dans le trafic routier urbain,
- Utilisation de méthodes de *machine learning* pour classier les véhicules du trafic routier suivant des catégories perceptivement homogènes à partir de leur signal audio,
- Implémentation et validation d’algorithmes d’estimation de rugosité et de force de fluctuation,
- Estimation du niveau sonore de véhicules en mouvement en fonction de leurs catégories perceptives,
- Estimation des gênes sonores spécifiques à chaque véhicule au sein du trafic routier au cours d’une journée grâce à l’ensemble de la méthode développée dans la thèse.

Plan de la thèse

Pour répondre à notre objectif ce manuscrit rapporte les différentes étapes de conception de l'outil d'estimation de la gêne sonore un à un avant de l'appliquer au milieu urbain. Le contenu des chapitres est détaillé ci-dessous.

Chapitre 1 Caractérisation de l'exposition sonore et modélisation de la gêne sonore. Les différents indicateurs utilisés pour quantifier l'exposition au bruit ainsi que les sensations sonores sont exposés. Des modèles disponibles dans la littérature pour relier le signal acoustique à la gêne sonore ressentie sont ensuite présentés. Un modèle est alors préféré, nécessitant la connaissance du signal sonore et de la catégorie perceptive du véhicule pour estimer sa gêne spécifique.

Chapitre 2 Identification et séparation de source à l'échelle du véhicule. Le développement d'antennes de microphones permettant de séparer les sources de bruit fixes ou mobiles est présenté. La capacité de filtrage spatial des antennes est analysée. Le signal audio de chaque véhicule peut ainsi être utilisé dans le modèle de gêne sonore. On présente enfin une analyse des sources de bruit à l'échelle du véhicule en fonction de son type, de sa motorisation et de son allure.

Chapitre 3 Classification typologique d'une scène sonore. Pour connaître la catégorie perceptive du véhicule, une méthode de classification automatique des éléments du trafic routier utilisant le signal audio du véhicule (extrait par antennerie acoustique) et son allure (obtenue par traitement d'image) est proposée. Les paramètres pertinents et les algorithmes utilisés y sont alors détaillés. Ces derniers permettent une estimation sans erreur de la catégorie perceptive.

Chapitre 4 Implémentation d'un modèle de gêne sonore : situation contrôlée. Le chapitre montre l'implémentation du modèle de gêne sonore sélectionné dans l'outil global. Le calcul d'indices psychoacoustiques pour l'alimenter est présenté et l'outil d'estimation de la gêne sonore est appliqué sur une scène sonore contrôlée : une piste d'essai.

Chapitre 5 Implémentation d'un modèle de gêne sonore : situation réelle. Vu les bons résultats sur piste d'essai, l'implantation du système dans l'environnement urbain de Paris est présentée. Elle fait appel à une antenne acoustique aisément déployable dans le contexte urbain. Une robustesse accrue est alors recherchée pour la classification du trafic routier et donc l'estimation des gênes spécifiques à chaque véhicule. Ces dernières sont estimées pendant une journée et les résultats sont analysés.

À l'issue de ce travail un outil de mesure de la gêne sonore spécifique du trafic routier en ville aura été mis au point. Il reste cependant perfectible et nous tracerons quelques perspectives pour mener à bien les améliorations requises.

Chapitre 1

Modélisation de la gêne sonore

Sommaire

Introduction	13
1.1 Indices énergétiques et psychoacoustiques	14
1.1.1 Le niveau équivalent	14
1.1.2 Pondération fréquentielle	15
1.1.3 Pondération temporelle	17
1.1.4 Niveaux statistiques	18
1.1.5 Indices événementiels	18
1.1.6 La sonie	20
1.1.7 L'acuité	21
1.1.8 La force de fluctuation	22
1.1.9 La rugosité	23
1.2 Gêne sonore et modélisation	24
1.2.1 Facteurs influençant l'estimation de la gêne sonore	25
1.2.2 Modèles de gêne	27
Conclusion	33

Introduction

La question de l'estimation de la gêne sonore se pose depuis que l'on cherche à quantifier l'exposition sonore. Elle a eu de nombreuses réponses qui seront développées dans ce chapitre, mais la définition même de la gêne sonore a longtemps fait débat.

“Noise annoyance is a psychological concept which describes a relation between an acoustic situation and a person who is forced to do things he/she does not want to do, who cognitively and emotionally evaluates this situation and feels partly helpless”¹

1. La gêne sonore est un concept psychologique qui décrit une relation entre une situation acoustique et une personne qui est forcée à faire des choses qu'il/elle ne veut pas faire, qui cognitivement et émotionnellement évalue cette situation et se sent en partie désemparée.

Après plusieurs décennies de flou et de concepts différents (désagrément, gêne, etc.), c'est finalement en ces termes que Guski *et al.*[51] définissent en 1999 le concept de gêne sonore en recueillant l'avis d'experts internationaux. Ils mettent en avant le fait que le son est ressenti comme une gêne à partir du moment où il nous perturbe dans notre action en cours. Un exemple familier : l'impossibilité de lire en présence de bruit persistant ou saillant.

R.M. Schafer [94] affirme que, pour que nos législations soient efficaces, elles doivent cibler la réduction des sons les plus déplaisants d'un lieu donné. L'estimation de la gêne sonore en fonction du type de sources et du lieu doit donc être un point important dans la définition de politique publique de réduction des nuisances sonores. Mais la gêne sonore telle que définie par Guski *et al.* [51] n'est pas une métrique mais un concept. En revanche, la directive européenne 2002/49/CE relative à l'évaluation et à la gestion du bruit dans l'environnement [2] introduit une notion de quantification en définissant la gêne comme étant "le degré de nuisance généré par le bruit dans l'environnement, déterminé par des enquêtes sur le terrain."

Dans ce chapitre nous exposerons dans un premier temps les indices énergétiques utilisés pour quantifier l'environnement sonore, l'exposition au bruit. Les indices psychoacoustiques introduits par Zwicker seront ensuite présentés. Ils permettent de modéliser des sensations sonores comme la sonie ou la rugosité. Ensuite, nous présenterons des modèles de gêne sonore disponibles dans la littérature, utilisant les indices cités précédemment.

1.1 Indices énergétiques et psychoacoustiques liés à l'exposition sonore

Depuis près d'un siècle, le décibel (dB) est utilisé pour quantifier les événements sonores. Il a été initialement défini par les laboratoires Bell pour quantifier l'affaiblissement du signal lors de sa transmission, d'abord sous le nom de *Transmission Unit*, puis sous le nom de décibel (en 1923 ou 1924) à la suite de la mort de Alexander Graham Bell. Il est défini comme :

$$X_{dB} = 10 \log \left(\frac{X}{X_0} \right), \quad (1.1)$$

où X est une intensité ou puissance mesurée et X_0 une quantité de référence.

1.1.1 Le niveau équivalent

L'application à l'acoustique a été faite en posant X comme l'intensité du champ acoustique I et l'intensité de référence X_0 définit le seuil d'audition d'un son mono-fréquentiel à 1000 Hz, soit $I_{ref} = 10^{-12} \text{ W.m}^{-2}$. En effet, le décibel respecte alors la loi psycho-physique de Weber-Fechner qui décrit la relation entre un stimulus sensoriel perçu et l'excitation suivant une loi logarithmique. Le niveau de l'intensité acoustique est alors donné par :

$$L = 10 \log \left(\frac{I}{I_{ref}} \right). \quad (1.2)$$

Le niveau équivalent $L_{eq,T}$, aussi appelé niveau SPL (pour *Sound Pressure Level* en anglais), est plus fréquemment défini à partir de la pression acoustique efficace (notée $p_{eff,T}$) et une référence notée p_{ref} de $20 \mu\text{Pa}$ qui correspond au seuil d'audition à 1 kHz. On définit alors le niveau équivalent sur une période d'intégration T comme :

$$L_{eq,T} = 10 \log \left(\frac{p_{eff,T}^2}{p_{ref}^2} \right) = 20 \log \left(\frac{p_{eff,T}}{p_{ref}} \right) \quad \text{avec} \quad p_{eff,T}^2 = \frac{1}{T} \int_0^T p(t)^2 dt. \quad (1.3)$$

Le niveau équivalent L_{eq} , exprimé en dB, est un outil de quantification de la sensation sonore très répandu car facile à calculer et prenant en compte la dynamique de l'oreille humaine. Cette quantification physique, ou énergétique, a apporté une première réponse mais ne suffit pas à représenter le système auditif dans toute sa complexité. C'est pourquoi on a tenté de l'améliorer depuis le milieu du XX^e siècle.

1.1.2 Pondération fréquentielle

Les travaux de Fletcher & Munson, dans les années 1930, ont montré que l'oreille humaine ne perçoit pas avec la même intensité des sons de même niveau sonore à des fréquences différentes. A la suite de ces études, ils ont établi les courbes d'isophonie (de même sensation sonore) visibles sur la figure 1.1. Comme on peut le voir, on mesure la sonie dans une unité spécifique, le phone, dont la référence est la suivante : 40 phones est le niveau sonore équivalent perçu d'un signal de 40 dB à 1000 Hz. Pour avoir la même sensation sonore de 40 phones à 100 Hz, il faut que le signal soit à 65 dB. A l'inverse, à 3000 Hz il ne doit être que de 35 dB, notre oreille étant alors plus sensible.

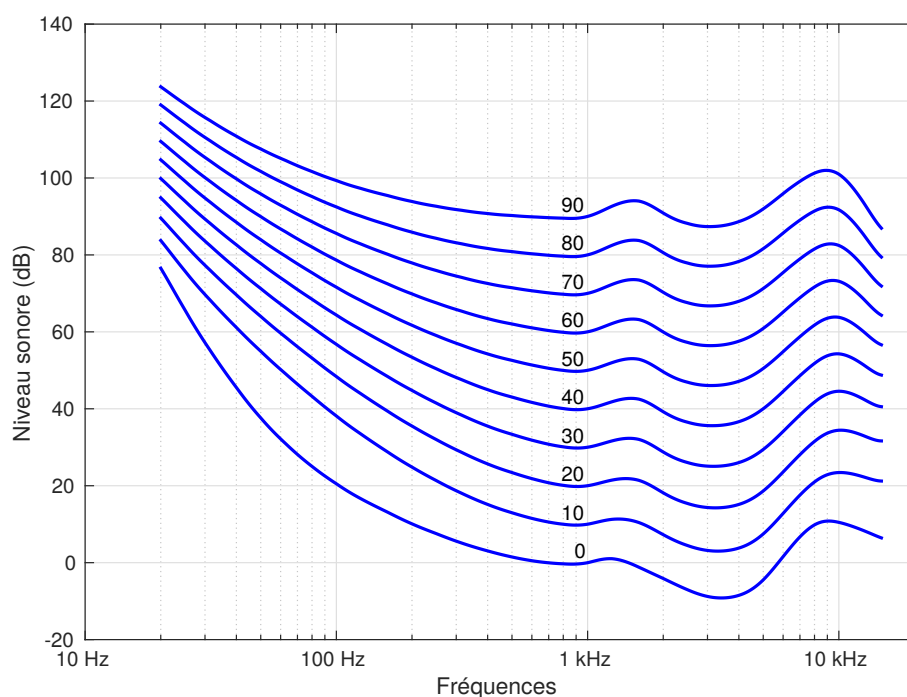


FIGURE 1.1 – Courbes d'isophonie de Fletcher - Niveau sonore perçu équivalent exprimé en phones en fonction du niveau acoustique et de la fréquence du son - ISO 226 [55].

Niveaux pondérés en fréquence

Les courbes présentées en figure 1.1 ont servi de support à la formalisation de différentes pondérations fréquentielles définies en fonction du niveau de la source. L'oreille est alors considérée comme un filtre dont la pondération varie en fonction de la fréquence et du niveau. L'allure des courbes de ces filtres de pondération est sensée reproduire, sous forme inversée, les courbes d'isotonie. Ces courbes de pondération sont représentées sur la figure 1.2.

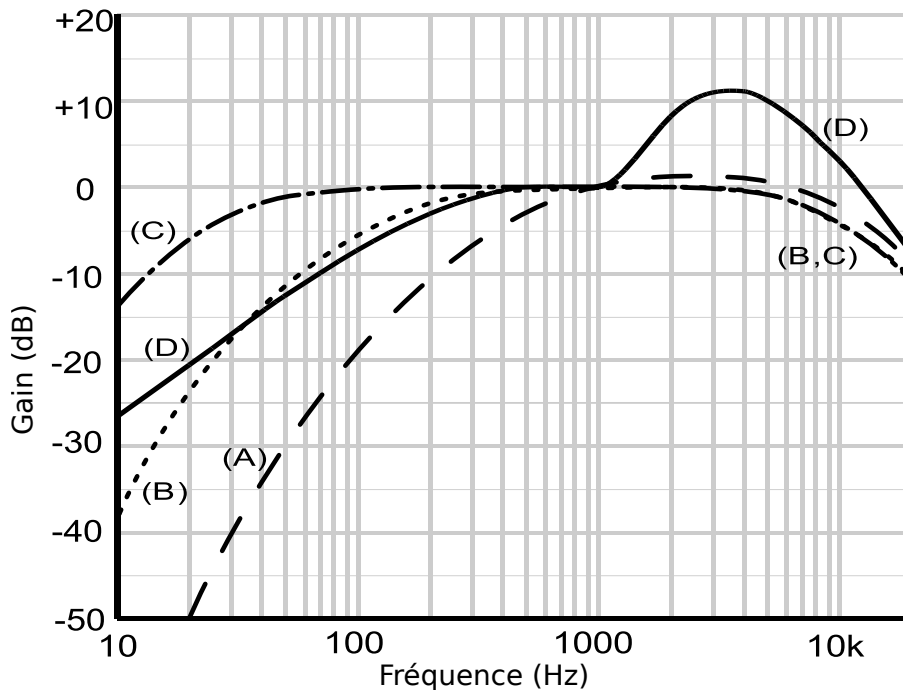


FIGURE 1.2 – Pondérations fréquentielles A (— —), B (- · -), C (— · —) et D (en trait plein) - D'après les normes IEC 60537 (pondération D) et IEC 60651 (pondérations A, B, C).

La courbe de pondération A correspond à l'allure des courbes d'isotonie entre 20 et 40 phones environ, la courbe de pondération B à celles entre 50 et 70 phones et la courbe de pondération C à celles entre 80 et 90 phones. Dans bien des normes ou préconisations, les seuils sont indiqués en dB(A), même pour des sons de sonie supérieure à 40 phones. Une courbe de pondération D a été introduite pour les nuisances provoquées par les avions (donc pour des niveaux supérieurs à 100 dB). Elle prend en compte le gain en sensibilité présent sur toutes des courbes d'isotonie (figure 1.1) autour de 3 500 Hz.

Niveau sonore perçu

Noté L_{PN} , pour *Perceived Noise Level*, le niveau sonore perçu a été développé dans les années 1960 par la communauté aéronautique. Il a été élaboré dans l'objectif de lier le caractère bruyant, la gêne sonore et les caractéristiques du bruit d'avion. Il s'exprime en PNdB et est calculé comme suit :

$$L_{PN} = 40 + 33,3 \log \mathbb{N}, \quad (1.4)$$

avec \mathbb{N} , exprimé en noys, donné par :

$$\mathbb{N} = n_{max} + 0,15 \left[\sum_{i=1}^{24} n(i) - n_{max} \right]. \quad (1.5)$$

Les $n(i)$ sont alors les nombres de noys pour chacune des 24 bandes de tiers d'octave comprises entre 50 et 10 000 Hz et n_{max} la valeur maximale de noys sur les 24 bandes. Pour chaque bande d'octave ou de tiers d'octave, le nombre de noys est calculé à partir du niveau de pression grâce aux courbes présentées en figure 1.3.

Ce niveau sonore perçu L_{PN} est classiquement évalué toutes les demi-secondes. Bien qu'intéressant, on peut voir que cet indice est assez fastidieux à calculer.

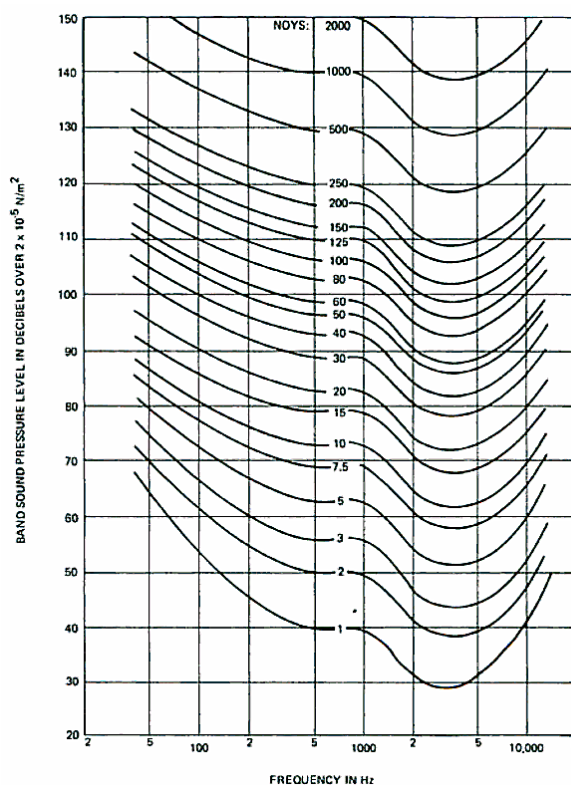


FIGURE 1.3 – Isocontours de noys en fonction du niveau en décibels et de la fréquence en Hertz, d'après [85].

1.1.3 Pondération temporelle

Dérivé du L_{dn} (*Day, Night*) utilisé par l'autorité de régulation de l'aviation américaine depuis les années 1970 pour prendre en compte la gêne accrue la nuit, le parlement européen a fixé l'utilisation du L_{den} (*Day, Evening, Night*) comme outil de mesure de l'exposition au bruit dans les villes avec la directive 2002/CE/49 [2]. Il est calculé ainsi :

$$L_{den} = 10 \log \left[\frac{12}{24} \cdot 10^{L_d/10} + \frac{4}{24} \cdot 10^{(L_e+5)/10} + \frac{8}{24} \cdot 10^{(L_n+10)/10} \right]. \quad (1.6)$$

Il s'agit donc de prendre en compte la gêne accrue due au bruit en soirée et la nuit. Les niveaux de jour (L_d , de 6h à 18h), de soirée (L_e , de 18h à 22h) et de nuit (L_n , de 22h à 6h)

sont des niveaux équivalents (cf. 1.1.1) pondérés A (cf. 1.1.2) et majorés de +5 dB pour L_e et +10 dB pour L_n .

1.1.4 Niveaux statistiques

Ces niveaux sont issus de l'analyse du niveau équivalent évalué toutes les secondes (*slow*) ou 125 ms (*fast*). On peut définir différents niveaux correspondant au pourcentage de temps dépassé. Par exemple :

- L_{95} : niveau dépassé pendant 95% du temps, assimilé au bruit de fond,
- L_{50} : niveau dépassé pendant 50% du temps, assimilé au bruit médian,
- L_5 : niveau dépassé pendant 5% du temps, assimilé aux émergences.

Un exemple de calcul du L_5 et L_{95} est présenté en figure 1.4 pour dix minutes de mesure dans une grande artère parisienne, proche de feux de signalisation. Ces indices statistiques servent de base pour la construction d'indices de gêne sonore qui seront détaillés en section 1.2.2.

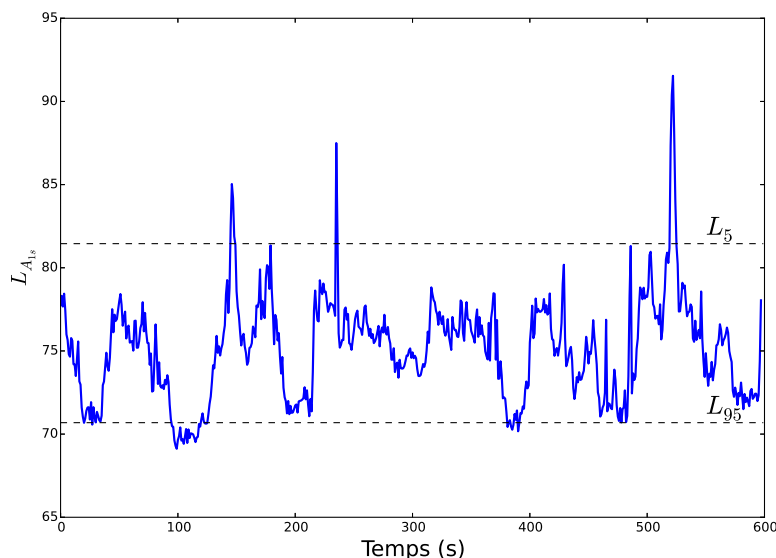


FIGURE 1.4 – Niveau acoustique calculé toutes les secondes pondéré A ($L_{A,1s}$) et niveaux L_{95} et L_5 - Enregistrement de 10 minutes dans une grande artère parisienne.

1.1.5 Indices événementiels

Indice de masquage

Cet indice, noté $MI_{\alpha_{max}}$, pour *Mask Index*, est un rapport de temps (il est donc sans dimension). L'indice de masquage peut être appliqué à différents indices, α peut alors être le niveau pondéré A ($L_{A,eq}$), le niveau dépassé pendant 1 % du temps (L_1), etc. Il est donné par :

$$MI_{\alpha_{max}} = \frac{T_{\alpha > \alpha_{max}}}{T_0} \quad (1.7)$$

où $T_{\alpha > \alpha_{max}}$ est la période pendant laquelle la valeur seuil (notée max) de l'indice α est dépassée et T_0 la période d'enregistrement. L'indice de masquage aide ainsi à décrire l'évolution temporelle d'un événement sonore via un indice comme le $L_{A,1s}$.

Niveau de pollution sonore

L'idée de cet indice conçu en 1971 par Robinson [91] est de prendre en compte la gêne sonore accrue lorsque le niveau de bruit fluctue beaucoup. Ce niveau, noté L_{NP} pour *Noise Pollution Level*, est donné par

$$L_{NP} = L_{A,eq} + K\sigma, \quad (1.8)$$

où σ est l'écart type du niveau instantané pondéré A ($L_A(t)$) et K une constante évaluée à 2,56 après des études britanniques sur l'évaluation du bruit de trafic routier et aérien [50]. Robinson indique que ce facteur se trouve être quasiment égal à $(L_{10} - L_{90})$ lorsque la distribution du niveau instantané est gaussienne, comme c'est souvent le cas pour le bruit urbain.

Niveau d'exposition sonore

Le niveau d'exposition sonore (noté SEL, pour *Sound Exposure Level*) consiste à définir un niveau sonore équivalent ramené à une durée d'une seconde pour un événement sonore de durée quelconque. Il est exprimé en décibels pondéré A et est donné par :

$$SEL = 10 \log \frac{1}{t_0} \int_{t_1}^{t_2} 10^{L_A(t)/10} dt, \quad (1.9)$$

où t_0 est le temps de référence, égale à 1 s, et t_1 et t_2 les temps de début et de fin de l'événement sonore. En pratique, d'après [70], les bornes d'intégration correspondent au temps pendant lequel le niveau dépasse un seuil, défini à -10 dB du niveau maximum. Les événements sonores étant souvent plus longs qu'une seconde, le SEL est plus élevé que le niveau équivalent. Cet indice est recommandé pour la prise en compte des bruits aériens.

Indice du bruit de trafic

S'inscrivant dans la caractérisation de l'environnement sonore via la prise en compte des événements sonores saillants, Griffiths et Langdon [50] ont développé l'indice de bruit de trafic, noté TNI pour *Traffic Noise Index*, qu'ils définissent à partir de niveaux statistiques pondérés A :

$$TNI = 4(L_{10} - L_{90}) + L_{90} - 30 \quad (1.10)$$

Griffiths et Langdon présentent des corrélations assez bonnes entre les valeurs de TNI et d'insatisfactions sonores perçues (coefficient de corrélation : $r = 0,81$). Cet indice a été utilisé dans plusieurs études, notamment en Grande-Bretagne [50, 96, 109].

Les indices présentés précédemment traduisent dans des métriques physiques des propriétés de l'environnement sonore et, dans le cas de sources isolées, des sources de bruit. Malgré la prise en compte d'études psychoacoustiques dans la construction de certains indices, Genuit et Fiebig [44] estiment que l'événement sonore peut être décrit de manière plus détaillée par des indices psychoacoustiques tels que la sonie, l'acuité, la rugosité et la

force de fluctuation. Ils insistent sur la capacité de ces indices à caractériser chaque source du paysage sonore.

1.1.6 La sonie

La sonie est l'unité de mesure de l'intensité sonore perçue, différente de l'intensité physique mesurée par exemple par un sonomètre. On l'exprime en sone (1 sone correspond à l'intensité perçue d'une tonalité pure à 1000Hz et 40 dB). La sonie, notée N , dépend évidemment du niveau de pression sonore, mais également d'autres facteurs, comme le contenu fréquentiel ou la durée [38].

On utilise également le niveau d'isononie présenté précédemment (et utilisé dans la figure 1.1), noté L_N et exprimé en phones. Il est obtenu à partir de la sonie (exprimée en sonnes) par la formule suivante :

$$L_N = 10 \log_2 N + 40 \quad (1.11)$$

Certains modèles (dont le modèle de Fastl et Zwicker [38] utilisé dans la norme ISO 532B) permettant de calculer la sonie d'un bruit, proposent de découper l'échelle des fréquences en 24 bandes critiques exprimées en Bark. Un Bark correspond à une bande de fréquence appelée bande critique dimensionnée sur des bases physiologiques et fonctionnelles de l'appareil auditif (essentiellement l'oreille interne).

z	f_l	f_u	f_c	Δf	z	f_l	f_u	f_c	Δf
Bark	Hz	Hz	Hz	Hz	Bark	Hz	Hz	Hz	Hz
0	0	100	50	100	12	1720	2000	1850	280
1	100	200	150	100	13	2000	2320	2150	320
2	200	300	250	100	14	2320	2700	2500	380
3	300	400	350	100	15	2700	3150	2900	450
4	400	510	450	110	16	3150	3700	3400	550
5	510	630	570	120	17	3700	4400	4000	700
6	630	770	700	140	18	4400	5300	4800	900
7	770	920	840	150	19	5300	6400	5800	1100
8	920	1080	1000	160	20	6400	7700	7000	1300
9	1080	1270	1170	190	21	7700	9500	8500	1800
10	1270	1480	1370	210	22	9500	12000	10500	2500
11	1480	1720	1600	240	23	12000	15500	13500	3500

TABLE 1.1 – Bandes critiques avec z le numéro de la bande en Bark, f_l et f_u respectivement les fréquences basse et haute de la bande, f_c la fréquence centrale et Δf la largeur de la bande critique. La limite haute de modèle est 24 Bark = 15500 Hz - D'après [38].

La sonie dans chaque bande critique est appelée sonie spécifique. On la note N' et l'exprime en sone/Bark. La sonie totale est obtenue en intégrant les sonies spécifiques sur les 24 bandes critiques :

$$N = \int_0^{24\text{Bark}} N' dz. \quad (1.12)$$

Les étapes du modèle de sonie le plus courant, proposé par Fastl et Zwicker [38], sont décrites dans la figure 1.5. Elles tendent à modéliser les différentes parties de l'oreille via

un banc de filtres centrés sur l'échelle de Bark. On peut voir que le modèle se base sur un filtrage du spectre du signal sur les bandes critiques. Ce spectre par bande est filtré et un seuil de détection est appliqué. Puis, pour chaque bande critique, un masque est appliqué pour calculer les sonies spécifiques qui sont ensuite intégrées pour estimer la sonie globale.

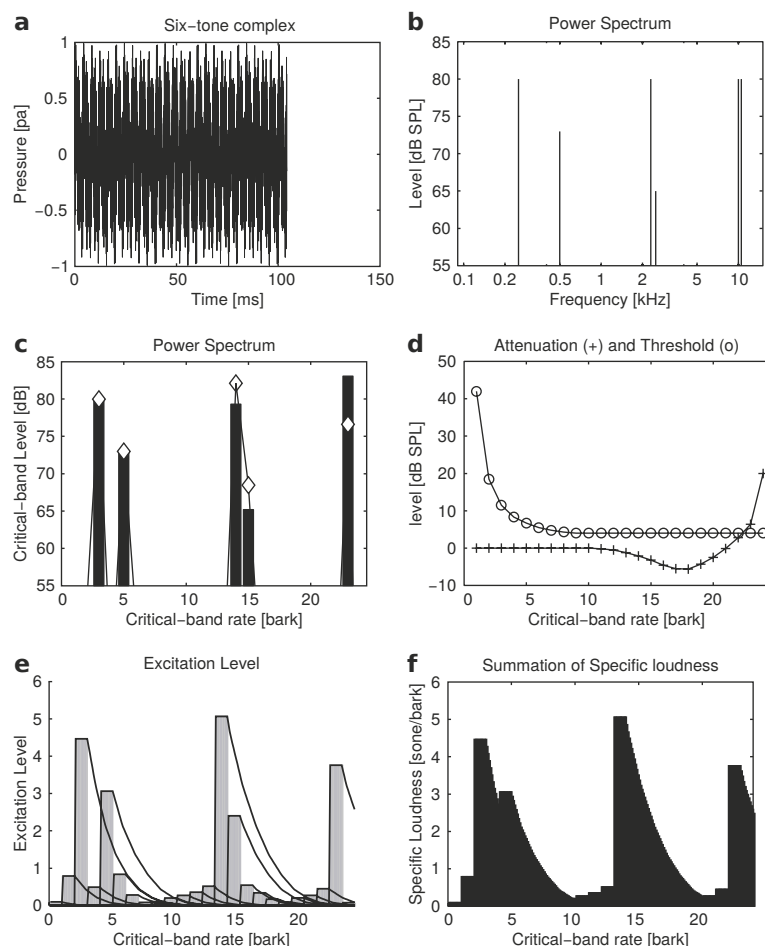


FIGURE 1.5 – Étapes du modèle de Zwicker : (a) signal temporel, (b) spectre en bande fine, (c) spectre filtré par les bandes critiques, (d) atténuation par deux filtres linéaires, (e) niveau d'excitation et sonie spécifique et (f) intégration des sonies spécifiques - D'après [41].

1.1.7 L'acuité

L'acuité est une mesure de la densité spectrale de sonie d'un son [38]. En ce sens elle représente une mesure de l'équilibre entre basses et hautes fréquences. On la note S et on l'exprime en acum. Fastl et Zwicker proposent de calculer l'acuité à partir des sonies spécifiques N' selon le modèle suivant [38] :

$$S = 0,11 \frac{\int_0^{24 \text{Bark}} N' g(z) z dz}{\int_0^{24 \text{Bark}} N' dz} \quad (1.13)$$

où $g(z)$ est un facteur de pondération qui est fonction des bandes critiques [38]. Il est unitaire de la première à la 17^{ème} bande critique puis augmente de 1 à 4 de la 17^{ème} à la 24^{ème}

bande critique. On peut voir l'acuité comme un centre de gravité spectral "perceptif". En effet, le centre de gravité spectral (CGS) est une mesure physique de la balance spectrale d'un son. On le définit alors comme la moyenne des fréquences f_i du spectre pondérée par leur amplitude A_i :

$$CGS = \frac{\sum_i A_i f_i}{\sum_i A_i} \quad (1.14)$$

1.1.8 La force de fluctuation

D'après Fastl et Zwicker [38], des bruits non stationnaires modulés en amplitude, peuvent faire apparaître deux sensations différentes. Pour des fréquences de modulation en amplitude jusqu'à 20 Hz, la sensation ressentie est appelée force de fluctuation. La force de fluctuation, notée F , est exprimée en vacil.

La figure 1.6 présente les résultats des expériences de Fastl et Zwicker. Ils mettent en évidence que cette sensation de force de fluctuation est maximum à 4 Hz pour différents types de bruit, c'est donc un phénomène très basse fréquence.

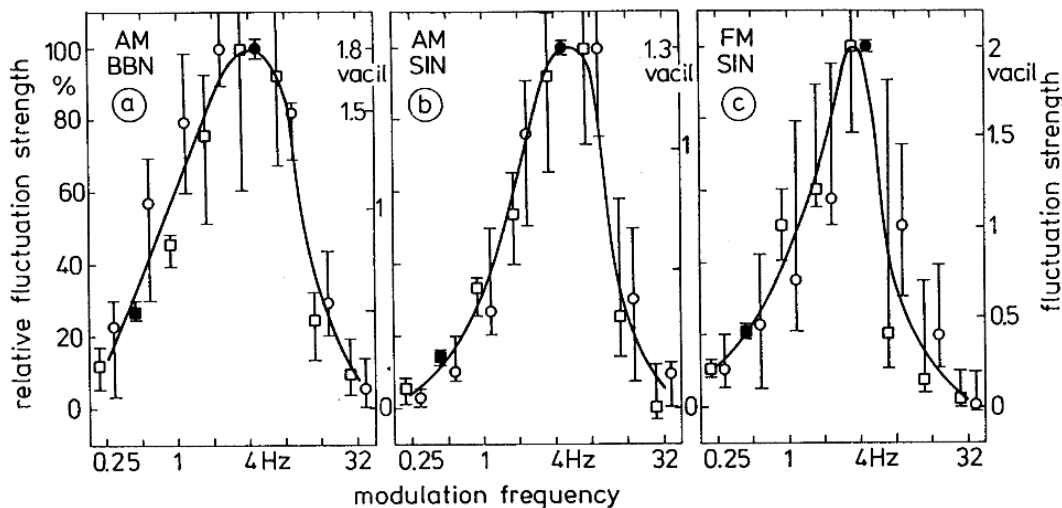


FIGURE 1.6 – Force de fluctuation estimée en fonction de la fréquence de modulation pour trois sons différents. (a) bruit large bande à 60 dB modulé en amplitude de 40 dB - (b) sinus à 1 kHz à 70 dB modulé en amplitude de 40 dB - (c) sinus modulé en fréquence, de fréquence centrale à 1,5 kHz et de 700 Hz d'écart-type à 70 dB - D'après [38].

Fastl et Zwicker proposent un modèle théorique de force de fluctuation avec la relation suivante :

$$F \approx \frac{\Delta L}{\frac{f_{mod}}{4} + \frac{4}{f_{mod}}} \quad (1.15)$$

où f_{mod} est la fréquence de modulation du son et ΔL est l'amplitude de modulation du niveau sonore perçu. L'explication des paramètres est donnée dans la figure 1.7 représentant le logarithme d'un sinus modulé sinusoidalement. Le modèle se base sur la recherche d'une sinusoïde qui aurait la même fréquence de modulation, l'amplitude de modulation ΔL en découle alors et elle est appelée profondeur de masquage temporel perçu.

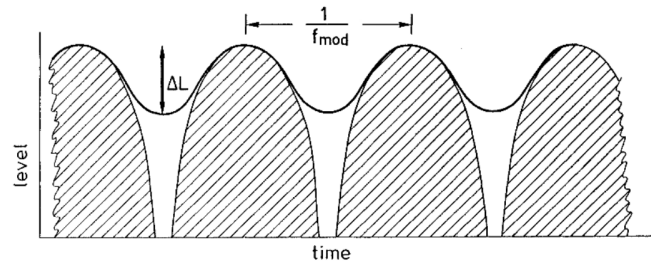


FIGURE 1.7 – Modèle de force de fluctuation : motif de masquage temporel d'un signal modulé en amplitude de façon sinusoidale conduisant à la profondeur de masquage temporel perçue ΔL - D'après [38].

1.1.9 La rugosité

La rugosité est une sensation différente du même phénomène de modulation d'un signal. On la ressent à plus hautes fréquences : entre 20 et 300 Hz. Elle est notée R , exprimée en asper et elle est normalisée à 1 asper pour un sinus à 1 kHz de 60 dB modulé en amplitude à 100% (indice de modulation $m=1$) par un sinus à 70 Hz. La figure 1.8 présente l'évolution de la rugosité en fonction de la fréquence centrale du signal et de la fréquence de modulation. On peut voir qu'en fonction de la fréquence centrale, le maximum de rugosité est atteint pour des fréquences de modulation comprises entre 40 et 80 Hz.

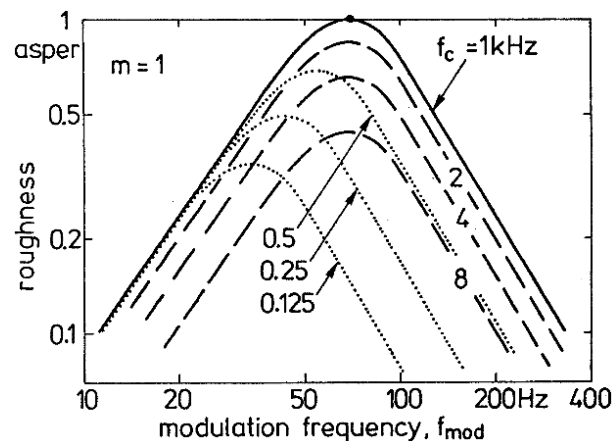


FIGURE 1.8 – Rugosité induite par l'écoute de sinus à 60 dB modulé en amplitude à 100 % ($m=1$) pour différentes fréquences centrales et de modulation.

Un modèle approximatif est proposé par Zwicker et Fastl[38] :

$$R \approx \Delta L f_{mod}. \quad (1.16)$$

On peut voir que la force de fluctuation et la rugosité sont des indices qui dépendent de la profondeur de masquage temporel perçue ΔL . Ce point est la partie délicate des modèles qui sont présentés en section 4.1.

Alors que les indices énergétiques présentés en section 1.1 cherchent à quantifier des paramètres physiques du son, les indices psychoacoustiques cherchent à modéliser des sensations perçues par ces sons. Ils nous permettent donc de caractériser une source de bruit avec

des métriques plus proches du ressenti humain. Ces indices psychoacoustiques font partie des indicateurs utilisés pour modéliser la gêne sonore associée à une source de bruit (section 1.2.2).

1.2 Gêne sonore et modélisation

Les premières études psychoacoustiques sur la gêne sonore ont cherché à définir le lien entre une quantité mesurable, le niveau sonore équivalent, et le pourcentage de la population gênée ou très gênée [98, 6, 39, 71, 56, 48], en introduisant la notion de relation dose-effet. La figure 1.9 présente un exemple issu de l'étude de Lercher *et al.* [63]. Cette notion a aussi été étendue à l'étude des risques sur la santé par rapport à l'exposition sonore [15, 42, 16].

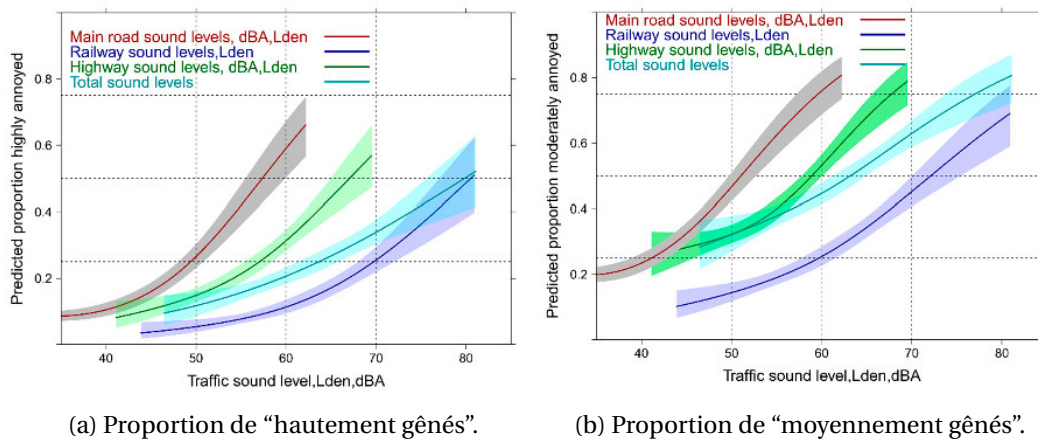


FIGURE 1.9 – Estimation de la proportion de gêne élevée et modérée associée au niveau sonore de chaque type de source ou de la somme des sources.

Une autre approche consiste à définir de nouvelles métriques pour estimer la gêne sonore ressentie. Certaines études ont cherché à faire le lien entre des indices énergétiques ($L_{A,eq}$, L_{10} , etc.) et la gêne exprimée. Lors d'études psychoacoustiques cherchant à quantifier la gêne sonore, les sujets doivent, par exemple, juger l'environnement (ou la scène) sonore auquel ils sont exposés sur une échelle de 0 ("pas gênant du tout") à 9 ("extrêmement gênant") [116, 117], ou sur une échelle relative par rapport à un son de référence [21, 20]. Cette évaluation est réalisée en demandant aux sujets de s'imaginer chez eux (fenêtre fermée, ouverte ou dans le jardin en fonction des études) et en train d'effectuer une activité (lire, jardiner, regarder la télévision, etc.). Lors d'études plus récentes qui utilisent un protocole expérimental similaire [78, 59], des modèles de gêne sonore sont proposés à partir d'indices énergétiques et psychoacoustiques comme la sonie ou la rugosité directement calculés sur le signal acoustique.

Gêne spécifique ou partielle. Les études citées précédemment cherchent souvent à estimer la gêne spécifique. Ce terme désigne la gêne associée à la source de bruit écoutée seule. La gêne partielle, assez peu utilisée dans la littérature, a été introduite par Berglund et Nils-

son [22]. Elle représente la part de gêne due à une source dans une scène sonore complexe composées de plusieurs sources.

Gêne totale. La gêne totale est définie comme la gêne ressentie à l'exposition à une scène sonore complexe composée de plusieurs sources. De nombreuses études, comme par exemple [104, 79, 116, 87, 20, 78], portent sur le lien entre gêne spécifique et gêne totale. D'autres études étudient le lien entre les gênes partielles et la gêne totale (comme [25, 68]). Mais aucune ne s'accorde sur un modèle qui serait utilisable avec n'importe quel type de source ou de situation. Les principaux modèles sont rappelés dans le tableau 1.2.

Modèle de gêne totale	Variables indépendantes	Exemple
Sommation de l'énergie	L_{tot} ou N_{tot}	$A_{tot} = \alpha L_{tot} + \beta$
Effets indépendants	A_i ou L_i	$A_{tot} = \alpha A_1 + \beta A_2 + \gamma$
Différence d'énergie	$L_{tot}, L_1 - L_2 $	$A_{tot} = L_{tot} + \alpha L_1 - L_2 + \gamma$
Sommation des réponses [79]	L_i	$A_t = \alpha [L_{tot} + \sum_{i=1}^n D_i 10^{(L_i - L_{tot})/10}] + \beta$
Sources dominantes	A_i ou L_i	$A_{tot} = \max(A_1, A_2)$
Sommation pondérée [116]	L_i, k	$A_{tot} = \alpha L_t + \beta, L_t = k \log(\sum_i 10^{L_i/k})$
Sommation et inhibition [87]	L_i, A_i	$A_{tot} = \alpha(L_{tot} + E) + \beta$
Sommation vectorielle [20]	A_i	$A_{tot} = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\alpha_{1,2}}$
Modèle mixte [76]	L_i, A_i	$A_{tot} = \alpha L_1 + \beta L_2 + \gamma L_1 - L_2 + \delta$

TABLE 1.2 – Principaux modèles de gêne totale selon Marquis-Favre (2005) [69] - A_{tot} désigne la gêne totale, A_i la gêne due à la $i^{ème}$ source, c_{ij} le facteur correctif prenant en compte la similarité ou non des sources i et j , L_i (ou N_i) le niveau (ou la sonie) de la $i^{ème}$ source, L_{tot} (ou N_{tot}) le niveau (ou la sonie) total et k est un coefficient de pondération.

1.2.1 Facteurs influençant l'estimation de la gêne sonore

Marquis-Favre *et al.* [69] proposent plusieurs facteurs pouvant influencer la sensation de gêne. Ils indiquent par exemple que l'évaluation à long-terme de la gêne sonore peut être surestimée par rapport à la moyenne des gênes sonores évaluées à court-terme.

Les auteurs listent aussi les facteurs acoustiques et non-acoustiques reconnus pouvant influencer l'évaluation de la gêne sonore *in situ* ou en laboratoire. Ils précisent que les différents facteurs prennent des proportions différentes suivant la méthode d'étude. Les effets acoustiques listés dans le tableau 1.3 sont des effets qui sont quantifiables grâce au signal audio ou à la vision de la scène. Parmi ceux-ci, on peut citer par exemple la fluctuation en amplitude, le caractère tonal ou impulsionnel ou encore la périodicité de la présence de la source.

Les effets non-acoustiques comme l'activité effectuée pendant l'exposition à la source de bruit ou la perception du voisinage, présentés en tableau 1.4, peuvent varier de façon importante pour chaque individu. En effet, ils traduisent notre histoire, notre relation passée et vécue, avec la source du bruit. Si le sujet a une bonne entente avec son voisinage, la gêne exprimée pour un son de type "bruit de voisinage" sera inférieure à celle exprimée par un sujet en conflit permanent.

Facteurs liés à la source de bruit	<ul style="list-style-type: none"> - Exposition au bruit - Large fluctuation en amplitude - Nombre d'événements - Émergence d'événements ponctuels - Caractère impulsionnel - Caractère tonal - Contenu en basses fréquences
Facteurs liés au contexte dans lequel apparaît le bruit	<ul style="list-style-type: none"> - Bruit ambiant - Changement d'exposition au bruit - Introduction durable d'une nouvelle source de bruit
Facteurs liés à la nature de la source de bruit	<p>Effet du mode de transport : pour un même niveau sonore la gêne perçue est différente entre le trafic aérien, routier et ferroviaire</p>
Facteurs liés à l'aspect qualitatif des bruits	<ul style="list-style-type: none"> - Sonie - Rugosité - Force de fluctuation - Périodicité

TABLE 1.3 – Exemples de facteurs acoustiques pouvant influencer les jugements de gêne sonore - d'après Morel [76].

Facteurs liés à l'attitude	<ul style="list-style-type: none"> - Crainte de la source de bruit - Possibilité d'adaptation par rapport à la source de bruit - Sensibilité au bruit - Activité effectuée durant l'exposition au bruit - Perception du voisinage - Perception globale de l'environnement
Facteurs socio-démographiques	<ul style="list-style-type: none"> - Age - Statut marital - Niveau d'éducation - Catégorie socio-professionnelle
Facteurs situationnels	<ul style="list-style-type: none"> - Temps passé au domicile - Isolation acoustique du domicile - Être propriétaire ou locataire du domicile

TABLE 1.4 – Exemples de facteurs non-acoustiques pouvant influencer les jugements de gêne sonore - d'après Morel [76] et Flindell *et al.* [40].

Certaines études, comme celle de Miedema et Vos [72], cherchent à utiliser les facteurs acoustiques et non-acoustiques dans l'estimation de la gêne sonore. Dans cet exemple, la gêne est modélisée comme une combinaison du niveau sonore et la sensibilité au bruit. Ce dernier facteur étant estimé via des informations comme l'âge et le genre.

De nombreuses autres études cherchant à estimer la gêne sonore se focalisent uniquement sur l'utilisation des facteurs acoustiques, à savoir les caractéristiques mesurables et objectives du son. Elles cherchent donc à estimer la part de la gêne sonore commune à tous et non les variations inter-individus.

1.2.2 Modèles de gêne

Dans cette section nous présenterons des modèles disponibles dans la littérature proposant de nouvelles métriques pour estimer la gêne sonore. Ces modèles utilisent le signal audio de la source sonore et cherchent à le relier à la gêne exprimée par des riverains ou via des tests d'écoute en laboratoire. L'objectif de chacun de ces indices est d'avoir la meilleure corrélation entre les données physiques et perceptives.

Harmonica

L'indice Harmonica est issu du projet européen du même nom qui s'est terminé en décembre 2014 dont les porteurs français étaient BruitParif et Acoucité [73]. Le but était le développement d'un outil d'information plus facile à appréhender que les cartes stratégiques du bruit et plus proche de la réalité des nuisances sonores telles qu'elles peuvent être ressenties par les riverains.

Ce nouvel indice propose de se baser sur des niveaux statistiques (pondérés A), notamment parce qu'ils sont disponibles dans la majeure partie des stations fixes des observatoires de bruit urbain. Il est constitué de la somme de deux sous-indices représentant la part de bruit de fond (BGN) et la part d'émergence (EVT), et dont la somme est comprise entre 0 et 10. Tel que défini dans l'équation (1.17), l'indice Harmonica est celui jugé le plus pertinent par les sujets d'une étude en laboratoire parmi quatre indices proposés dans une étude. Les deux sous-indices sont calculés comme suit :

$$\text{BGN} = 0,2 \times (L_{A95} - 30) \quad , \quad \text{EVT} = 0,25 \times (L_{Aeq} - L_{A95}). \quad (1.17)$$

L'indice Harmonica est normalisé de sorte qu'un son ayant un niveau constant de 30 dB(A) ait une valeur de 0. Il est, par ailleurs, non borné, ce qui peut engendrer des valeurs supérieures à 10 dans certaines situations.

Selon les auteurs [73], la représentation est assez intuitive (voir exemple figure 1.10) et fidèle aux objectifs de clarté du projet : l'indice BGN est représenté par un rectangle dont la hauteur dépend de sa valeur; celui-ci est surmonté d'un triangle dont la hauteur dépend de la valeur de l'indice EVT. Un code couleur est introduit pour quantifier l'indice total et dépend de la période de la journée, comme indiqué dans le tableau 1.5. On peut voir que

tous les seuils sont minorés de 1 point pour la période nocturne par rapport à la période diurne.

Couleur	Période diurne (de 6 à 22h)	Période nocturne (de 22 à 6h)
	de 0 à 4	de 0 à 3
	de 4 à 8	de 3 à 7
	8 et +	7 et +

TABLE 1.5 – Code couleur de l'indice Harmonica en fonction de la période de la journée.

La figure 1.10 présente un exemple de l'évolution de l'indice Harmonica sur une journée. On peut voir différentes situations d'exposition avec globalement un indice de bruit de fond assez faible mais une émergence assez importante dans la journée. Un exemple de différence d'appréciation (application du code couleur) est visible pour deux valeurs équivalentes de l'indice : à 5h (considéré comme période nocturne) et 6h (considéré comme période diurne) du matin.

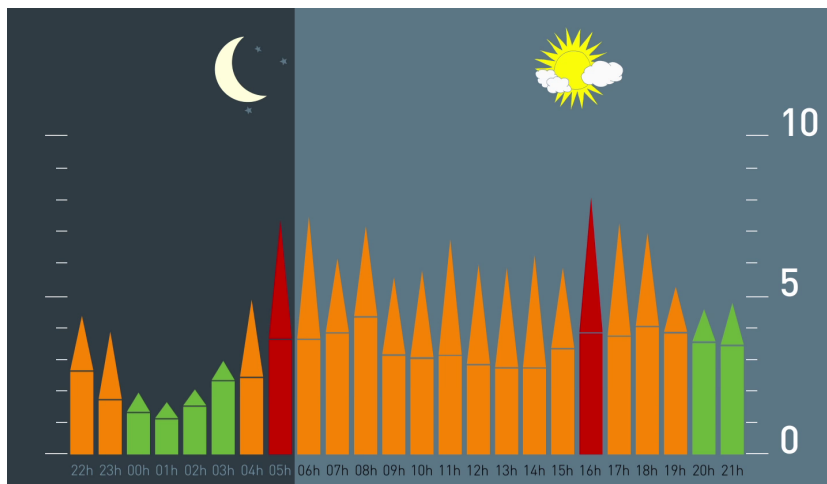


FIGURE 1.10 – Exemple d'évolution sur 24 heures de l'indice Harmonica - Source: Mietlicki *et al.*[73].

Les résultats de l'indice ont été comparés aux données issues de 246 questionnaires de résidents sur huit sites de mesures répartis dans les agglomérations lyonnaises et parisiennes. Les résultats pour six sites sont présentés en figure 1.11 où les auteurs montrent que le coefficient de détermination entre les gènes exprimées *in situ* et les valeurs de l'indice est de $R^2 = 0,96$ [73].

Ce modèle se base sur le niveau équivalent et sur le niveau de bruit de fond. Il résume donc l'évolution de l'exposition au bruit par deux valeurs constantes sur une période donnée. D'autres approches consistent à analyser plus finement l'évolution du niveau sonore instantané.

Modèle d'événements détectés - De Coensel *et al.* (2009)

Les travaux de De Coensel *et al.*[35] se basent sur l'hypothèse que la gêne à long-terme est déterminée principalement par les événements sonores détectés par le sujet. Ils proposent

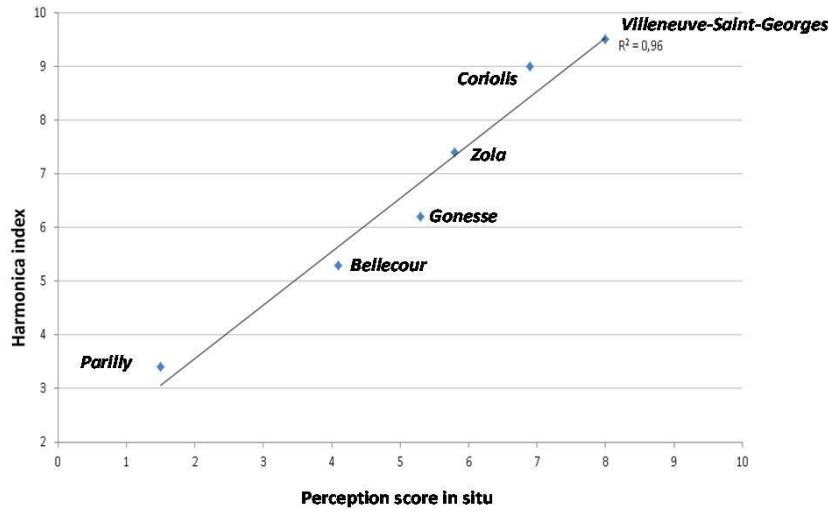


FIGURE 1.11 – Valeurs de l'indice Harmonica en fonction des gênes exprimées *in situ* - De Mietlicki *et al.* [73].

alors un modèle cognitif basé sur une vision très physique du cerveau humain et, là encore, sur le $L_{A,eq}$. Ainsi, un modèle d'attention est proposé sur une base d'exponentielle décroissante dépendant lui-même d'un modèle d'accoutumance. Ce dernier est basé sur l'intégration du niveau instantané L_f passé pondéré comme suit :

$$h(t) = \frac{1}{\tau_h} \int_{-\infty}^t L_f(u) \cdot \exp\left(-\frac{t-u}{\tau_h}\right) du, \quad (1.18)$$

avec τ_h une constante de temps.

Ces modèles d'attention et d'accoutumance sont utilisés pour définir les instants de début et de fin pendant lesquels le bruit est détecté par l'auditeur. Ils définissent alors un niveau sonore d'exposition des sources détectées noté SEL_{thr} (pour *Noticed Sound Exposure Level*) qui est une variante du SEL exposé en section 1.1.5. Il est déterminé par

$$SEL_{thr} = 10 \log \left[\int_{t_b^i}^{t_e^i} 10^{L_{thr}(t)/10} dt \right], \quad (1.19)$$

où $[t_b^i, t_e^i]$ sont les temps de début et de fin pour chaque événement sonore détecté grâce au modèle d'attention et L_{thr} est donné par

$$L_{thr}(t) = L_f(t) - L_b(t) - L(t_e), \quad (1.20)$$

avec $L_b(t)$ le niveau de bruit de fond et $L(t_e)$ le niveau de détection de l'événement.

La figure 1.12 montre que ce modèle, basé sur des connaissances en cognition des événements sonores, permet de distinguer des types de trafic différents à partir des pentes différentes dans l'évolution du SEL_{thr} en fonction du L_{DEN} en dessous de 45 dB(A). Au-delà de 55 dB(A), les pentes sont équivalentes et le trafic ferroviaire a un SEL_{thr} plus élevé que le trafic routier.

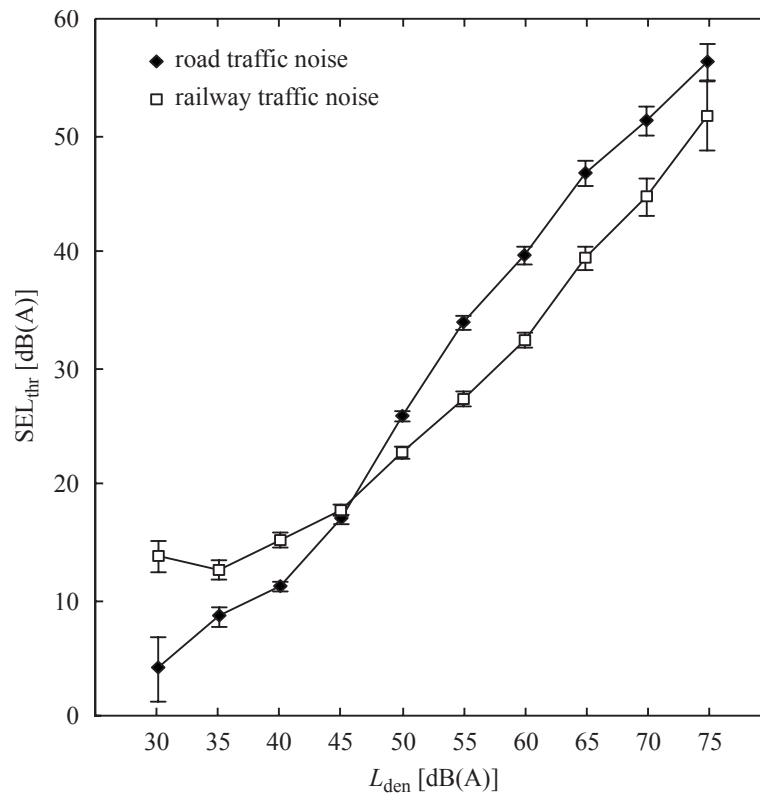


FIGURE 1.12 – Niveau d'exposition sonore détecté SEL_{thr} causé par l'exposition au bruit de la circulation routière et ferroviaire en fonction de leur niveau L_{DEN} simulé. Source : De Coensel *et al.* [35].

Ce modèle a servi de base à la création d'un modèle d'attention basé sur la saillance des sons [32] qui a été implémenté et testé par Terroir *et al.* [105]. Dans cette étude, ils comparent le nombre d'événements sonores détectés par ce modèle et ceux détectés par des sujets en laboratoire. Ces résultats montrent de bons accords. En revanche, bien que disponibles, ils ne comparent pas les données de gêne sonore perçue avec un modèle.

Modèle de gêne sonore de Morel *et al.* (2016)

L'approche de Morel *et al.* [77] citée précédemment, cherche à décrire le trafic routier par les sources qui le constituent. L'objectif de cette étude est de proposer une typologie perceptive et cognitive des bruits de passage de véhicules routiers. Pour cela, il a été demandé aux 58 participants de l'étude d'écouter les 57 bruits de passage disponibles dans leur corpus et de réaliser une tâche de catégorisation libre, donc de les regrouper par groupe de sons similaires.

Ce corpus a été obtenu par des mesures stéréophoniques dans différentes rues, pour différents types de véhicules. Le corpus était présenté aux sujets via une reproduction stéréophonique en laboratoire avec un niveau équivalent de 60 dB(A) au point d'écoute, ceci dans le but de chercher les aspects du bruit qui influent sur cette catégorisation en plus du facteur important qu'est le niveau du son. Une tâche de verbalisation était aussi demandée aux participants pour qu'ils disent quelles étaient les similitudes intra-groupe et les différences inter-groupes.

A la suite de cette expérience Morel *et al.* proposent une typologie perceptive et cognitive des bruits de passage de véhicules basée sur sept catégories :

1. Deux-roues passant à vitesse constante,
2. Deux-roues en accélération,
3. Bus, véhicules lourds et légers passant à vitesse constante,
4. Deux-roues en décélération,
5. Bus, véhicules lourds et légers en décélération,
6. Véhicules légers en accélération,
7. Bus et véhicules lourds en accélération.

Cette étude a ensuite été complétée par une seconde [78] portant sur la gêne sonore induite par le trafic routier. Les auteurs proposent alors, sur la même base de données de bruits de trafic routier regroupés en sept catégories perceptives, de concevoir un modèle liant la gêne ressentie par trente sujets et les indices énergétiques ou psycho-acoustiques mesurables sur le signal audio de chaque véhicule appartenant aux sept catégories perceptives proposées précédemment.

Pendant les expériences qui ont permis d'établir le modèle, les sujets doivent s'imaginer chez eux en train de faire une activité. Chaque stimulus (issus du même corpus de son que précédemment [77]) est présenté aux sujets pour sept niveaux sonores entre 50 et 62 dB(A). Il leur est alors demandé de noter la gêne due à ce bruit entre 0 (pas du tout gênant) et 10 (extrêmement gênant). Les indices utilisés pour construire le modèle ont été calculés grâce au logiciel dBsonic de 01dB² qui n'est plus commercialisé par le fournisseur. La sonie est calculée avec le modèle de Zwicker basé sur la norme DIN 45631. Le calcul de l'acuité et de la rugosité sont calculés grâce au modèle d'Aures [11, 12].

Le modèle de gêne pour chaque catégorie est présenté en tableau 1.6. On peut constater que la sonie est l'un des facteurs qui est le plus utilisé dans ce modèle. Elle est complétée par la force de fluctuation, la rugosité ou encore le niveau sonore en moyennes fréquences. Par ailleurs, les coefficients de détermination sont assez élevés.

Dans ce tableau, p est le facteur déterminant la probabilité de validation de l'hypothèse nulle, celle-ci disant que la variable n'est pas significative. La valeur $p < 0,05$ est classiquement admise comme seuil de significativité.

Ce modèle montre l'importance des variations spectrales et temporelles du bruit dans l'estimation de la gêne sonore, confirmant ainsi les observations de Paviotti *et al.* [83], Vos [118] ou encore Trollé *et al.* [107].

Modèle de gêne sonore de Klein *et al.* (2015)

Selon Vos [118], les variations spectrales et temporelles, traduites par les indices perceptifs que sont la rugosité, la force de fluctuation ou la sonie, peuvent expliquer 51% de la va-

2. dbkes.com.tr/brosur/dbsonic.pdf

Catégorie	Modèle de gêne sonore A	R ²	Écart-type
1ère Cat.	$A_1 = 1,03N^a + 0,18$	0,95	0,32
2ème Cat.	$A_2 = 16,99N_{15-18}^a + 0,10F^a + 1,45$	0,91	0,37
3ème Cat.	$A_3 = 1,32N^a - 0,32\Delta N^{-a} - 0,36$	0,97	0,23
4ème Cat.	$A_4 = 0,89N^a + 0,02R_{max}^c + 0,33$	0,96	0,29
5ème Cat.	$A_5 = 1,07N^a + 0,08F_{max}^a - 1$	0,95	0,34
6ème Cat.	$A_6 = 0,29L_{MF}^a - 8,5$	0,94	0,34
7ème Cat.	$A_7 = 0,95N^a + 0,10F^b - 0,5$	0,94	0,34

$${}^a p < 0,001 \quad {}^b p < 0,01 \quad {}^c p < 0,05$$

TABLE 1.6 – Indicateurs de gêne sonore. Les régressions entre les indices pertinents et les indices de gêne moyen sont données, avec le coefficient de détermination R² et l'écart-type de l'estimation. N est la sonie, N₁₅₋₁₈ est la sonie intégrée entre 15 et 18 Barks, R_{max} est la rugosité maximale, F est la force de fluctuation, F_{max} est la force de fluctuation maximale, ΔN⁻ est le taux de diminution de la sonie dans le temps, L_{MF} est le niveau L_{Aeq} calculé pour les bandes de tiers d'octave de fréquences moyennes (entre 315 et 1250 Hz) - p est la probabilité de rejeter l'hypothèse de non-significativité - D'après [78].

riance dans l'estimation de la gêne due au trafic routier et ferroviaire. Cependant, analysant les résultats de Morel *et al.* [78], Klein *et al.* [59] mettent en avant une corrélation assez limitée dans certains cas entre la rugosité, la force de fluctuation mesurée et la gêne exprimée due au trafic routier urbain, notamment pour les deux-roues motorisés passant à vitesse constante où uniquement la sonie est conservée dans le modèle (voir tableau 1.6).

Klein *et al.* [59] présentent un modèle de gêne sonore utilisant des indices psychoacoustiques innovants caractérisant les variations temporelles et spectrales des même stimuli que Morel *et al.* [77, 78]. Ce modèle global, ne différenciant pas les types de véhicules ou les allures, estime la gêne sonore par :

$$A = 0,50^a N_{mean} + 2,85^a m_{sputt,10} + 3,51^a m_{nas,10} + 0,026^a TETC - 0,79, \quad {}^a p < 0,01, \quad (1.21)$$

où N_{mean} est la sonie moyenne, m_{sputt,10} est la valeur dépassée 10% du temps d'un indice proposé modélisant le facteur "pétaradant" du son, m_{nas,10} est la valeur dépassée 10% du temps d'un indice proposé modélisant le facteur "nasal" du son et TETC (pour *Total Energy of Tonal Components*) est un indice proposé par Trollé *et al.* [107] caractérisant le contenu tonal hautes fréquences d'un signal acoustique. Les facteurs m_{sputt} et m_{nas} sont calculés pour chaque trame temporelle i par:

$$m_{sputti} = \left[\frac{2 \cdot |P_{max}(2 - 100\text{Hz})|}{P(0)} \right], \quad m_{nas_i} = \left[\frac{2 \cdot |P_{max}(100 - 200\text{Hz})|}{P(0)} \right], \quad (1.22)$$

avec P le spectre de l'enveloppe du signal audio et P(0) = P(f = 0) : la composante continue du spectre.

La figure 1.13 montre le lien entre les indices proposés (m_{sputt,10} et m_{nas,10}) et la gêne exprimée. On distingue notamment deux nuages de valeurs : les voitures (tous types d'al-

lure confondus) qui ont une gêne ressentie faible et les autres véhicules (deux-roues, bus et poids-lourds) plus gênants. Les coefficients de corrélation sont donc assez faibles. Cependant, ce modèle couple ces deux indices avec le TETC et la sonie moyenne. Cela augmente sensiblement la corrélation entre le modèle et les gênes exprimées puisque le coefficient de détermination du modèle est $R^2 = 0,94$ et l'erreur-type est assez faible à 0,35. Avec ce mo-

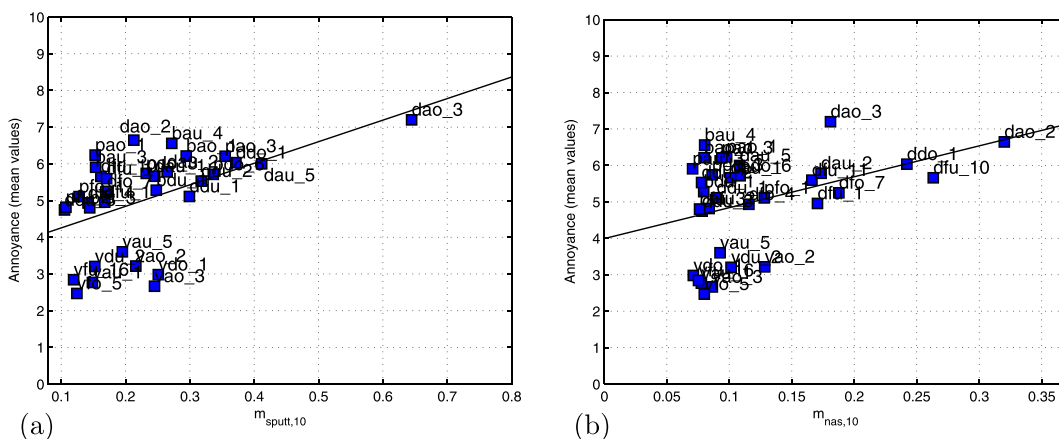


FIGURE 1.13 – Valeurs moyenne de gêne exprimée pour 33 bruits urbains de véhicules routiers en fonction de l'indice (a) $m_{sputt,10}$ (coefficient de corrélation $r = 0,5$, $p < 0,01$) et (b) $m_{nas,10}$ ($r = 0,4$, $p < 0,05$) - D'après Klein *et al.*[59].

dèle, les auteurs proposent des indices ($m_{sputt,10}$, $m_{nas,10}$ et TETC) qui représentent 47% de la valeur de gêne (et 53% pour la sonie). Ils mettent en avant ici le fait que, comme Paviotti *et al.* [83], la rugosité et la force de fluctuation, notamment utilisées par Morel *et al.* [78], ne sont pas adaptées à la caractérisation de la gêne induite par les deux-roues motorisés. Ils apportent une première explication qui serait que le mode de calcul de ces deux indices psychoacoustiques sont établis avec des fluctuations sinusoïdales alors que l'enveloppe du signal de bruits routiers est bien souvent plus complexes.

Conclusion

Dans ce chapitre, nous avons présenté des modèles disponibles dans la littérature pour quantifier la gêne sonore à partir de mesures acoustiques. Certains se basent sur le niveau instantané (données sonométriques), d'autres proposent une approche plus évoluée, nécessitant le signal audio pour en extraire des caractéristiques temporelles ou fréquentielles en fonction du temps.

Le premier indice présenté, Harmonica, propose une approche globale et énergétique de l'estimation de la gêne. En effet, la note de l'indice est donnée toutes les heures et générée par l'utilisation du niveau instantané pondéré A. Cet indice s'est révélé être assez efficace en termes de communication des résultats et corrèle assez bien avec le ressenti des habitants des six points de mesures réalisées dans les agglomérations lyonnaise et parisienne. En revanche, il simplifie l'expression de la gêne à deux informations que sont le bruit de fond et le bruit émergents. Cet indice ne permet donc pas de distinguer les sources de bruit et leurs quantités respectives.

Cette question de l'émergence, et donc de la détection d'un événement sonore, est au centre du modèle de De Coensel *et al.*. Posant l'hypothèse que la gêne sonore est reliée à la détection de l'événement sonore et de son émergence par rapport au bruit de fond, les auteurs proposent une adaptation du niveau d'exposition sonore (SEL) ne prenant en compte que les événements consciemment détectés. Dans ce modèle encore, les différentes sources de bruit ne sont pas différenciées. En effet, seule la différenciation entre le trafic ferroviaire et routier est proposée. Les résultats sont présentés sur des simulations de trafics et concluent bien à une différence de l'évolution du SEL_{thr} en fonction du niveau sonore, notamment à des niveaux faibles, mais l'étude ne pousse pas plus loin la segmentation des trafics et ne valide pas le modèle proposé avec une expérience perceptive.

Le modèle de Morel *et al.* [78] a ensuite été présenté. Il propose d'estimer la gêne sonore spécifique induite par chaque véhicule du trafic routier connaissant son appartenance à l'une des sept catégories perceptives (proposées précédemment par Morel *et al.* [77]). Ces catégories dépendent du type de véhicule et de son allure. L'estimation de la gêne est faite à partir de la somme pondérée d'indices énergétiques et psychoacoustiques qui dépendent de la catégorie du véhicule. Le modèle, utilisant des indices comme la sonie, la force de fluctuation ou encore la rugosité, donne des estimations ayant de bonnes corrélations avec les gênes exprimées.

Klein *et al.* [59], se basant sur les observations de Paviotti *et al.* [83] indiquant que la rugosité ne peut expliquer la gêne sonore, proposent de nouveaux indices pour exprimer les sensations induites par la fluctuation de l'enveloppe du signal de la source de bruit ($m_{sputt,10}$ et $m_{nas,10}$). Un modèle de gêne est alors proposé couplant ces indices à la sonie. Le coefficient de détermination ne semble pas meilleur que pour le modèle de Morel *et al.* [78], d'autant plus que le modèle global tel qu'il est proposé ne permet pas de représenter les différences perçues entre les types de véhicule.

Morel *et al.* [78] et Klein *et al.* [59] proposent deux modèles de gêne sonore intéressants issus d'expériences perceptives pendant lesquelles des bruits de véhicule au passage sont étudiés. Ils proposent respectivement d'utiliser des indices connus ou en suggèrent de nouveaux. Dans la suite de ce manuscrit nous avons privilégié l'utilisation du modèle de Morel *et al.* [78] car il permet de fournir une information plus détaillée.

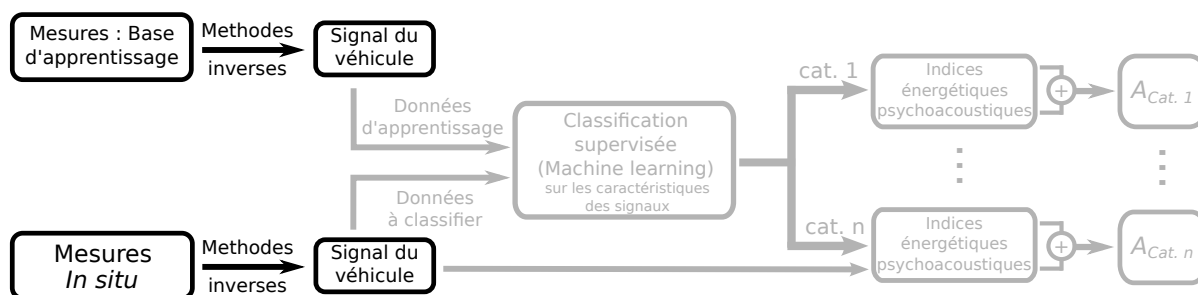
Les sons de véhicules au passage utilisés lors des écoutes en laboratoires de Morel *et al.* [78] et Klein *et al.* [59] sont issus de véhicules passant seuls dans la ville. Or, pour pouvoir utiliser ces modèles et systématiser l'estimation de la gêne sonore en milieu urbain, il faut pouvoir extraire le signal d'un véhicule dans une scène sonore complexe. Le chapitre suivant présentera des techniques permettant d'isoler chaque source d'une scène sonore complexe grâce à l'utilisation de grand réseaux de microphones.

Chapitre 2

Imagerie acoustique et extraction du signal audio de sources mobiles

Sommaire

Introduction	35
2.1 Méthodes inverses en antennerie acoustique	38
2.1.1 Modélisation du processus direct	39
2.1.2 La formation de voies, ou <i>beamforming</i>	40
2.2 Mégamicros, un système d'acquisition modulaire	44
2.2.1 Le système <i>Megamicros</i>	45
2.2.2 Une géométrie d'antenne pour chaque application	46
2.3 Localisation et quantification des sources de bruit	50
2.3.1 Véhicules légers	52
2.3.2 Autres véhicules routiers	58
2.4 L'antennerie acoustique, outil de séparation de sources	60
2.4.1 Extraction du signal audio d'une source fixe	60
2.4.2 Extraction du signal audio d'un véhicule en mouvement	61
Conclusion	67



Introduction

Dans les chapitres précédents, nous avons pu voir que la description de la source sonore est primordiale pour estimer son impact sur son environnement. Or les systèmes d'observation du bruit dans la ville généralement utilisés – les sonomètres – ne permettent d'accéder ni à la position ni à la nature de la source de bruit.

Des observatoires du bruit commencent à se doter de stations de mesure permettant de séparer deux types de bruits : le trafic aérien et les bruit terrestres. C’est notamment le cas de Bruitparif qui avait installé 19 sites de mesures permanentes aux abords des aéroports franciliens fin 2015 [3]. Ces instruments (RION NA-37) sont constitués d’une antenne de quatre microphones disposés selon un repère orthonormé permettant de détecter si le bruit vient d’au-dessus (aéronef) ou d’en dessous (source terrestre) du point de mesure situé au moins à 6 m du sol. Ils permettent aussi d’estimer l’angle d’arrivée de la source de bruit. Le système est présenté en figure 2.1 où l’on peut voir que l’antenne acoustique est couplée à un microphone de mesure complémentaire et à une station météorologique. Bruitparif a aussi commencé le déploiement de petites antennes composées de 4 microphones en septembre 2016 [4] appelées “méduses” en raison de leur forme. Ces systèmes ont été installés pour déterminer la provenance des sons enregistrés dans les quartiers animés de Paris, permettant ainsi de différencier les bruits des usagers des lieux (voix humaines essentiellement), ceux liés à la diffusion de musique ou encore ceux liés au trafic routier. Ces systèmes permettent une première segmentation de l’environnement sonore mais restent assez limités en précision de la détection et en capacité de séparation de plusieurs sources simultanées.

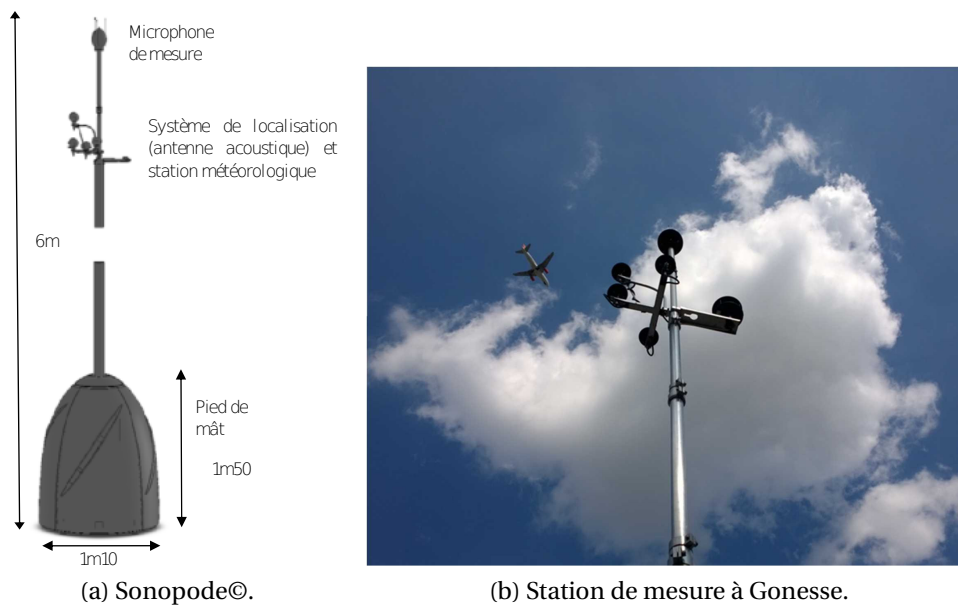


FIGURE 2.1 – Station de mesure du programme SURVOL de Bruitparif permettant la distinction du trafic aérien et routier.

D’autres études, comme celle de Marmaroli *et al.* [66], ont montré la capacité d’antenne de microphones à déterminer la vitesse et l’empattement de véhicules à deux essieux – donc le type de véhicule – dans le trafic routier, et cela grâce à l’inter-corrélation faite entre les signaux des deux microphones de l’antenne qui fait apparaître deux pics dus au bruit de contact pneu/chaussée des deux essieux. Ce système est très facile à installer mais ne permet pas de séparer les signaux des véhicules au passage et ne semble pas différencier les deux-roues des voitures.

Il est pourtant important de connaître la composition du trafic routier en terme de type de véhicule et d’allure pour pouvoir apporter une information riche sur le paysage sonore et estimer la gêne induite par chacun des véhicules du flux.

Or le rayonnement acoustique est différent pour chaque type de véhicule et chaque type d'allure. Comme on peut le voir sur la figure 2.2, grâce à l'antennerie acoustique (ici par formation de voies), il est possible de localiser les sources de bruit sur le véhicule et la composition spectrale de chacune. On peut voir qu'en sur-régime (en deuxième rapport à 50 km/h) c'est le moteur qui domine le bruit alors que dans le cas d'un rapport adapté (4^{ème} rapport à 50 km/h) les sources principales sont les contacts pneu/chaussée (voir figure 2.2b). En accélération, les sources de bruits sont similaires à la figure 2.2a : le moteur est la source dominante.

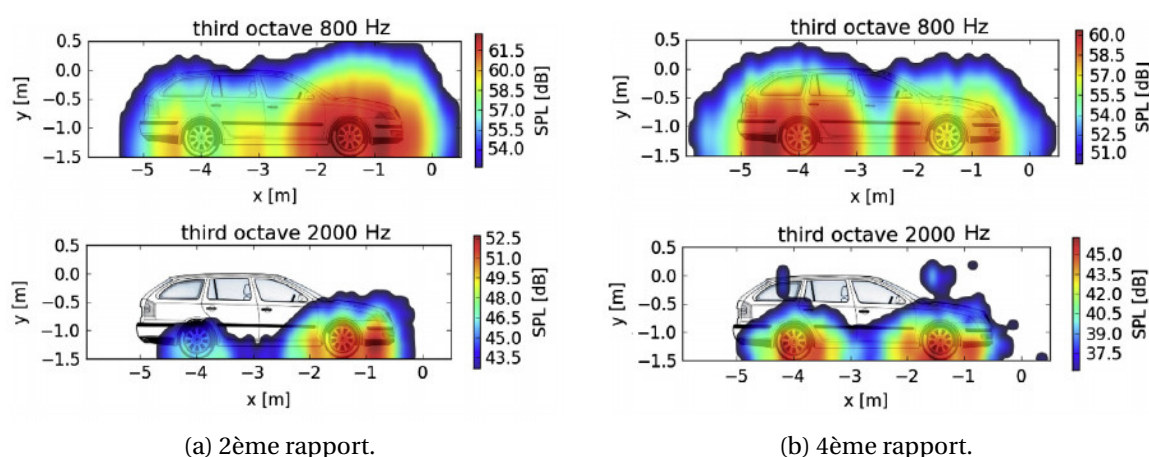


FIGURE 2.2 – Carte du niveau de bruit d'une voiture à essence passant à 50 km/h pour deux bandes de tiers d'octave différentes - De Ballestero *et al.* [17]

Les sources sonores présentes sur un véhicule routier aux différents régimes moteurs sont différentes et ont des signatures spectrales propres. Le bruit rayonné par un véhicule est donc caractéristique de son type et de son allure et son analyse doit permettre de l'identifier.

Par ailleurs, Weintein *et al.* [119] ont montré que l'extraction du son rayonné par une source de bruit fixe dans un environnement sonore complexe était possible grâce à l'utilisation d'antennes à grand nombre de microphones. Leur antenne comporte 1020 microphones répartis de façon régulière sur une surface de 1,77 m×0,96 m. Ils comparent les rapports signal sur bruit (RSB) du signal mesuré par un microphone proche de la source (Sennheiser HMD-410) et celui obtenu grâce au traitement des signaux de l'antenne. Leur microphone proche obtient un RSB de 35 dB et celui du signal issu de l'antenne passe de 17,2 dB avec un seul microphone activé à 30,9 dB avec les 1020 capteurs activés. De plus, ils proposent une expérience¹ dans laquelle plusieurs locuteurs évoluent dans le champ proche de l'antenne de 1020 microphones en lisant un texte. La formation de voies réalisée au niveau de chacun d'eux permet d'apprécier la capacité de filtrage de l'antenne.

1. <http://groups.csail.mit.edu/cag/mic-array/videos/microphone-array-switching.wmv>

Dans le même but, Hafizovic *et al.*[52] ont montré que ce filtrage spatial était possible sur des sources en mouvement et à plus grande distance de l'antenne. En effet, ils cherchent à extraire le son venant d'un joueur lors d'un match de basket-ball à l'aide d'une antenne circulaire de 300 microphones de 2,1 m de diamètre.

Il est donc possible d'extraire le signal audio d'une source mobile grâce à des antennes microphoniques. Hafizovic *et al.*[52] rappellent que l'efficacité du filtrage spatial dépend du nombre de microphones ce qui constitue un verrou lorsqu'on utilise des capteurs analogiques classiques.

Une autre approche est la séparation de source mono-canal, opérée sur un seul signal audio. Cette technique a été mise au point avant les années 2010 grâce à des algorithmes comme la factorisation de matrice non-négative (NMF). Elle fait désormais appel à l'apprentissage machine (*machine learning*) qui lui procure une meilleure efficacité. Emmanuel Vincent [114] synthétise les résultats du troisième challenge 'CHIME' (détails dans Barker *et al.* [18]) dont l'objectif est la reconnaissance vocale dans un environnement multi-sources bruité. Il y compare les résultats de différentes équipes de recherche utilisant des réseaux de neurones profonds (noté DNN, *Deep Neural Network*) ou encore du *beamforming* via l'estimation du taux d'erreur de reconnaissance de mots (WER). Le taux d'erreur du *beamforming* est de 26% contre 37% pour un DNN en monocanal. Seul un DNN appliqué à un réseau de microphones (multi-canaux) obtient de meilleurs résultats avec 20% d'erreur. La formation de voies est donc un bon outil de séparation de sources et c'est sur cette technique que nous avons fondé notre étude.

Ce chapitre présente, pour commencer, l'algorithme de formation de voies et explique son choix et son adaptation à notre étude. Le système d'acquisition *Megamicros* permet de relever le défi technologique de la construction d'antennes à grand nombre de microphones et de grandes dimensions. Il sera ensuite présenté au travers des deux antennes mises en œuvre pendant notre étude. En troisième partie nous nous intéresserons à identifier les sources de bruit présentes sur un véhicule en fonction de son type et de son allure. Enfin, la procédure d'extraction du signal de véhicules en mouvement sera présentée après avoir prouvé la faisabilité sur sources fixes.

2.1 Méthodes inverses en antennerie acoustique

Les antennes de microphones sont utilisées depuis quelques décennies pour estimer l'angle d'arrivée ou la position des sources de bruit dans l'espace. Elles permettent aussi d'identifier leurs caractéristiques spectrales en niveau et en fréquence. Pour cela, il faut utiliser des méthodes dites "inverses" qui consistent à inverser le processus direct de rayonnement des sources aux récepteurs suivant des modèles plus ou moins complexes.

2.1.1 Modélisation du processus direct

Il s'agit d'exprimer la pression acoustique $p_m(t)$ mesurée au $m^{\text{ème}}$ microphone d'une antenne résultant de la superposition des champs acoustiques issus de I sources potentielles localisées sur les points d'un maillage donné de l'espace (voir figure 2.3). Chaque source i est caractérisée par un signal propre $s_i(t)$. Dans le cadre classique de l'acoustique linéaire en milieu homogène, la pression au microphone m s'obtient par convolution de la pression rayonnée $s_i(t)$ avec la fonction de Green $G_{im}(t)$ qui décrit le trajet acoustique entre la source et le microphone.

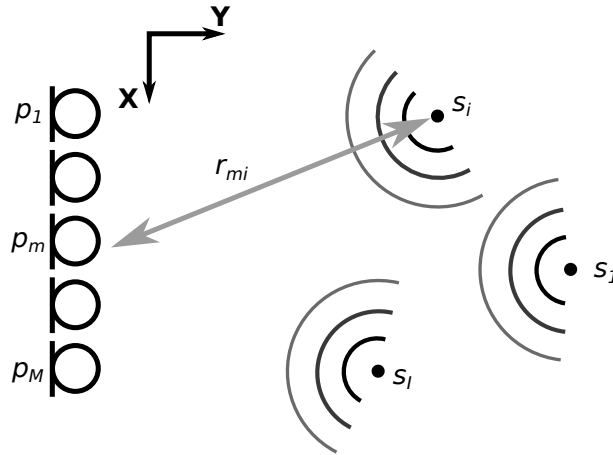


FIGURE 2.3 – Modèle de propagation en champ libre entre I sources et M microphones.

On suppose que les I sources sont monopolaires, que le milieu de propagation est homogène, sans écoulement et libre de tout obstacle. La fonction de Green de champ libre entre la source au point i et le micro m s'écrit

$$G_{im}(t) = \frac{1}{r_{mi}} \delta\left(t - \frac{r_{mi}}{c_0}\right), \quad (2.1)$$

où la distance microphone-source est $r_{mi} = \|\mathbf{x}_m - \mathbf{x}_i\|$, avec $\mathbf{x}_m = [x_m, y_m, z_m]$ et $\mathbf{x}_i = [x_i, y_i, z_i]$ les vecteurs position respectif du microphone m et de la source i . La célérité du son dans le milieu est notée c_0 . La pression $p_m(t)$ au microphone m est donc donnée par la convolution des signaux des I sources $s_i(t)$ avec la fonction de Green :

$$p_m(t) = \sum_{i=1}^I \frac{1}{r_{mi}} s_i(t - \tau_{mi}), \quad \tau_{mi} = \frac{r_{mi}}{c_0}, \quad (2.2)$$

où τ_{mi} est le temps de propagation associé à la distance r_{mi} .

Ce modèle de propagation peut aussi s'écrire dans le domaine fréquentiel, ce qui transforme le décalage temporel en déphasage. On obtient ce modèle de propagation $\hat{p}_m(f)$:

$$\hat{p}_m(f) = \sum_{i=1}^I \frac{1}{r_{mi}} \hat{s}_i(f) e^{-j\omega\tau_{mi}} \quad (2.3)$$

$$= \sum_{i=1}^I \frac{1}{r_{mi}} \hat{s}_i(f) e^{-jk r_{mi}}, \quad (2.4)$$

avec $\hat{s}_i(f)$ la transformée de Fourier du signal émis et $k = \frac{\omega}{c_0} = \frac{2\pi f}{c_0}$ le nombre d'onde. Cette expression peut s'écrire sous forme matricielle pour chaque fréquence en posant

$$\hat{\mathbf{p}} = [\hat{p}_1(f), \dots, \hat{p}_m(f), \dots, \hat{p}_N(f)]^T \text{ et } \hat{\mathbf{s}} = [\hat{s}_1(f), \dots, \hat{s}_i(f), \dots, \hat{s}_I(f)]^T :$$

$$\hat{\mathbf{p}} = \mathbf{A} \cdot \hat{\mathbf{s}}, \quad (2.5)$$

avec \mathbf{A} la matrice de propagation, ou fonction de Green, de dimensions $M \times I$ défini par :

$$\mathbf{A} = \begin{pmatrix} \frac{1}{r_{11}} e^{-jkr_{11}} & \dots & \frac{1}{r_{1i}} e^{-jkr_{1i}} & \dots & \frac{1}{r_{1I}} e^{-jkr_{1I}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{1}{r_{m1}} e^{-jkr_{m1}} & \dots & \frac{1}{r_{mi}} e^{-jkr_{mi}} & \dots & \frac{1}{r_{mI}} e^{-jkr_{mI}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{1}{r_{M1}} e^{-jkr_{M1}} & \dots & \frac{1}{r_{Mi}} e^{-jkr_{Mi}} & \dots & \frac{1}{r_{MI}} e^{-jkr_{MI}} \end{pmatrix}. \quad (2.6)$$

Dans le cas d'une propagation dans un milieu réverbérant ou non homogène, la fonction de Green doit être adaptée (voir par exemple [47, 7, 115, 84]).

2.1.2 La formation de voies, ou *beamforming*

Pour localiser et identifier les sources du bruit de trafic urbain perçues par les riverains, il faut inverser le modèle précédent. On cherche à estimer le champ acoustique au lieu supposé des sources en résolvant le problème inverse. Dans les études citées plus haut, c'est l'algorithme de formation de voies (*beamforming*) qui est employé. Cette méthode (autrement appelée "*Delay and Sum*") combine les signaux de pression enregistrés par l'antenne microphonique pour estimer le signal $s_i(t)$ au point i et à l'instant t comme :

$$s_i(t) = \frac{\sum_{m=1}^M \frac{1}{r_{mi}} p_m(t + \tau_{mi})}{\sum_{m=1}^M \frac{1}{r_{mi}^2}}, \quad (2.7)$$

avec $p_m(t)$ la pression à l'instant t et M le nombre de microphones. En pratique, l'opération peut aussi être réalisée dans le domaine des fréquences. Le spectre du signal source au point i et à la fréquence f s'écrit alors comme :

$$\hat{s}_i(f) = \frac{\sum_{m=1}^M \frac{1}{r_{mi}} \hat{p}_m(f) e^{j\omega r_{mi}/c_0}}{\sum_{m=1}^M \frac{1}{r_{mi}^2}}. \quad (2.8)$$

Caractéristiques et performances de la formation de voies

Cette section rappelle brièvement les paramètres fondamentaux et caractéristiques principales de la formation de voies avec une antenne linéaire. On s'intéresse en particulier à la résolution spatiale et à la dynamique en fonction de la longueur de l'antenne et du nombre de microphones qu'elle utilise.

Les fronts d'onde issus d'une source située dans le champ lointain d'une antenne peuvent être considérés comme plan, la fonction de Green est donc modifiée. Comme on peut le voir

sur la figure 2.4, le décalage temporel entre deux microphones vaut $\tau = (d \sin \theta_i) / c_0$. Il est donc uniquement fonction de l'angle d'arrivée θ_i et de la position relative des microphones.

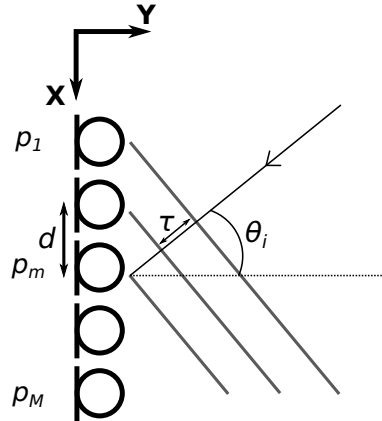


FIGURE 2.4 – Propagation acoustique d'une source située en champ lointain.

A partir de l'équation (2.8) et en prenant la pression reçue par les microphones comme unitaire, on obtient alors le diagramme de rayonnement de l'antenne D_i pour la $i^{\text{ème}}$ source en fonction de l'angle θ . Le diagramme, qui est alors unitaire au niveau de la source, est donné par

$$D_i(k, \theta) = \frac{1}{M} \sum_{m=1}^M e^{jkx_m(\sin\theta - \sin\theta_i)}. \quad (2.9)$$

Or, dans le cas d'un réseau linéaire de microphones à pas constant d , x_m peut s'écrire $x_m = (m-1)d - \frac{1}{2}$. On obtient :

$$|D_i(k, \theta)| = \frac{1}{M} \frac{\sin(Mkd(\sin\theta - \sin\theta_i)/2)}{\sin(kd(\sin\theta - \sin\theta_i)/2)} \quad (2.10)$$

La figure 2.5 montre des diagrammes de directivité (en décibel) en fonction de l'angle d'incidence θ pour une source harmonique de fréquence $f=500$ Hz située en $\theta_i = 0^\circ$ et une antenne de 20 m de long. La figure 2.5a montre le diagramme de directivité d'une antenne composée de 64 microphones. L'ouverture à -3 dB du lobe principal est de $2\theta_{3\text{dB}} = 1,7^\circ$. Cette valeur se retrouve en cherchant, à partir de l'équation (2.10) :

$$20 \log \left[\frac{1}{M} \frac{\sin(Mkd(\sin\theta_{3\text{dB}} - \sin\theta_i)/2)}{\sin(kd(\sin\theta_{3\text{dB}} - \sin\theta_i)/2)} \right] = -3 \text{ dB}, \quad (2.11)$$

On peut alors exprimer l'ouverture comme

$$2\theta_{3\text{dB}} \approx \frac{0,9\lambda}{L \cos\theta_i}, \quad (2.12)$$

avec $\lambda = \frac{c_0}{f}$ la longueur d'onde. Cette expression montre que la résolution est meilleure en face de l'antenne et dégradée sur les côtés. La figure 2.5a montre aussi des lobes secondaires – intrinsèques à la méthode – à -13 dB du lobe principal.

La figure 2.5b présente le diagramme de directivité pour deux antennes : l'une est composée de 64 microphones (courbe bleue), l'autre de 16 (courbe orange), toujours pour une longueur $L=20$ m. On confirme alors que la résolution ne dépend pas du nombre de microphones mais bien de l'envergure de l'antenne et de la fréquence de l'onde. En revanche, dans ce cas, des lobes de réseaux apparaissent et sont synonymes d'un échantillonnage spatial insuffisant. Le critère de Shannon pour l'espacement des microphones n'est pas respecté : d doit être suffisamment petit pour discrétiser une longueur d'onde. Il faut donc respecter :

$$d \leq \frac{\lambda}{2}, \quad (2.13)$$

ce qui n'est pas le cas pour $M=16$ où l'écart inter-microphones $d = 1,96\lambda$ mais qui est le cas pour $M=64$ où $d \approx \frac{\lambda}{2}$.

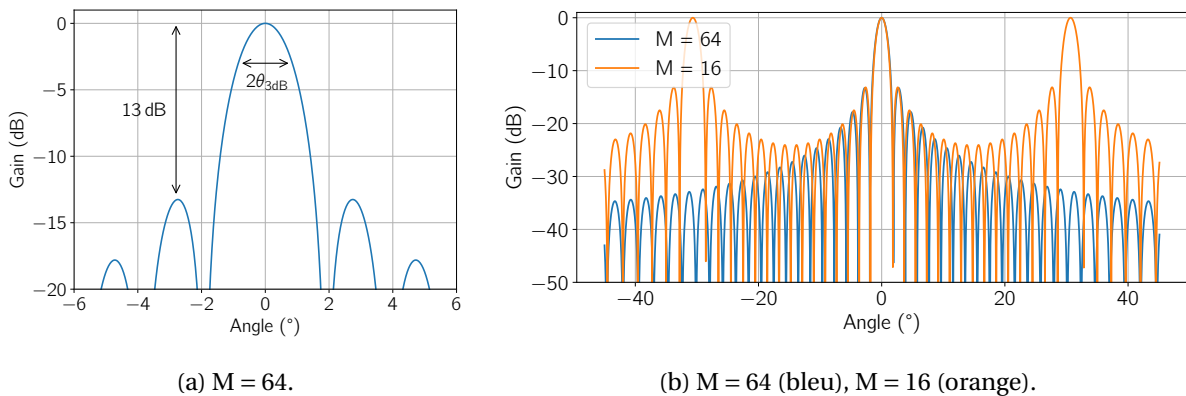


FIGURE 2.5 – Diagramme de directivité d'une antenne linéaire de longueur $L = 20$ m composée de M microphones en fonction de l'angle d'incidence.

Le nombre de microphones n'a donc pas d'influence sur la résolution de l'antenne si la longueur de l'antenne est fixe mais sur la bande fréquentielle d'utilisation de l'antenne. De plus, le nombre de microphones influence la dynamique de l'antenne. En effet, la sommation incohérente (de l'énergie) des diagrammes suivant les fréquences implique que ces lobes de réseaux (dont la position change en fonction de la fréquence) s'additionnent de façon destructive. Ainsi, la figure 2.6 montre le résultat en sommant l'énergie des diagrammes de 20 Hz à 25 kHz (fréquence maximale de notre système d'échantillonnage) pour un nombre de microphones allant de 32 à 1024 et une longueur d'antenne de 20 m. En s'intéressant au maximum des "lobes" secondaires, on remarque que le bruit de fond diminue de 3 dB quand le nombre de microphones double. La formation de voies nous permet donc de séparer les sources de bruit avec un RSB proportionnel $10 \log_{10}(M)$.

Méthodes haute-résolution

Comme on l'a vu précédemment, la formation de voies est aisée d'implémentation mais a certaines limites. On peut citer notamment la résolution, qui dépend de l'étendue spatiale de l'antenne et de la fréquence, mais aussi le rapport signal sur bruit qui dépend du nombre de microphones.

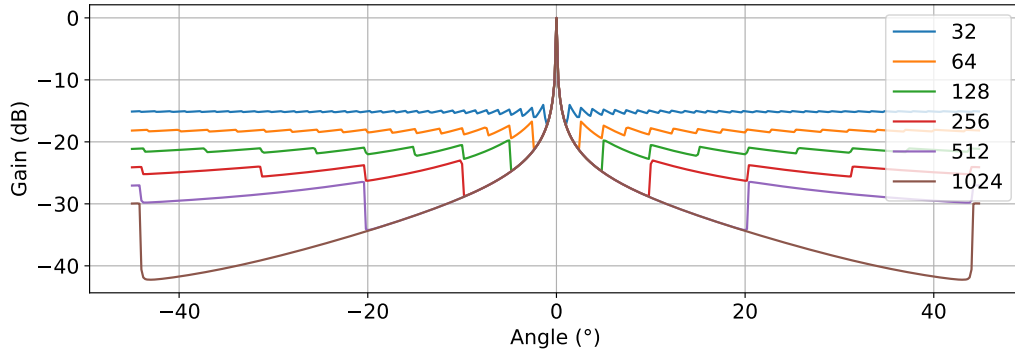


FIGURE 2.6 – Diagramme de directivité large bande [20 Hz - 25 kHz] d’une antenne linéaire de longueur $L = 20\text{ m}$ en fonction de la fréquence pour différents nombres de microphones ($M = \{32, 64, 128, 256, 512, 1024\}$).

Des méthodes dites à “haute-résolution” ont été proposées depuis la fin des années 1960 pour améliorer cette technique d’imagerie. Elles se basent sur l’estimation de la matrice interspectrale de puissance notée \mathbf{R} . Pour chaque fréquence f , \mathbf{R} est une matrice de dimensions $M \times M$:

$$\mathbf{R} = \langle \hat{\mathbf{p}}_m \cdot \hat{\mathbf{p}}_m^T \rangle \quad (2.14)$$

Ces techniques, telles que Capon (du nom de l’auteur) [30], MUSIC (*MUltiple Signal Classification*) [95] ou encore DAMAS (*Deconvolution Approach for the Mapping of Acoustic Sources*) [27], nécessitent soit d’inverser la matrice \mathbf{R} , soit de chercher ses valeurs et vecteurs propres. Le bon conditionnement de la matrice est requis, ce qui est généralement réalisé en moyennant plusieurs observations des interspectres.

Ces méthodes sont très efficaces pour localiser les sources de bruit dans des situations complexes (rapport signal sur bruit très faible, plusieurs sources rapprochées, etc.). Cependant, l’obtention d’une matrice \mathbf{R} de rang plein, via sa moyenne temporelle, a plusieurs inconvénients dans notre champs d’application. En effet, l’utilisation de telles méthodes est difficilement envisageable sur des sources de bruits en mouvement. Par ailleurs, la phase des signaux moyennée temporellement ne permet plus la reconstruction du signal audio de la source de bruit. En revanche le *beamforming* classique (hors sommation incohérente) permet de conserver cette information de phase. C’est pour cela que cette technique a été choisie, malgré ses limites, comme outil d’imagerie et d’extraction de signaux (développé en section 2.4) dans cette étude.

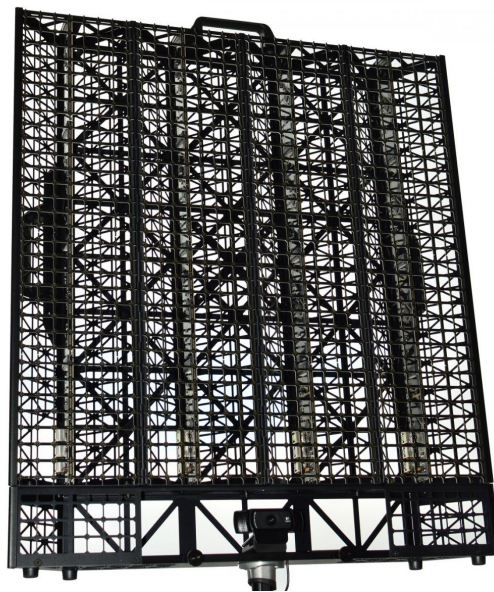
Le *beamforming* nécessite un grand nombre de microphones répartis sur la plus grande étendue spatiale possible pour obtenir les meilleurs résultats de séparation de source. Il faut donc pour cela des systèmes permettant l’acquisition synchrone de plusieurs centaines de voies.

2.2 Mégamicros, un système d'acquisition modulaire à grand nombre de microphones

Les systèmes à grand nombre de microphones cités précédemment (Weinstein *et al.* [119] et Hafizovic *et al.* [52]) ont prouvé leur intérêt pour la séparation de sources et le filtrage spatial et ont montré l'importance du nombre de microphones sur les résultats de leurs systèmes. Les systèmes commerciaux intégrant un grand nombre de microphones se basent désormais sur la technologie des microphones numériques MEMS (pour *MicroElectroMechanical Systems*) qui offrent une qualité d'enregistrement suffisante et un encombrement réduit. Le microphone intègre un convertisseur analogique-numérique. Il est petit (en moyenne de l'ordre du millimètre de côté) et consomme peu d'énergie, ce qui permet l'intégration d'un grand nombre de capteurs au sein d'un même système. Parmi ces systèmes, on peut citer par exemple Norsonic qui propose son système Nor848A composés de 128, 256 ou 384 microphones MEMS. Mais c'est Sorama qui propose actuellement à la vente le plus grand réseau : le CAM^{1k}, composé de 1024 microphones.



(a) Norsonic Nor848A - 384 microphones (1,6 m de diamètre) et 128 microphones (40 cm de diamètre).



(b) Sorama CAM^{1k} - 1024 microphones - Rétine rectangulaire de 64 × 64 cm.

FIGURE 2.7 – Exemples de grandes antennes de microphones disponibles dans le commerce.

L'antenne CAM^{1k} de Sorama se veut être utilisable pour plusieurs techniques d'imagerie : la formation de voies (détaillée dans la section 2.1) et l'holographie acoustique de champ proche. C'est pour pouvoir utiliser cette dernière méthode que les microphones sont espacés régulièrement de 2 cm. Les antennes de Norsonic présentées ainsi que l'antenne de Hafizovic *et al.* [52] ont été développées uniquement pour une utilisation avec la formation de voies. L'espacement radial des microphones est alors optimisé pour cette utilisation. Ces antennes sont donc conçues pour faire une image acoustique sur un plan.

Les systèmes commerciaux présentés ont ainsi un défaut majeur : le manque de flexibilité dans la géométrie de l'antenne. Cela ne permet pas de s'adapter aux différents besoins. En particulier, aucun système à notre connaissance ne permet l'acquisition synchrone de microphones espacés de plusieurs dizaines de mètres.

2.2.1 Le système *Megamicros*

Megamicros est un système d'acquisition à très grand nombre de voies conçu et développé à l'Institut Jean le Rond d'Alembert depuis 2012. Associé, lui aussi, à des microphones MEMS numériques², il permet aujourd'hui de déployer des antennes de très grandes dimensions et de géométries arbitraires pour capter les champs acoustiques et étudier une grande variété de phénomènes : diagnostic dynamique, aéroacoustique, bulles, etc. Le système est particulièrement bien adapté à notre étude, et nous en donnons ici un bref descriptif technique avant d'exposer les adaptations qui en ont été faites pour pouvoir extraire les sources de bruit du trafic urbain.

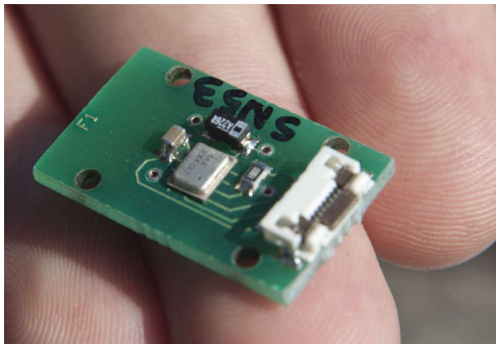
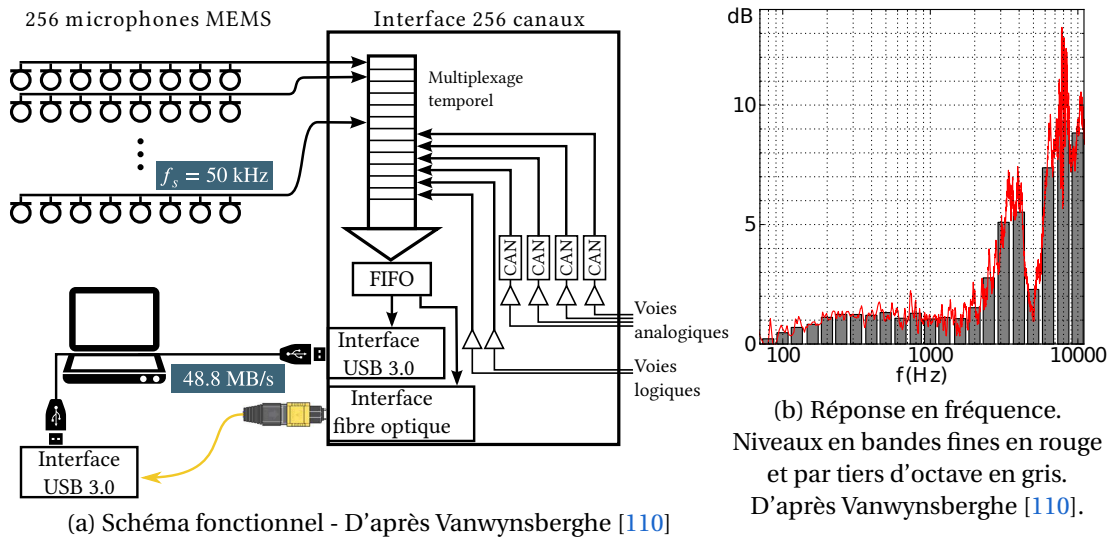
L'électronique du système. Comme le montre la figure 2.8a, les microphones sont assemblés par faisceaux de huit et connectés au concentrateur par câble RJ45 grâce au protocole I²S. Le système est également muni de 4 entrées analogiques supplémentaires, destinées à des capteurs de référence. Des voies logiques sont aussi acquises. Elles peuvent servir pour le déclenchement de l'acquisition par un système extérieur (barrières optiques par exemple).

Le premier système *Megamicros* pilotait 128 microphones et nécessitait un débit de 24,4 Mo/s pour la fréquence d'échantillonnage maximum de $f_e = 50$ Hz. Le passage à 256 microphones (et un débit associé de 48,8 Mo/s) a nécessité un changement d'interface. Le système à 256 microphones a été conçu avec deux interfaces : une USB3 (permettant un débit maximum théorique 500 Mo/s) et une interface fibre optique pour pouvoir déployer le système à distance du poste de pilotage d'une expérience. Ces évolutions d'interface permettent de concevoir un système à 1024 microphones. Ce système est en cours de développement et nécessitera un taux de transfert de 195,2 Mo/s.

Caractéristiques principales. Le capteur acoustique de base de *Mégamicros* est un microphone MEMS ADMP441 produit par Analog Device. On le voit figure 2.8c localisé au centre d'un circuit imprimé élémentaire. La figure 2.8d montre des faisceaux de microphones sur leur support, prêt à être installés dans différentes configurations (antennes 1D, 2D ou 3D).

D'après les résultats de Charles Vanwynsberghe [112], les microphones sont omnidirectionnels. La réponse en fréquence (figure 2.8b) montre une variation moyenne inférieure à 2 dB entre 80 et 2000 Hz. Dans cette plage le niveau est sur-estimé de 1 dB. De plus, un pic de 9 dB est présent entre 2 et 4 kHz. Bien que ces variations soient importantes sur certaines plages fréquentielles, elles sont très homogènes sur l'ensemble des microphones et peuvent donc être compensées à l'aide d'une réponse moyenne.

2. commercialisés en 2011



(c) Microphone MEMS



(d) 256 microphones montés sur profilé et équipés de leur protection au vent.

FIGURE 2.8 – Système Mégamicros à 256 microphones.

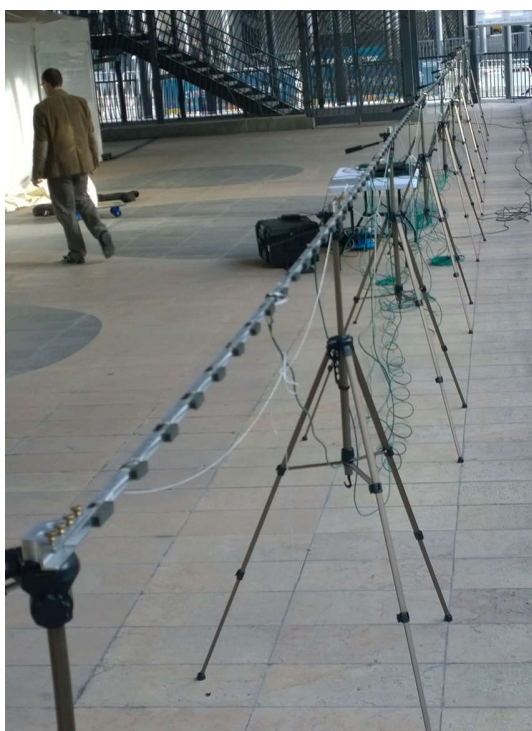
Les faisceaux de microphones peuvent être déportés du système d'acquisition avec des câbles RJ45 pouvant atteindre 20 m de long (au-delà le signal numérique est dégradé). On peut donc envisager de déployer une antenne de 40m de long, ce qui est bien plus important que les réseaux synchrones existant dans le commerce ou en laboratoire.

2.2.2 Une géométrie d'antenne pour chaque application

Le projet *Mégamicros* propose d'explorer différentes formes d'antennes pour différents usages allant de l'holographie acoustique de champ proche à la caméra acoustique en passant par la mesure du diagramme de rayonnement d'une source. Dans cette thèse deux antennes ont été réalisées (visibles en figure 2.9 et 2.11). Pour chacune d'elles l'objectif était d'avoir la meilleure résolution dans les basses fréquences. Comme vu précédemment (section 2.1.2), la résolution est déterminée par la taille de l'antenne.

La première antenne a été utilisée pour des mesures *in situ* sur le quai Saint-Bernard à Paris. Elle sera nommée "QSB-128". La seconde antenne, avec 256 microphones, a été utilisée pour des mesures sur une piste d'essai de PSA Peugeot-Citroën à La Ferté-Vidame. Elle sera référencée par "LFV-256".

Performance de l’antenne “QSB-128”. L’antenne présentée en figure 2.9 est linéaire (ou 1D) composée de 128 microphones et fait 21,6 m de long. Elle a été développée pour faire des images acoustiques dans un plan horizontal en ayant une bonne résolution en basses fréquences. Elle fournit une résolution angulaire de $2\theta_{3dB} = 1,6^\circ$ à 500 Hz. A titre de comparaison, l’antenne de 384 microphones de Norsonic offre une résolution angulaire $2\theta_{3dB} = 22^\circ$ à cette fréquence. Cette antenne peut être installée en trente minutes à deux personnes, qu’elle soit installée sur pied (figure 2.9a) ou attachée à une structure (figure 2.9b). Cela rompt avec les temps d’installation très importants nécessaires avec les systèmes analogiques utilisés antérieurement.



(a) Antenne sur pied à 1,6 m du sol imageant une rue en contre-bas .



(b) Antenne accrochée sur support en haut de la tour Zamansky (90 m d’altitude).

FIGURE 2.9 – Antenne linéaire de 128 microphones de 21,60 m de long avec un espacement régulier de $d = 17$ cm.

Pour tester l’adaptation de cette antenne à notre problème, on calcule son diagramme de rayonnement dans un plan désaxé figurant une voie urbaine en contrebas. Une source monopolaire émettant un bruit blanc est modélisée en $x = 0$ m, $y = 12$ m et $z = 0$ m et l’antenne linéaire est installée parallèlement à l’axe x en $y = 0$ m et $z = 8,5$ m d’altitude. Les résultats sont présentés en figure 2.10. On calcule l’image théorique dans le plan $z = 0$ intégrée sur la bande d’octave 500 Hz, d’une part (fig 2.10a et 2.10c) et sur la bande utile (60 Hz à 15 kHz - fig. 2.10b et 2.10d) d’autre part.

À 500 Hz, la résolution suivant x est de 44 cm et de 3,4 m suivant y . On voit donc qu’en basses fréquences il est impossible de distinguer les voies de circulations. En revanche, chaque véhicule sur une voie peut être isolé.

Le résultat de la formation de voies sur la bande utile montre une résolution à -3 dB de 4 cm. On dispose donc d'une très bonne résolution longitudinale (suivant x). Transversalement (suivant y) en revanche, la tache de focalisation atteint 25 cm. La figure 2.10d montre une résolution large bande correspondant à une cible en champ lointain (figure 2.6). Mais le niveau de bruit de fond augmente en s'écartant du centre de l'antenne. Cela étant dû à l'utilisation d'une fonction de Green en champ proche contrairement à l'exemple théorique (figure 2.6) en champ lointain.

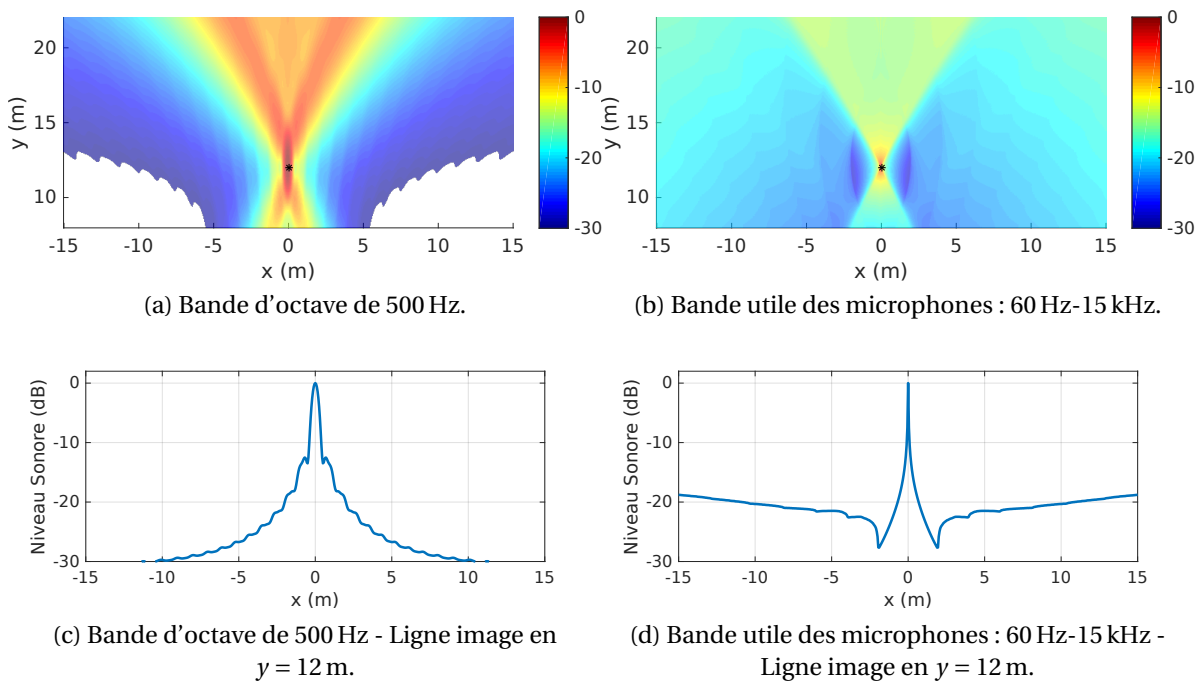


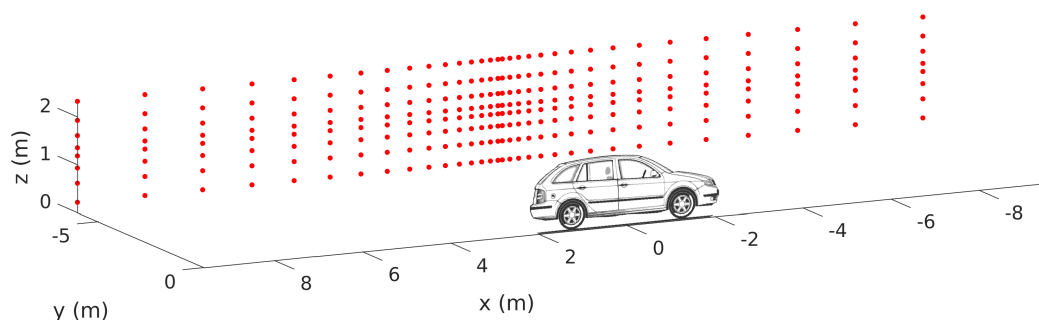
FIGURE 2.10 – Diagramme de rayonnement de l'antenne *Megamicros* linéaire de 128 microphones placée à $z = 8,5$ m. Source (croix noire) monopolaire large bande positionnée en $\mathbf{x} = [0, 12, 0]$. Plan image horizontal positionné en $z = 0$ m (a et b) et ligne image $y = 12$ m en $z = 0$ m (c et d). Dynamique : 30 dB.

Performances de l'antenne "LFV-256". L'expérience menée à La Ferté-Vidame, réalisée plus tard dans le projet, a bénéficié de développements plus récents de *Megamicros* et a utilisé une antenne de 256 microphones. L'antenne présentée en figure 2.11a a été conçue pour mesurer le champ acoustique par formation de voies sur un plan parallèle à celui de l'antenne. Les microphones ont été disposés sur des profils verticaux avec une distribution optimisée pour assurer la meilleure dynamique en réduisant le niveau des lobes secondaires tout en augmentant la capacité de résolution verticale. On a ainsi opté pour une répartition logarithmique en longueur (suivant x) et hauteur (suivant z), cette configuration permettant d'avoir une antenne grande ($19,6 \times 2,25$ m) tout en atténuant les lobes de réseau inhérents aux grands espacements latéraux (jusqu'à 1,5 m entre les faisceaux extrêmes et ceux adjacents).

La figure 2.11b décrit l'utilisation de cette antenne pour l'imagerie sur véhicules au passage. Elle est positionnée à 7,5 m du centre des véhicules, conformément à la position des sonomètres définie dans la norme de mesurage de bruit de véhicule au passage ISO 362.



(a) L'antenne sur la piste d'essais de PSA Peugeot-Citroën à La Ferté Vidame.



(b) Schéma d'implantation de l'antenne sur piste d'essai - Points rouges : positions des microphones.

FIGURE 2.11 – Antenne de 256 microphones réparti horizontalement et verticalement suivant une loi logarithmique par rapport au centre sur une surface rectangulaire de $19,6 \times 2,25$ m.

La figure 2.12 présente l'image théorique obtenue pour une source monopolaire large bande positionnée en $\mathbf{x} = [0, 0, 1, 27]$ et pour l'antenne située en $y = -7,5$ m. On peut voir que sur la bande d'octave de 500 Hz (figure 2.12a), la résolution dans le sens d'avancement des véhicules (suivant x) est de 40 cm et que le niveau de bruit est à -20 dB. Ce dernier est dû à l'intégration des lobes de réseau sur la bande. Contrairement à l'antenne linéaire de 128 microphones, cette antenne permet d'avoir une résolution verticale significative. Sur cette bande fréquentielle, la résolution à -3 dB est de 2,18 m.

La figure 2.12b présente l'image acoustique large bande obtenue par intégration de la formation de voies sur la bande fréquentielle utile des microphones. On note, au niveau de la source, une résolution horizontale de 3 cm et verticale de 16 cm (suivant z). Par ailleurs, la présence de lobes de réseaux à partir de $f = 110$ Hz limite la dynamique à 16 dB.

D'un point de vue opérationnel, l'encombrement de cette antenne, la distribution plus complexe de ses éléments compliquent notablement son installation et la rendent moins adaptée à une utilisation en milieu urbain par rapport à l'antenne linéaire décrite précédemment. Elle répond en revanche parfaitement à l'objectif visant à constituer une base de données de signaux caractéristiques des véhicules urbains, comme l'expose le paragraphe suivant.

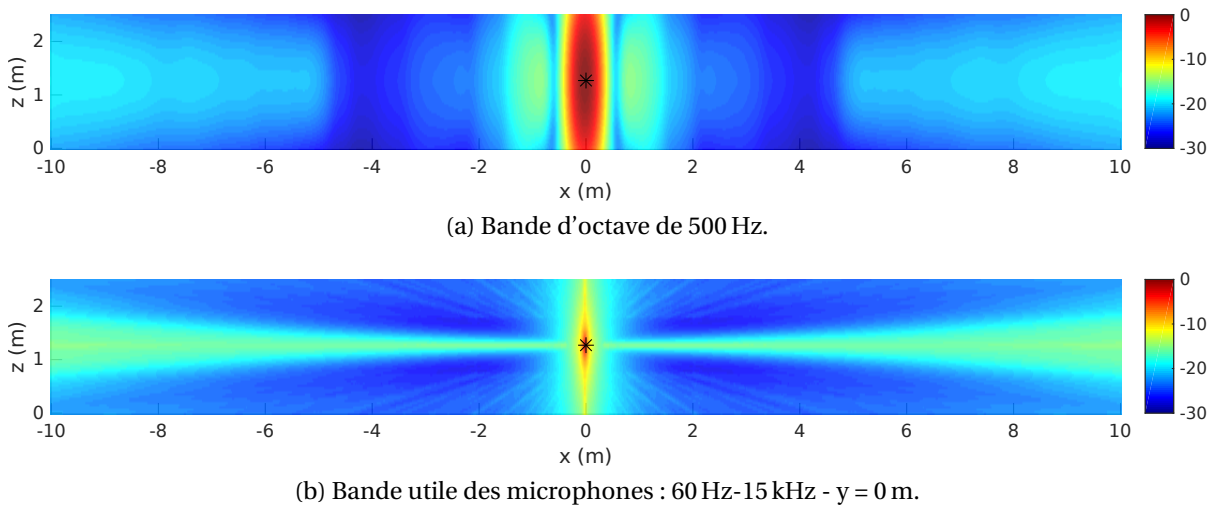


FIGURE 2.12 – Diagramme de rayonnement de l’antenne de 256 microphones placé à $y = -7,5$ m. Source (croix noire) monopolaire large bande positionné en $\mathbf{x} = [0, 0, 1,27]$. Plan image vertical positionné à $y = 0$ m. Dynamique : 30 dB.

2.3 Localisation et quantification des sources de bruit des véhicules routiers

L’antenne “LFV-256” (figure 2.11) est utilisée ici pour imager des sources acoustiques sur des véhicules routiers au passage à différentes allures. L’objectif a été d’identifier les sources acoustiques associées à des véhicules routiers, en termes de position et de contenu spectral, dans des conditions de conduite similaires au milieu urbain.

La figure 2.13 présente le schéma du dispositif expérimental utilisé sur le site de La Ferté-Vidame de PSA Peugeot-Citroën en janvier 2016. Ces essais ont été réalisés en se basant sur les recommandations de la norme ISO 362 relative à l’évaluation du bruit au passage de véhicule. Cependant les essais réalisés ont été adaptés pour tenir compte de la spécificité du trafic urbain.

Comme indiqué en figure 2.13a, l’antenne était placée à 7,5 m de la trajectoire du véhicule. Comme on peut le voir sur la figure 2.13b, un microphone de référence (Brüel et Kjør 4190) est placé au centre de l’antenne. En vue d’études ultérieures, une tête artificielle Head Acoustics a aussi été disposée au centre de l’antenne pour avoir des signaux binauraux. Le système d’acquisition était piloté par le “PC Master” via une fibre optique. Un anémomètre sonique, permettant de connaître la vitesse et la direction du vent, était installé en haut d’un coin de l’antenne de sorte à rejeter des mesures trop bruitées. Un radar Sagem MESTA 208 a été utilisé pour connaître précisément la vitesse des véhicules lors des passages. Des barrières optiques ont été placées à 20 m en amont et en aval du centre de l’antenne. Leur signal logique a été enregistré pour synchroniser le signal vidéo et les signaux acoustiques. Enfin, une webcam Genius WideCam F100 avec une ouverture angulaire de 120° a été utilisée pour avoir une image des véhicules sur toute la longueur de l’antenne. Les données de ces trois derniers outils de mesures ont été acquies par un ordinateur dit “PC Slave” qui recevait un ordre de départ du “PC Master” par ligne série RS232 émulée sur port USB.

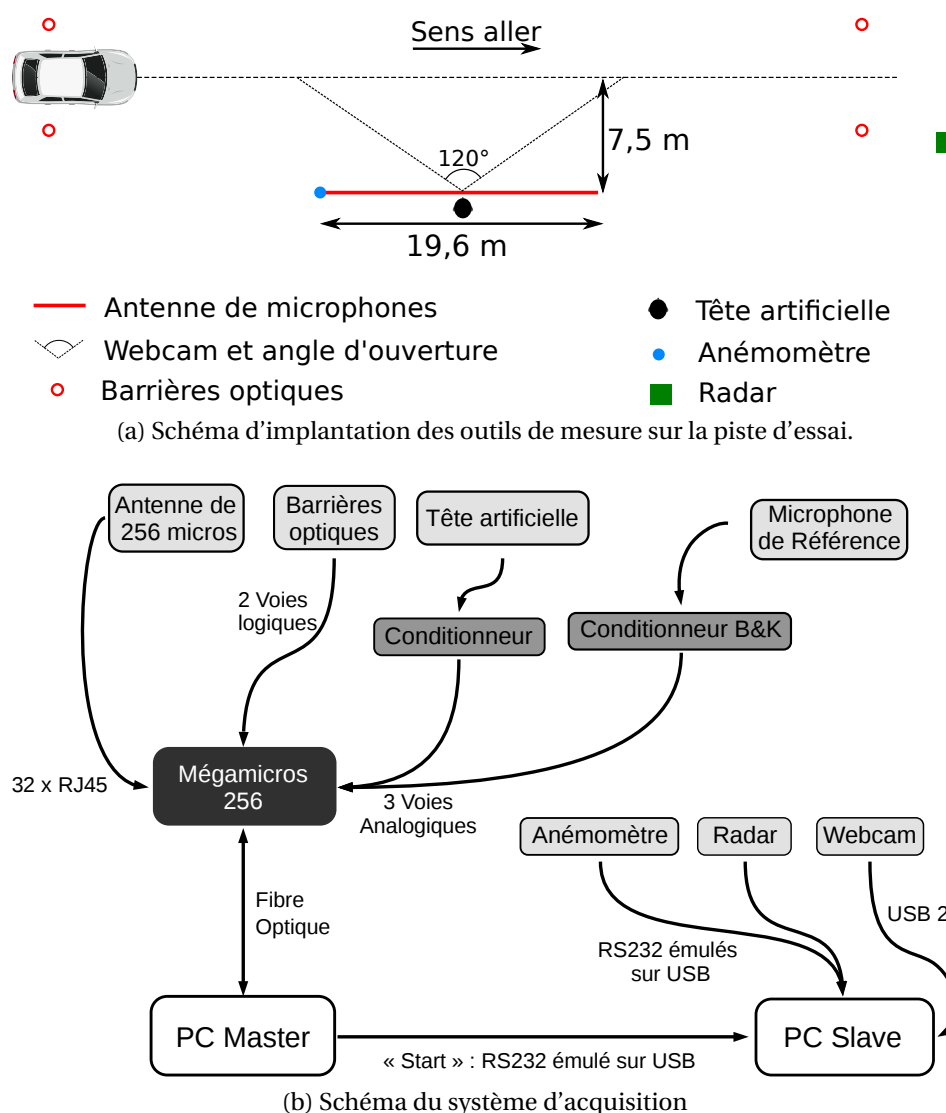


FIGURE 2.13 – Schémas d'implantation et informatique de l'expérience à La Ferté-Vidame.

L'ensemble des véhicules utilisés pour l'expérience est présenté dans le tableau 2.1. Les différents modèles³ couvrent une large gamme de type (deux-roues, véhicules légers, etc.) mais aussi de cylindrées, d'énergie et de gamme (berline, citadine ou monospace), définissant une typologie complète et représentative du trafic automobile urbain.

Ces véhicules étaient des véhicules de série déjà immatriculés. Les valeurs présentées dans ce manuscrit dépendent donc de l'histoire du véhicule (entretien, conducteur). Ces mesures ne sont donc pas faites dans les conditions d'homologation. Par ailleurs, la piste était humide et des gravillons étaient présents, ce qui peut avoir une influence sur le spectre du bruit de contact pneu/chaussée et les niveaux.

3. Les véhicules Peugeot et Citroën ont été mis à disposition par PSA Peugeot-Citroën dans le cadre de la chaire "MoUVie"

Nom du véhicule	Énergie	Cylindrée	Gamme	Type
V1-D-1	Diesel	4 cyl.	berline	Veh. Léger
V1-Es	essence	3 cyl.	citadine	Veh. Léger
V1-E1	électrique		citadine	Veh. Léger
V1-D-2	Diesel	4 cyl.	monospace	Veh. Léger
V1-D-3	Diesel	4 cyl.	berline	Veh. Léger
VL-D	Diesel	4 cyl.	utilitaire	Veh. Lourd
DR-E1	électrique		scooter	Deux-roues
DR-Es-1	essence	50 cm ³ , 2 temps	scooter	Deux-roues
DR-Es-2	essence	400 cm ³ , 4 temps	scooter	Deux-roues

TABLE 2.1 – Caractéristiques et noms des véhicules utilisés lors des essais sur piste.

L'objectif a été d'enregistrer des passages dans les différentes conditions de roulement qui sont propres au milieu urbain, en prenant en compte une éventuelle différence de rayonnement acoustique en fonction du côté du véhicule. Ainsi, pour l'ensemble des véhicules participant à l'expérience, les scénarios suivants ont été enregistrés :

- 25 km/h à vitesse constante en second rapport (aller/retour, moteur allumé ou éteint);
- 50 km/h à vitesse constante en troisième rapport (aller/retour, moteur allumé ou éteint);
- situation "feu rouge" : décélération de 30 à 0 km/h, ralenti moteur devant le centre de l'antenne puis accélération de 0 à 30 km/h avec passage de rapport (aller/retour);
- accélération plein gaz sur les 20 m devant l'antenne (aller/retour).

Une comparaison des sources de bruit observées pour différents véhicules et allures est présentée par la suite. Pour chaque cas considéré, l'image acoustique est constituée sur un plan vertical de 3 m de hauteur et placé à 7,5 m devant l'antenne (il passe donc au centre des véhicules).

Notre objectif est de décrire le spectre de chaque source sur la gamme utile : 60 Hz-15 kHz. Par comparaison, Ballesteros *et al.* [17] présentent des résultats analogues pour des fréquences supérieures à 800 Hz.

2.3.1 Véhicules légers

Cette section s'intéresse aux véhicules légers. On y analyse le rayonnement du V1-D-1 (4 cylindres, Diesel), du V1-Es (3 cylindres, essence), du V1-D-2 (4 cylindres, Diesel) et du V1-D-3 (4 cylindres, Diesel).

Passage à vitesse constante

La figure 2.14 présente le spectrogramme du microphone de référence lors d'un passage à vitesse constante (50 km/h) et les résultats d'imagerie sur différentes bandes d'octave quand le véhicule est au centre de l'antenne (soit à $t = 2,4$ s). Ces images sont réalisées sur des trames temporelles de 100 ms alors que le véhicule est en mouvement. Les niveaux présentés sont absolus, comme l'a montré Vanwysberghe *et al.* [112] en utilisant le système *Megamicros*.

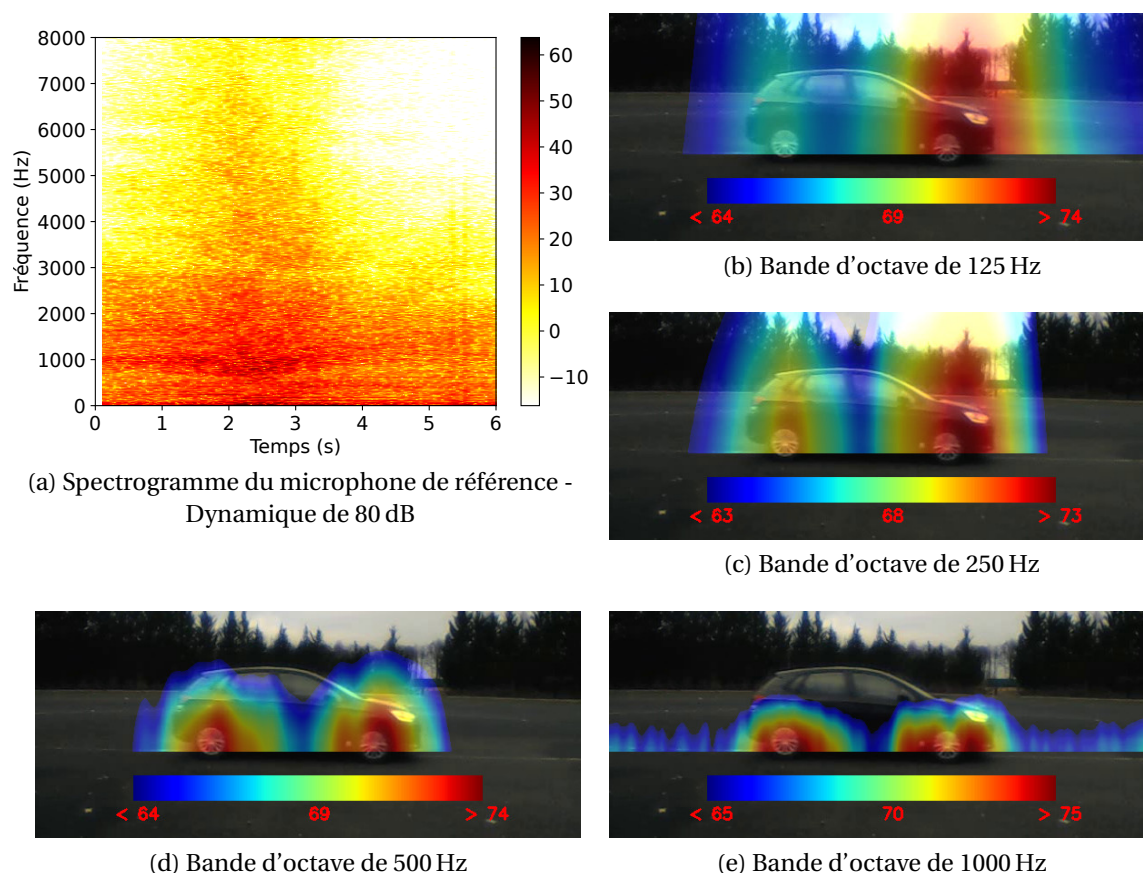


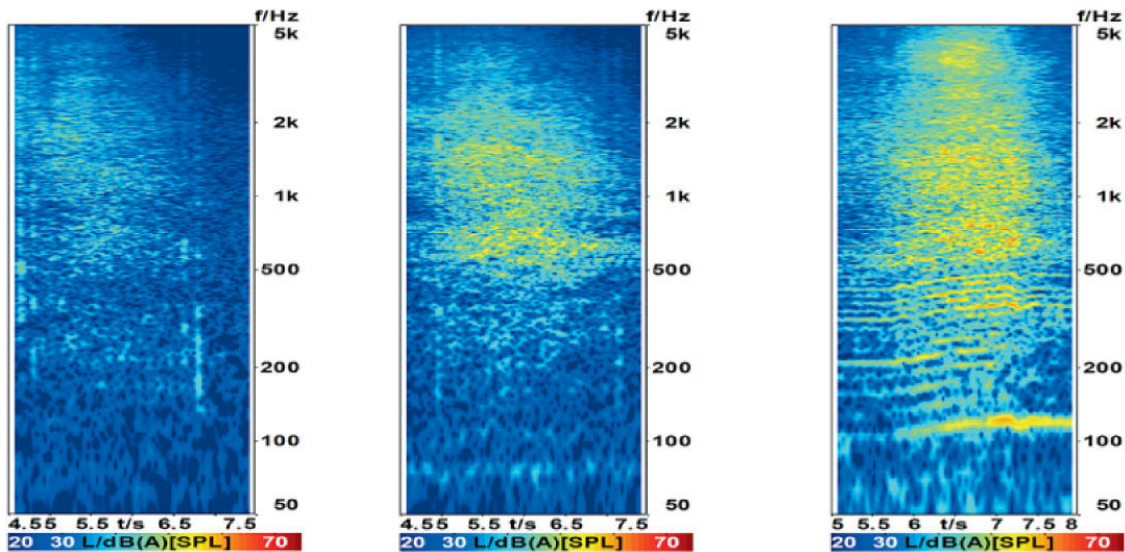
FIGURE 2.14 – Spectrogramme et images acoustiques à différentes fréquences du V1-D-1 – Vitesse constante de 50 km/h – $t = 2,4$ s

Le spectrogramme (figure 2.14a) révèle un bruit important jusqu'à 3 kHz dont la bande s'élargit au passage du centre de l'antenne ce qui suggère une directivité plus grande des sources issues du véhicule en hautes fréquences. Sandberg et Ejsmont [93] notent une directivité importante du bruit de contact pneu/chaussée. Le rayonnement de cette source est principalement vers l'avant en dessous de 1000 Hz, au delà c'est à l'arrière du pneu que le niveau est le plus important. Autour de 1000 Hz le rayonnement peut être considéré comme dipolaire.

Les figures 2.14b à 2.14e présentent les images acoustiques sur les bandes d'octave de 125 à 1000 Hz. On peut noter qu'en basses fréquences le moteur est la source dominante. En revanche, plus la fréquence augmente, plus le bruit dû au contact pneu/chaussée devient important. La figure 2.14d montre cette augmentation du niveau de la source de bruit de roulement. On voit aussi apparaître une source à l'arrière du pneu avec un niveau inférieur de 2 dB par rapport à la source moteur. Comme l'ont montré Sandberg et Ejsmont [93], on retrouve ici (figure 2.14e) le caractère dipolaire de la source de bruit de roulement autour de 1000 Hz.

Cela confirme les mesures de Alt *et al.* [8] qui constatent que le spectre du bruit de contact pneu/chaussée augmente de 200 à 500 Hz puis reste assez constant entre 500 Hz et 5 kHz (voir figure 2.15). On peut voir sur le spectrogramme présenté en figure 2.15c que, lors d'un essai selon la norme ISO 362 (accélération pleine charge en entrant dans la zone de mesure

à 30 km/h stabilisé), le contenu en basses fréquences est très tonal et se rajoute au bruit de roulement en hautes fréquences.

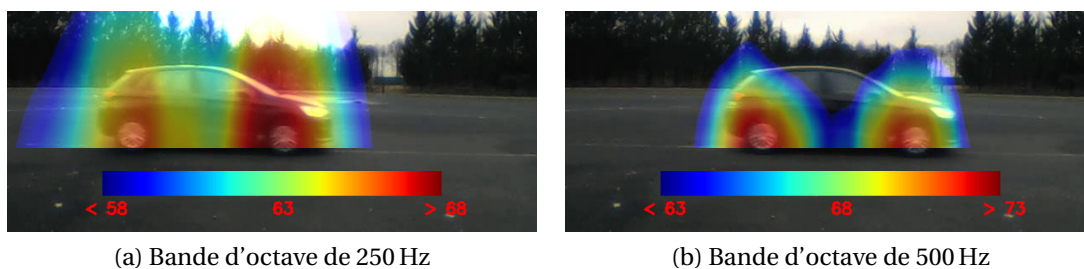


(a) Moteur éteint et pneus fins (b) Moteur éteint et pneus standards (c) Test selon la norme ISO 362

FIGURE 2.15 – Spectrogrammes de différents enregistrements de bruits de passage – De Alt *et al.* [8]

La figure 2.14c montre une source principale : le bruit moteur estimé à 73 dB sur la bande d'octave de 250 Hz. La source induite par le contact pneu/chaussée est alors estimée à 68 dB. Pour confirmer l'interprétation de nos mesures, des essais ont été réalisés en coupant le moteur des véhicules en entrée de zone de mesure. Ainsi, la figure 2.16a montre une source de bruit située moins à l'avant du véhicule et 5 dB plus faible. On distingue donc deux sources pneumatiques dont les niveaux sont mesurés à 68 et 67 dB respectivement à l'avant et à l'arrière du véhicule.

La figure 2.16b révèle deux sources pneumatiques à 73 dB, soit le même niveau que la source pneu/chaussée arrière visible en figure 2.14d. Ce qui confirme que le bruit issu du moteur est dominant en basses fréquences mais que le bruit de roulement est la source principale en hautes fréquences à 50 km/h.



(a) Bande d'octave de 250 Hz

(b) Bande d'octave de 500 Hz

FIGURE 2.16 – Images acoustiques à différentes fréquences du V1-D-1 – Moteur coupé – Vitesse constante de 50 km/h.

L'abaissement de vitesse à 25 km/h (figure 2.17) induit une réduction du niveau des sources. Ainsi, à 250 Hz, le bruit moteur passe de 73 dB (figure 2.14c) à 68 dB (figure 2.17a). Les bruits pneumatiques diminuent aussi mais de façon moins significative (1 dB). En revanche, à 500 Hz (figure 2.17b) la différence entre les deux types de sources est bien nette : c'est le moteur qui est plus bruyant de 5 dB.

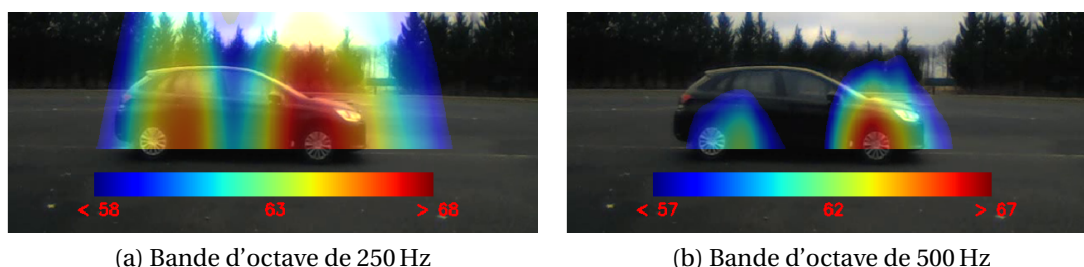


FIGURE 2.17 – Images acoustiques à différentes fréquences du V1-D-1 – Moteur allumé – Vitesse constante de 25 km/h.

Comparaison Diesel / Essence. Différentes motorisations ont été testées et un comparatif peut être fait grâce au V1-Es (3 cylindres, essence). Cette dernière peut être comparée aux V1-D-2 et V1-D-3 (4 cylindres, Diesel). On constate une très faible différence en hautes fréquences. En revanche, sur la bande d'octave de 125 Hz, elle est notable. En effet, on observe figure 2.18a que la source principale sur cette bande fréquentielle est le système d'échappement. En revanche, pour les véhicules Diesel (V1-D-2 sur la figure 2.18b et V1-D-3 sur la figure 2.18c), le moteur est la source dominante avec des niveaux bien plus élevés : 80 ou 70 dB pour ces derniers contre 63 dB pour le V1-Es, qui a la plus petite cylindrée.

A noter qu'en augmentant la vitesse à 50 km/h, la distribution des sources sur le V1-Es est en tout point équivalente avec celle présentée pour une vitesse de 25 km/h (figure 2.18a). La source d'échappement est alors estimée à 70 dB. Les résultats détaillés ne sont donc pas présentés. Ceci est à comparer avec le V1-D-1 en figure 2.14b et le V1-D-3 en figure 2.18d où on peut voir une légère augmentation du niveau des sources à l'arrière du véhicule tout en gardant comme source principale le moteur.

Véhicule électrique. Le V1-E1 est une petite citadine qui n'a pas de sonification additionnelle. La figure 2.19 montre des images acoustiques à différentes fréquences et vitesses. On étudie ici l'influence de la vitesse du véhicule sur le niveau des sources à l'exclusion du moteur thermique sur différentes plages fréquentielles. Sandberg [92] indique que le niveau du bruit induit par le contact pneu/chaussée (noté $L_{P/C}$) est une relation logarithmique de la vitesse V (en km/h) telle que :

$$L_{P/C} = A + B \cdot \log(V). \quad (2.15)$$

On peut voir que cette dépendance ne se retrouve pas particulièrement en basse fréquence entre les figures 2.19a et 2.19b, ce qui est dû au spectre de ce bruit. En effet, comme il a été dit précédemment (figure 2.15), le bruit de contact pneu/chaussée devient significatif à partir de 500 Hz. On voit bien l'augmentation du niveau du contact pneumatique de 9 dB sur l'octave de 500 Hz (figures 2.19c et 2.19d) et ainsi que sur l'octave de 1000 Hz (figures 2.19e et

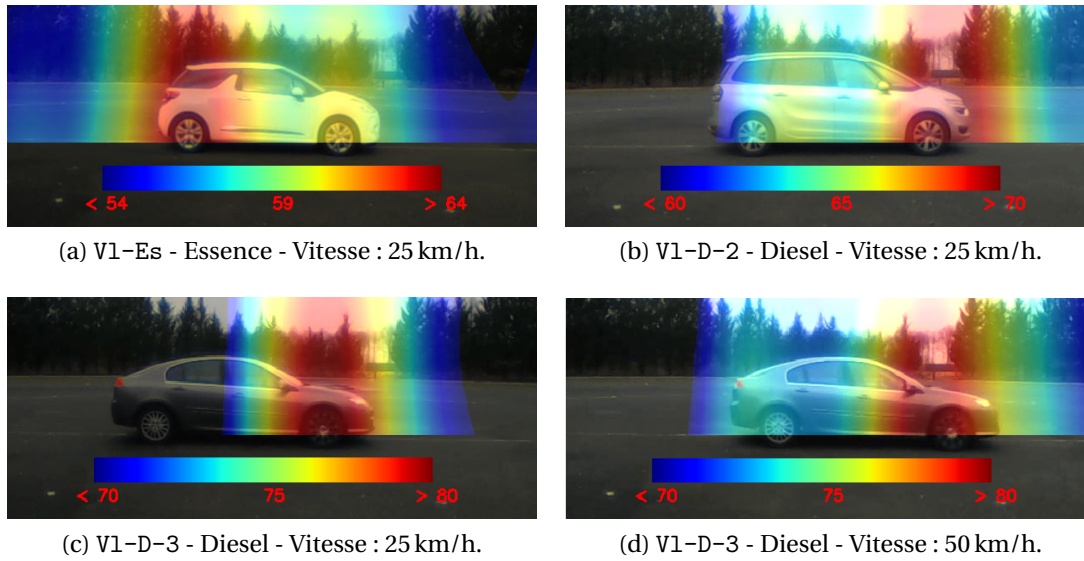


FIGURE 2.18 – Images acoustiques sur la bande de 125 Hz – Moteur allumé.

2.19f). On a donc ici $B \approx 30$, ce qui est cohérent avec la littérature pour ces bandes d'octaves [92, 81, 58].

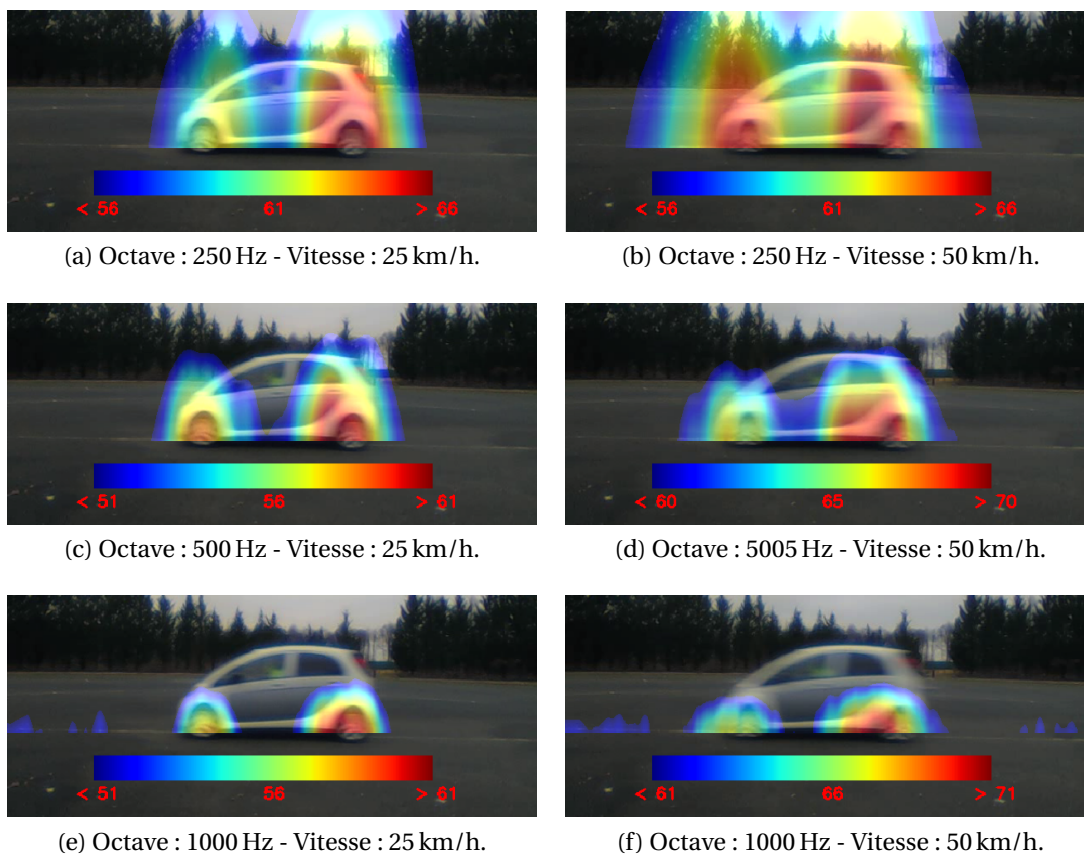


FIGURE 2.19 – Images acoustiques à différentes fréquences du V1-E1.

Sachant que le niveau maximal estimé sur le V1-Es à 50 km/h sur l'octave de 1000 Hz est de 72 dB, on peut noter le fait que le niveau de 71 dB atteint par ce véhicule n'est pas significativement différent. Ce qui confirme que l'intérêt des véhicules électriques dans la réduction des nuisances sonores ne se limite qu'aux faibles vitesses.

Passage en décélération

La figure 2.20 présente les résultats d'imagerie pour le V1-Es en décélération. Le spectrogramme présenté en figure 2.20a montre une composante basse fréquence du bruit importante. La figure 2.20b nous indique que ce contenu très basses fréquences est émis par l'échappement. Le bruit large bande au delà de 400 Hz (avant l'arrêt du véhicule à 7 s) semble lui provenir du contact pneu/chaussée comme le montre la figure 2.20c. L'image acoustique sur cette figure semble montrer aussi une composante du bruit induite par le moteur.

Cette répartition spatiale en fonction de la fréquence reste valable une fois le véhicule arrêté : à 125 Hz c'est l'échappement la source dominante, alors qu'à 1000 Hz c'est le bruit moteur (voir figures 2.20d et 2.20e).

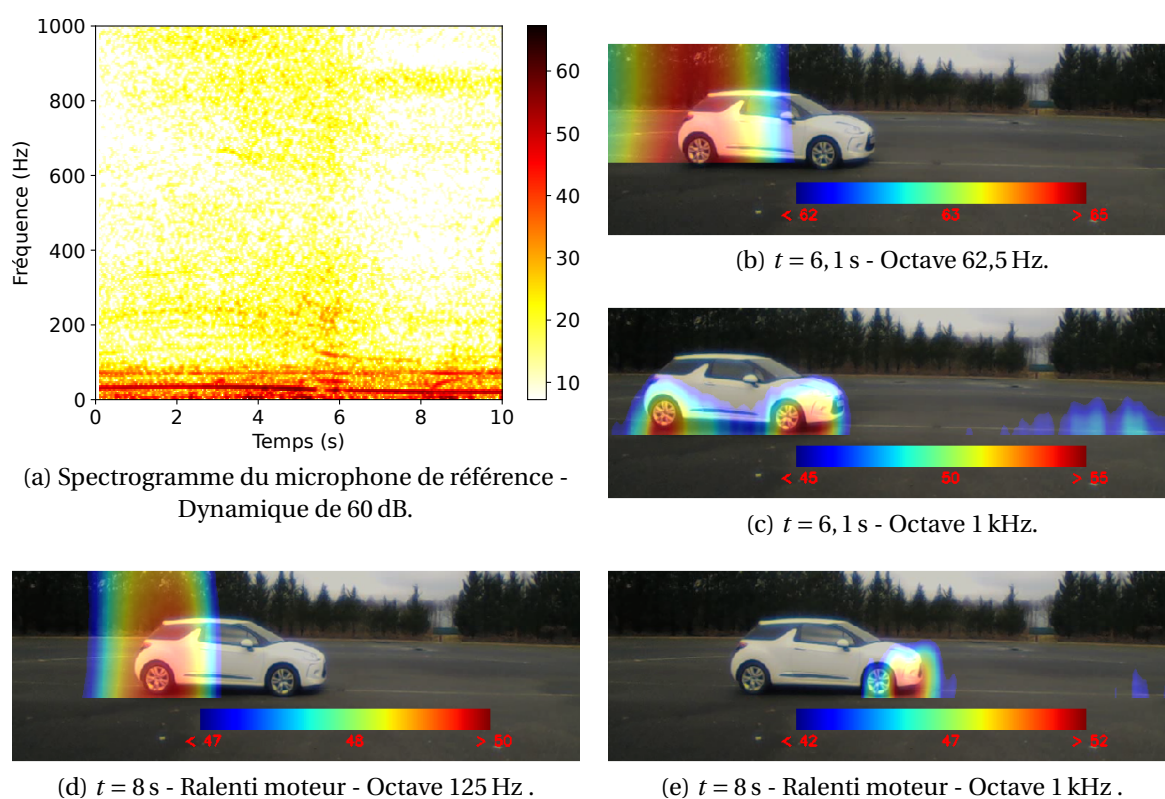


FIGURE 2.20 – Spectrogramme et images acoustiques lors d'une décélération du V1-Es.

Les véhicules Diesels disponibles lors de l'étude présentent eux une source dominante constante sur tout le spectre lors de la décélération : le bruit moteur (résultats non-présentés).

Passage en accélération

L'étude des sources acoustiques présentes lors de la phase d'accélération est faite avec le V1-Es. En effet, tous les véhicules légers testés ont montré les mêmes répartitions spatio-

fréquentielles de l'énergie acoustique.

Comme on peut le voir sur la figure 2.21a, le spectre du signal est très tonal dès que le véhicule accélère, à 12 s. Ce bruit est localisé en basses fréquences (figures 2.21b et 2.21c) au niveau de l'échappement. Aux plus hautes fréquences (octave de 500 Hz sur la figure 2.21e et au delà), la source dominante est située au niveau du moteur. Ce n'est que lors d'un passage de rapport (figure 2.21d) que l'échappement est une source importante de bruit.

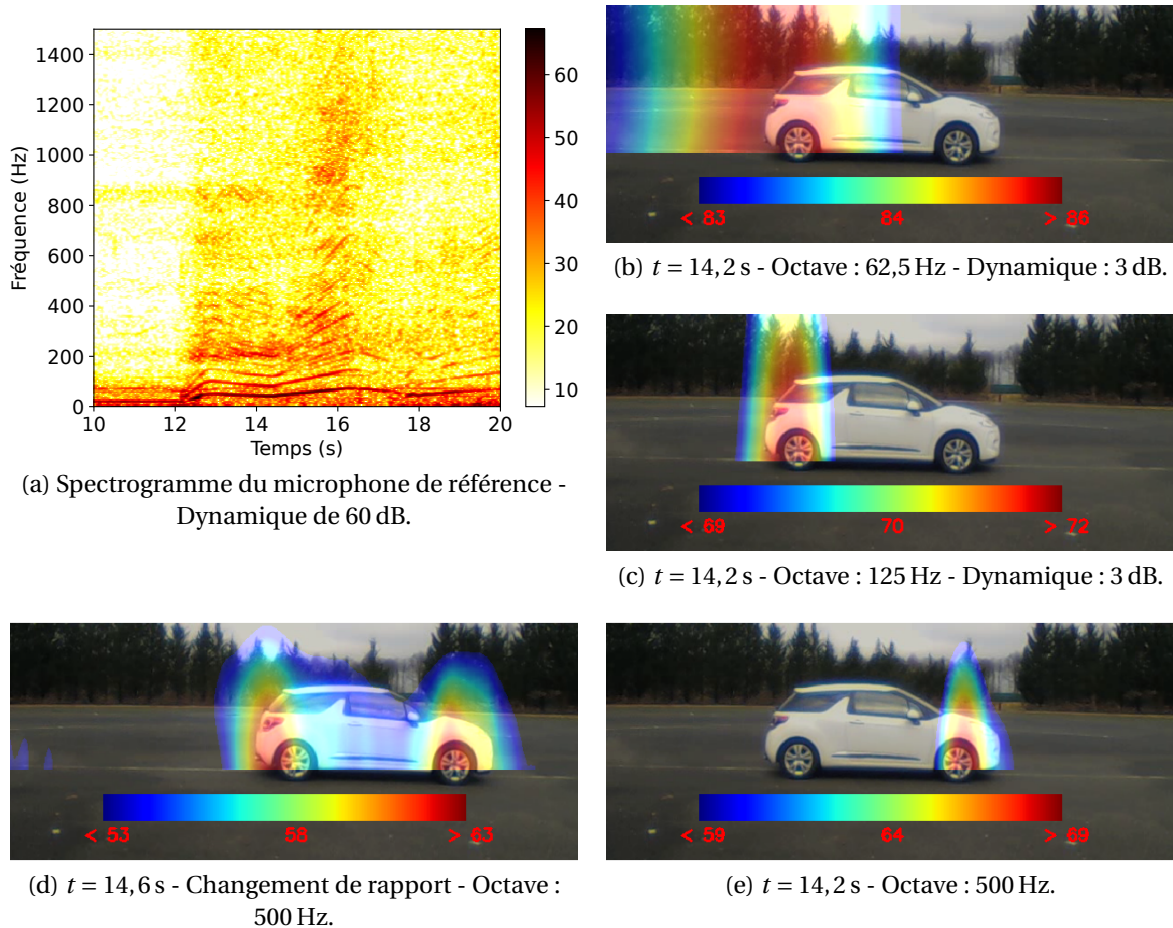


FIGURE 2.21 – Spectrogramme et images acoustiques du V1-Es – Accélération.

2.3.2 Autres véhicules routiers

Dans cette section nous décrivons succinctement le rayonnement acoustique des véhicules lourds et des deux-roues motorisés. Les résultats détaillés d'imagerie aux différentes allures de l'utilitaire et des différents scooters sont présentés en annexe, page 129.

Véhicules lourds. D'après les résultats de Pallas *et al.*[81], les sources acoustiques d'un véhicule lourd sont assez similaires – en fonction des plages fréquentielles et des allures – aux véhicules légers présentés précédemment, seule leur amplitude diffère. Nous observons aussi les mêmes sources. Il faut noter cependant la part faible du bruit moteur, lors d'un rapport élevé, même à une vitesse constante faible de 25 km/h (alors en 3^{ème} rapport).

Deux-roues motorisés. La figure 2.22 présente les résultats pour un deux-roues en accélération dans les deux sens sur une bande fréquentielle allant de 0 à 2 kHz. On note ici que la source principale de bruit est le moteur. En augmentant la fréquence d'étude la source semble se concentrer sur l'échappement. Cette analyse s'étend à tous les essais réalisés pour les deux-roues testés.

Nous observons aussi une différence de niveau sonore du bruit moteur entre les deux sens de passage. En effet, si le pot d'échappement est masqué par le scooter (figure 2.22a), le niveau est de 91 dB. Lorsqu'il est face à l'antenne (figure 2.22b), le niveau est de 94 dB.

Cela est en accord avec les observations de Chatagnon & Lelong [31] qui montrent une différence pouvant aller jusqu'à 3 dB entre les deux côtés d'un scooter. Cette différence est maximale pour un scooter de 50 cm³ et décroît avec l'augmentation de la cylindrée.

On observe aussi cette diminution pour le scooter de 400 cm³ testé (DR-Es-2). En effet, l'asymétrie de rayonnement de ce deux-roues se réduit à 2 dB en accélération, son niveau maximum est alors mesuré à 104 dB. En revanche l'estimation du contenu plus hautes fréquences ne semble pas être impactée par le côté d'enregistrement sur aucun des deux scooters.

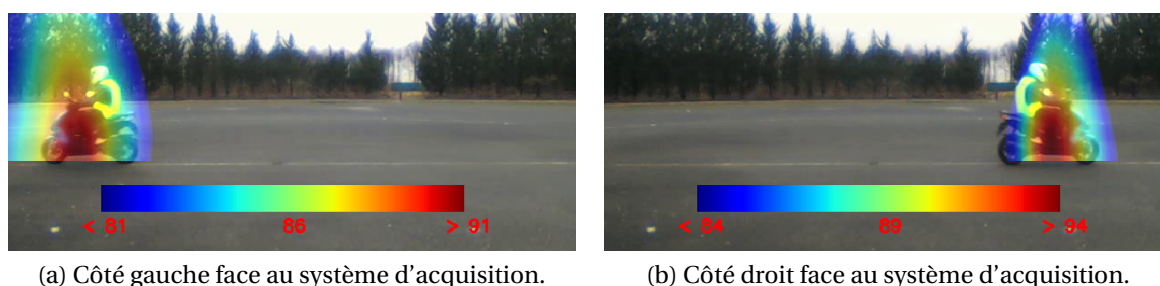


FIGURE 2.22 – Images acoustiques entre 0 et 2 kHz du DR-Es-1 (50 cm³) – Accélération – Dynamique : 10 dB.

Conclusion. L'utilisation du système *Megamicros* a permis de construire une antenne acoustique bien adaptée au besoin de l'imagerie de véhicule au passage. D'une envergure de 19,6 m de long sur 2,25 m de haut, elle nécessite la présence de 256 microphones pour avoir un nombre de capteurs suffisants. A notre connaissance, cette réalisation est unique. Elle permet d'obtenir des images acoustiques très précises pour des situations variées (différentes allures et véhicules) et avoir ainsi une inter-comparaison riche.

2.4 L'antennerie acoustique, outil de séparation de sources

Comme on l'a montré dans les sections précédentes, une antenne acoustique associée à la technique de formation de voies permet de visualiser et de caractériser les sources de bruit présentes dans le "champ de vision" de l'antenne avec une résolution plus ou moins grande suivant la fréquence et la géométrie de l'antenne. Les antennes jouent ainsi un rôle de filtre spatial des sources et on se propose d'adapter le traitement du signal de l'antenne pour isoler des sources en mouvement dans le trafic urbain et en extraire un signal propre dont l'analyse permettra d'alimenter ensuite les algorithmes de classification utiles à l'évaluation de la gêne. On peut citer deux études ayant déjà exploré cette capacité de l'antennerie : Weinstein *et al.* [120] ont extrait le signal audio provenant de chaque source acoustique dans une scène sonore complexe. Adcock [5] quant à lui s'est attaché à débruiter des signaux audio à l'aide de la formation de voies. Ces deux études étaient restreintes aux cas des sources fixes.

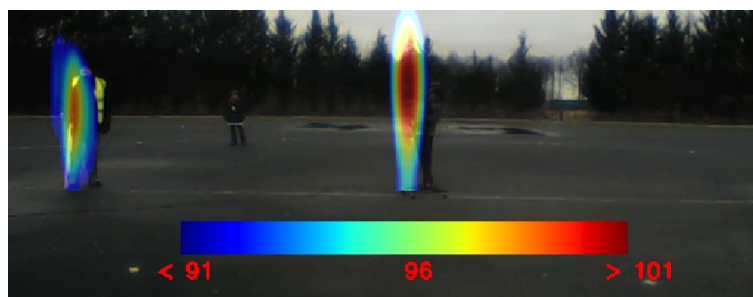
2.4.1 Extraction du signal audio d'une source fixe

Les méthodes citées précédemment [120, 52, 5] utilisent la formation de voies temporelle. Comme indiqué dans l'équation (2.7), les signaux temporels des microphones sont décalés temporellement et sommés. La précision du recalage temporel des signaux dépend de la période d'échantillonnage. Dans le domaine fréquentiel en revanche, le déphasage appliqué à chaque signal est aussi précis que l'estimation de la distance r_{mi} . Par ailleurs, la formation de voies peut conserver une information de phase, ce qui rend possible la reconstruction d'un signal temporel par transformée de Fourier inverse.

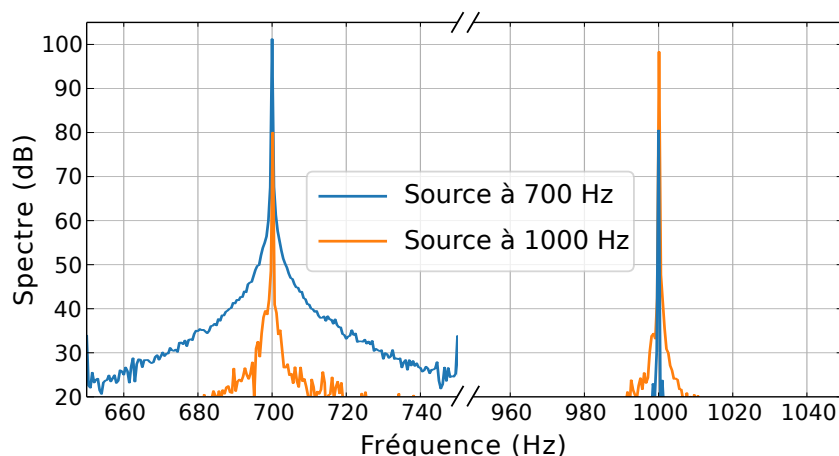
Dans sa formulation fréquentielle, la formation de voies est particulièrement bien adaptée à la parallélisation pour accélérer le calcul. En utilisant des processeurs de carte graphique (ou GPU pour *Graphics Processing Unit*) et un code parallélisé, Charles Vanwynsberghe [111] est ainsi parvenu à réaliser de l'imagerie acoustique large bande en temps réel. Nous avons adapté ses algorithmes pour extraire le signal des sources identifiées dans le trafic urbain.

Evaluation initiale sur sources harmoniques fixes. Cette technique a d'abord été testée sur deux sources fixes avec l'antenne de 256 microphones installée à La Ferté-Vidame (figure 2.11). Les deux haut-parleurs utilisés étaient placés à 7,5 m de l'antenne et écartés de 5 m l'un de l'autre. Les deux émettaient un signal monofréquentiel : l'un à 700 Hz à un niveau de 101 dB et l'autre à 1000 Hz à 98 dB (voir figure 2.23a). La figure 2.23b montre les spectres des deux signaux reconstruits au niveau des deux sources (position des maxima des lobes sur la figure 2.23a). On peut voir que lorsque le signal de la source principale est extrait (700 Hz, courbe bleu), le gain sur la composante à 1000 Hz du spectre est de 19 dB. Le gain est légèrement amélioré quand la formation de voies est faite au niveau de la source à 1000 Hz. En effet, l'atténuation de la composante 700 Hz est alors de 21 dB.

On peut donc espérer, avec un tel système, pouvoir séparer les contributions de deux sources fixes d'environ 20 dB.



(a) Imagerie acoustique sur tout le spectre des sources acoustiques - Espace entre-source : 5 m - Dynamique : 10 dB.



(b) Spectres des signaux reconstruits en deux points de l'espace : sur la source à 700 Hz (courbe bleu) et la source à 1000 Hz (courbe orange).

FIGURE 2.23 – Image et spectre des signaux issus des points de l'image correspondant aux deux sources harmoniques (700 et 1000 Hz) telles que $\Delta L = 3$ dB.

2.4.2 Extraction du signal audio d'un véhicule en mouvement

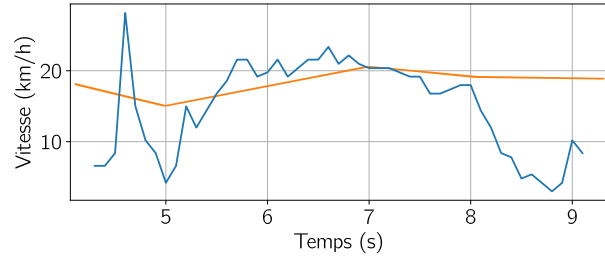
La méthode précédente est étendue aux sources en mouvement en réalisant la formation de voies dans le domaine fréquentiel sur de courtes trames temporelles. Ainsi, la position de la source doit être connue pour chaque trame traitée. Pour cela, un algorithme de suivi d'objet est utilisé. Enfin, le signal total de la source est reconstruit par somme des signaux de chaque trame temporelle.

Identification des trajectoires par détection d'objet. La trajectoire du véhicule est obtenue grâce la bibliothèque de traitement d'image OpenCV qui permet la détection de contour. La procédure de *tracking* s'effectue sur le flux vidéo issu de la caméra grand angle disposée au centre de l'antenne. Sur les images vidéo, on procède pour commencer à la détection des contours des véhicules. Le ou les contours présents sur l'image sont résumés à des rectangles les incluant si leur taille respecte les critères fixés, sinon ils ne sont pas considérés. La connectivité du ou des contours d'une image à ceux présents sur la suivante est simplement défini par un écart maximal des centroïdes des différents contours. Un exemple de la trajectoire du centroïde du contour obtenu est visible en figure 2.24a.

Après avoir converti les coordonnées de l'image en distances réelles, on peut estimer la trajectoire et la vitesse instantanée de la cible. La vitesse estimée du véhicule suivi en figure 2.24a est présentée en figure 2.24b. Pendant cet essai, le véhicule évoluait à 20 km/h. La courbe orange présente l'information délivrée par le radar, elle sert de référence. On peut voir sur la courbe bleue que, sur une plage de 2 s, cette vitesse est assez bien estimée, ce qui n'est pas le cas en début et fin de passage. Cela est dû à l'apparition et la disparition progressive du véhicule sur l'image. Les pics en début et fin de passages sont dus aux ombres.



(a) Détection d'un véhicule (réctangle vert) et obtention de sa trajectoire (ligne rouge).



(b) Vitesse estimée à partir de la trajectoire obtenue (bleu) et information du radar (orange).

FIGURE 2.24 – Exemple de détection de la trajectoire d'une véhicule et de l'obtention de sa vitesse - $V = 20\text{km/h}$.

Synthèse du signal audio par *Overlap-Add*. Afin de reconstruire le signal audio du véhicule en mouvement, nous proposons de s'inspirer d'une méthode utilisée en synthèse de signaux appelée *Overlap-Add*. Comme décrit par Georges et Smith [45], celle-ci consiste à générer un signal $\tilde{s}(t)$ par la somme pondérée avec recouvrement de trames $\tilde{s}^k(t)$:

$$\tilde{s}(t) = \sum_{\kappa=-\infty}^{\infty} w_s(t - \kappa T_s) \tilde{s}^k(t - \kappa T_s), \quad (2.16)$$

avec T_s le temps séparant deux trames de signal, w_s la fenêtre de pondération et $\tilde{s}^k(t)$ des trames de signaux générés par somme de sinusoides. Pour assurer la régularité des transitions entre les trames et ainsi conserver l'enveloppe du signal la fenêtre doit satisfaire la condition :

$$\sum_{\kappa=-\infty}^{\infty} w_s(t - \kappa T_s) = 1. \quad (2.17)$$

Cette méthode est adaptée à notre étude en remplaçant $\tilde{s}^k(t - \kappa T_s)$ dans l'équation (2.16) par $\tilde{s}_i^k(t - \kappa T_s)$: le résultat de la formation de voies temporelle au niveau de la source i . Ainsi pour chaque trame de 50 ms, le signal $s_i^k(t)$ de la source est obtenue par transformée de Fourier inverse du résultat de la formation de voies fréquentielle présentée en équation (2.8).

On se propose d'utiliser une fenêtre composée d'une demi-période d'un sinus au carré sur la longueur de la trame T_f et nulle ailleurs :

$$w_s(t) = \begin{cases} \sin^2\left(\frac{\pi}{T_f} t\right), & \text{si } t \in [0, T_f] \\ 0, & \text{si } t \notin [0, T_f] \end{cases} \quad (2.18)$$

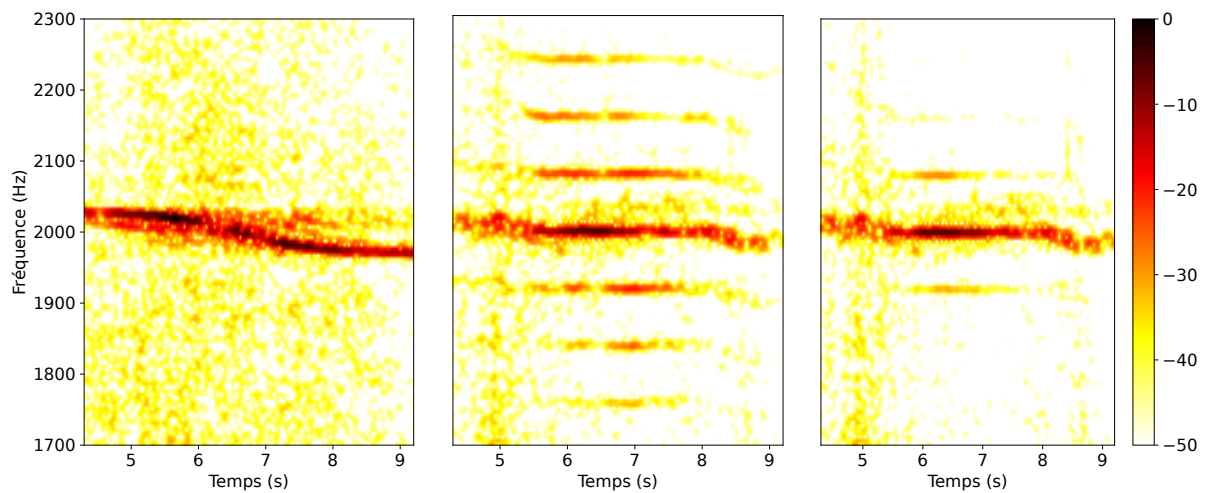
Ainsi, pour la fenêtre $w(t)$ et un recouvrement de 50% (durée de la trame $T_f = 2T_s$), on a

$$\sum_{\kappa=-\infty}^{\infty} w(t - \kappa T_s) = \sin^2\left(\frac{\pi}{T_f} t\right) + \sin^2\left(\frac{\pi}{T_f} (t - T_s)\right) = \sin^2\left(\frac{\pi}{T_f} t\right) + \cos^2\left(\frac{\pi}{T_f} t\right) = 1. \quad (2.19)$$

Pour de meilleurs résultats, le taux de recouvrement sera posé à 75 % ($T_f = 4T_s$), imposant un facteur $\frac{1}{2}$ à appliquer. En effet, en reprenant l'équation (2.19) et sachant qu'alors une durée de trame, on peut ainsi écrire :

$$\sum_{\kappa=-\infty}^{\infty} w(t - \kappa T_s) = \underbrace{\sin^2\left(\frac{\pi}{T_f} t\right) + \sin^2\left(\frac{\pi}{T_f} (t - 2T_s)\right)}_{=1} + \underbrace{\sin^2\left(\frac{\pi}{T_f} (t - T_s)\right) + \sin^2\left(\frac{\pi}{T_f} (t - 3T_s)\right)}_{=1} = 2. \quad (2.20)$$

Expérience de validation - Reconstruction d'un signal harmonique émis par une source en mouvement. Cette méthode est testée sur des données expérimentales de sources en mouvement. Le véhicule présenté en figure 2.24, roulant à 20 km/h, était équipé d'un haut-parleur émettant un signal harmonique à 2 kHz au niveau de la fenêtre avant passager. Le spectrogramme du signal enregistré par un microphone au milieu de l'antenne est présenté en figure 2.25a. On note que ce signal est modulé par l'effet Doppler et qu'un bruit de fond large bande s'y additionne. Remarquons que le spectrogramme est présenté avec une dynamique importante de 50 dB.



(a) Spectrogramme d'un microphone MEMS au centre de l'antenne. (b) Spectrogramme du signal reconstruit - Méthode initiale. (c) Spectrogramme du signal reconstruit - Méthode modifiée.

FIGURE 2.25 – Application de la méthode de reconstruction du signal à une source harmonique se déplaçant à vitesse constante ($V = 20$ km/h) - Dynamique : 50 dB.

La figure 2.25b montre le spectrogramme du signal reconstruit. On peut voir que le niveau du bruit de fond est abaissé, ce qui laisse supposer qu'il était dû aux bruits de roule-

ment. On constate également que l'effet Doppler est compensé. Cette dédopplerisation est peu marquée au début et à la fin du passage, ce qui est normal car la vitesse estimée (présentée en figure 2.24b) n'est pas correcte. On voit aussi apparaître de nombreuses composantes tonales parasites perturbant la reconstruction du signal audio de la source. Ces composantes tonales décalées de 80 Hz correspondent à la période de recouvrement $T_s = 12,5$ ms. Les premières harmoniques de ces composantes ont un niveau 20 dB inférieur au niveau du signal initial. Cela provoque un effet de battement sur le signal audio qui n'est pas acceptable.

Pour y remédier, nous proposons de modifier la pondération en la segmentant en deux temps : (i) pondérer le signal temporel de chaque microphone par une demi-période de sinus (ii) pondérer le signal issu de la formation de voies par cette même fonction.

La figure 2.26 présente la synthèse de la méthode utilisée. Pour une trame temporelle κ le signal de chaque microphone est pondéré par une fenêtre w_1 qui est définie par $w_1 = \sqrt{\frac{w_s}{2}} = \frac{1}{\sqrt{2}} \sin\left(\frac{\pi}{T_f}(t - T_s)\right)$. La formation de voies est ensuite réalisée sur ces trames temporelles, le signal de la source en mouvement est alors fenêtré par $w_2 = \sqrt{\frac{w_s}{2}}$ avant d'être sommé avec les trames précédentes et suivantes en conservant le décalage T_s entre les trames.

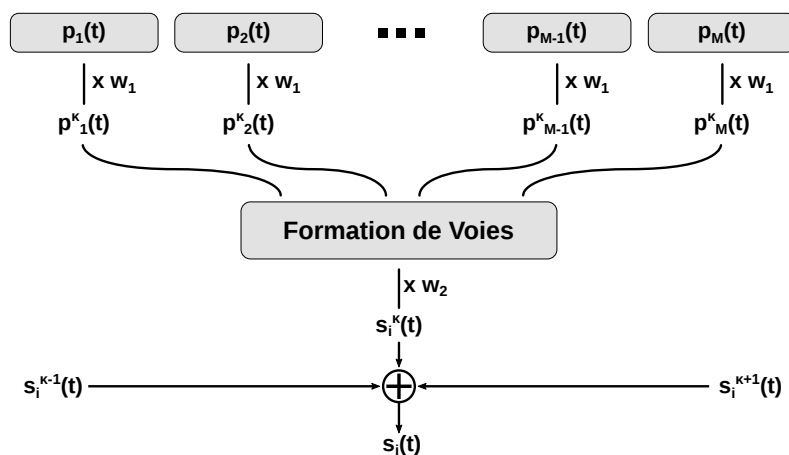


FIGURE 2.26 – Schéma de la méthode de reconstruction du signal de la source en mouvement i comme la somme avec recouvrement de κ trames temporelles issus du signal de M microphones pondérés par les fenêtrés w_1 et w_2 respectivement avant et après l'étape de formation de voies.

Le spectrogramme du signal ainsi reconstruit est présenté en figure 2.25c. On peut noter la diminution des artefacts de reconstruction. Il ne reste plus que deux fréquences constantes à 1920 et 2080 Hz à -30 dB de la source principale. On voit donc que des artefacts persistent dans la reconstruction de l'enveloppe du signal mais la sensation de battement induite reste très faible à l'écoute du fait de la différence d'amplitude.

Expérience de validation - Extraction du signal audio d'un véhicule en mouvement. Deux véhicules passant à 25 km/h sont étudiés ici pour appliquer la méthode : V1-D-2 et DR-Es-1. Les spectrogrammes des signaux initiaux et des signaux extraits le long de la trajectoire du centre des deux véhicules sont présentés en figures 2.27 et 2.28. Comme on peut le voir sur la figure 2.27a, le Picasso émet un bruit avec des composantes tonales à très basse fréquence (50 Hz, 100 Hz) et des harmoniques supérieures qui sont de moins en moins énergétiques

et se noient dans le bruit large bande, principalement dû au contact pneu/chaussée. La figure 2.27b montre le signal reconstruit le long de la trajectoire du centre du Picasso. On peut notamment voir qu'au milieu du passage, donc au centre de l'antenne (où la vitesse est la mieux estimée et la formation de voies la plus sélective) le bruit large bande est bien filtré et seul le bruit moteur est extrait.

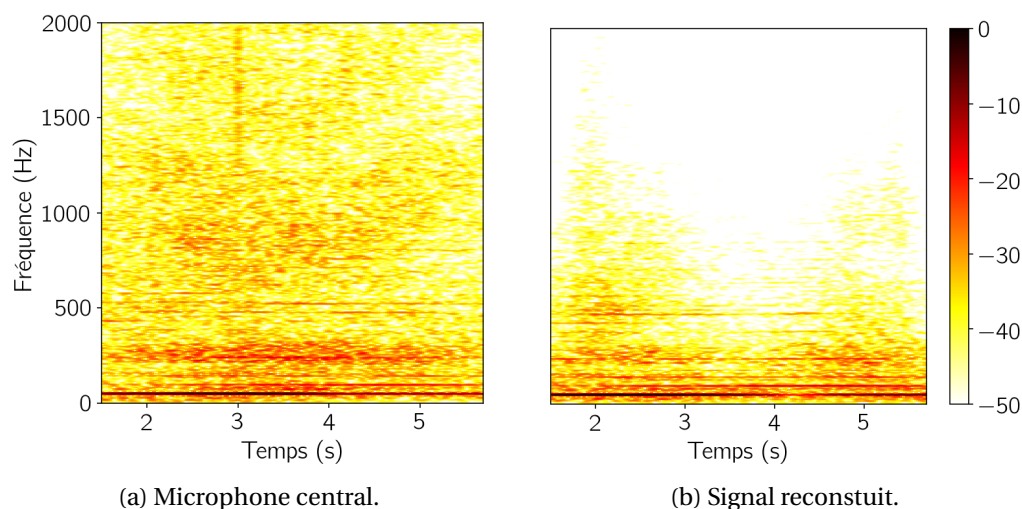


FIGURE 2.27 – Spectrogrammes des signaux du microphone central de l'antenne et ceux reconstruits issus du V1-D-2 - Vitesse : 25 km/h.

La figure 2.28a montre le spectrogramme du deux-roues utilisé pour l'expérience. On peut noter le comportement tonal très marqué de ce signal dont la fréquence fondamentale est à 100 Hz. Le spectrogramme du signal reconstruit au milieu du deux-roues présenté en figure 2.28b montre une baisse importante du niveau des harmoniques et du bruit large bande additionnel pour les fréquences supérieures à 700 Hz. Cela est dû à une trop grande précision de l'antenne en hautes fréquences ce qui empêche de reconstruire le signal audio comprenant toutes les sources de bruit du véhicule. La résolution pourrait être conservée constante en fonction des fréquences en "éteignant des microphones" comme le montre [52].

Expérience de validation - Extraction du signal audio pour plusieurs véhicules. Par mesure de sécurité, nous n'avons pas pu faire rouler plusieurs véhicules en même temps à La Ferté-Vidame. Pour valider la séparation de sources en mouvement en milieu multi-source, on se propose de réaliser une scène sonore synthétique à partir de deux passages de véhicules isolés. Les passages sélectionnés sont à vitesse constante : 25 km/h. On a choisi de traiter les enregistrements du V1-D-2 et du DR-Es-1 car les fichiers de mesures avaient un décalage initial intéressant pour simuler deux véhicules se suivant.

Les deux passages présentés sont utilisés pour créer une scène synthétique où les deux véhicules se suivent. L'équation (2.2) nous a montré que l'on pouvait écrire la pression mesurée par un microphone comme la somme des composantes de chaque source de bruit déphasée de la distance de propagation. Ainsi, considérant chaque véhicule comme une

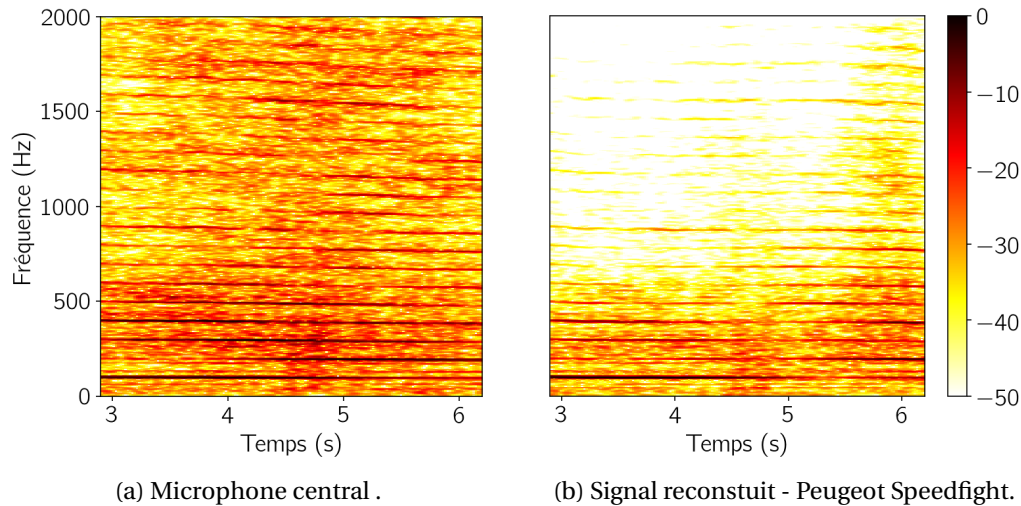


FIGURE 2.28 – Spectrogrammes des signaux du microphone central de l’antenne et ceux reconstruits du DR-Es-1 - Vitesse : 25 km/h.

source, les signaux utilisés dans l’étude ont été créés en sommant les 256 signaux enregistrés pour chaque expérience avec la même origine des temps. La vidéo, nécessaire pour l’obtention de la trajectoire des véhicules, a été réalisée avec le logiciel libre `kdenlive` en faisant simplement l’addition des valeurs pour chaque couleur primaire (rouge, vert et bleu) des images des deux vidéos.

La figure 2.29 montre le résultat de cette addition et les trajectoires obtenues. On constate que la scène est assez cohérente avec des situations réelles dans l’environnement urbain avec environ 5 m d’écart entre les deux véhicules.

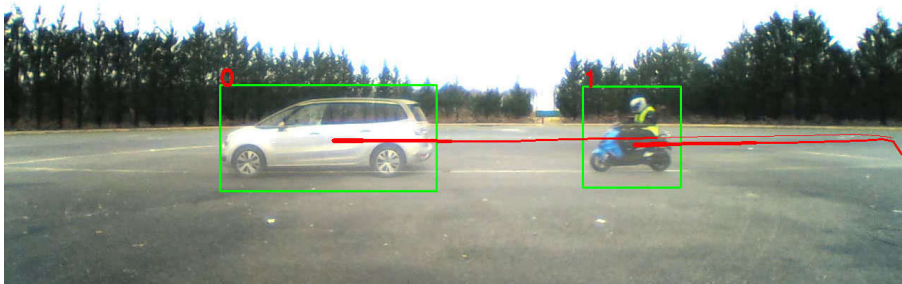


FIGURE 2.29 – Vidéo de la scène synthétique créée avec les trajectoires détectées des véhicules.

La figure 2.30a présente le spectrogramme de la pression synthétisée sur le microphone centrale lors de cette scène sonore virtuelle. On retrouve bien les deux sons initiaux. Les figures 2.30b et 2.30c présentent le résultat de la méthode d’extraction sur respectivement le V1-D-2 et le DR-Es-1. On peut voir que le signal reconstituit au niveau du V1-D-2 (figure 2.30b) contient les composantes fréquentielles du deux-roues mais qu’elles sont diminuées de 20 à 40 dB en fonction de la fréquence. Les composantes basses fréquences du véhicule léger sont bien retrouvées.

La figure 2.30c présente les résultats sur la trajectoire du deux-roues. Outre la bonne reconstruction des composantes tonales, on peut noter l'incapacité de la méthode à filtrer les très basses fréquences. En effet, on retrouve sur ce spectrogramme la composante tonale du Picasso à 50 Hz. Ce qui est cohérent avec l'ouverture angulaire (équation 2.12) à cette fréquence qui est d'environ $2\theta_{3\text{dB}} = 18^\circ$.

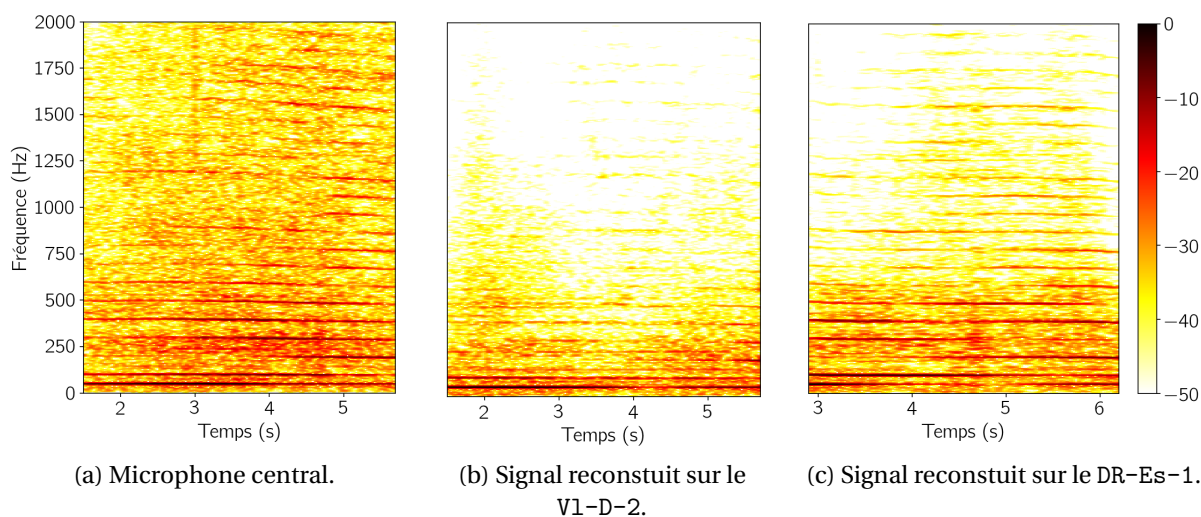


FIGURE 2.30 – Spectrogrammes des signaux initiaux et reconstruits au niveau des sources en mouvement.

Conclusion. Dans cette section, l'efficacité de la séparation des signaux de différentes sources dans une même scène sonore a été montrée. L'opération menée sur sources fixes permet d'évaluer la dynamique du filtrage à près de 20 dB. La technique a ensuite été adaptée pour des sources mobiles et a montré sa capacité à séparer des sons de deux véhicules roulant à vitesse constante et espacés d'environ 5 mètres. On peut donc envisager de l'appliquer en situation réelle dans la perspective d'isoler les signaux caractéristiques des diverses sources présentes dans le trafic urbain.

Conclusion

Les méthodes de séparation de sources et de localisation de sources ne sont pas nouvelles. Elles utilisent des antennes de microphones associées à des méthodes de traitement du signal qui permettent d'imager le paysage sonore. Parmi ces techniques, la formation de voies est la plus simple et la plus couramment utilisée. Elle permet l'estimation de la position de la source et de son spectre à chaque instant. C'est pourquoi notre travail s'est fondé sur cette technique qui présente des limites dont il faut tenir compte, en termes de résolution et de dynamique. L'une et l'autre dépendant essentiellement de l'étendue de l'antenne et du nombre de ses capteurs élémentaires.

Le projet *Megamicros* a été développé dans l'objectif de déployer des grandes antennes de microphone pour des applications diverses. Il nous a permis de disposer de réseau parfaitement adapté à nos objectifs comme ici avec la réalisation d'une expérience unique basée sur une antenne de 256 microphones MEMS avec une envergure de presque 20 m (l'antenne "LFV-256").

Cette antenne a été déployée sur la piste d'essais de PSA Peugeot-Citroën à La Ferté-Vidame pour enregistrer le bruit au passage de différents véhicules à différentes allures. Les sources principales de bruit ont pu être mises en évidence en fonction des allures et du type de véhicule.

Dans une deuxième partie, nous avons montré qu'il était possible d'utiliser la formation de voies pour extraire le signal d'une source de bruit dans une scène sonore complexe. L'application a d'abord été faite sur deux sources fixes, montrant alors une capacité de filtrage de 20 dB. Le cas de plusieurs sources en mouvement a également été traité et a montré l'efficacité à distinguer plusieurs sources mobiles dans une même scène sonore.

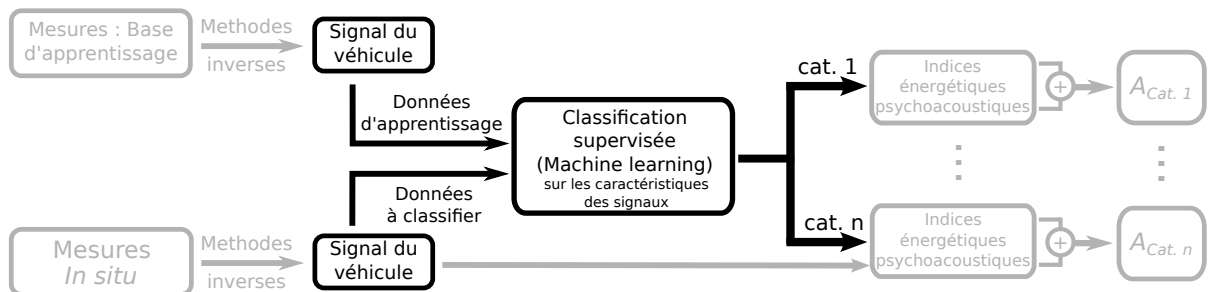
Dans les étapes suivantes du projet, les signaux de l'expérience à La Ferté-Vidame réunis dans une base de données seront utilisés au chapitre 3 comme signaux d'apprentissage pour des algorithmes de *machine learning* pour réaliser la classification des sources utile à la mise en œuvre d'un modèle de gêne acoustique spécifique du trafic automobile urbain, modèle qui fera l'objet du chapitre 4.

Chapitre 3

Classification du trafic routier par apprentissage supervisé

Sommaire

Introduction	69
3.1 Classification de signaux audio	70
3.1.1 Modèles et paramètres pertinents	70
3.2 Classification des véhicules par <i>Support Vector Machines</i>	72
3.2.1 Principe des SVM	72
3.2.2 Application des SVM à la base de donnée de véhicules isolés	74
Conclusion	79



Introduction

Les travaux de Klein *et al.* [59] présentés en section 1.2.2 montrent de grandes variabilités dans le ressenti de la gêne sonore en fonction du type de véhicule et des conditions de roulage. Morel *et al.* [77] ont défini des ensembles de sources sonores regroupées de manière pertinente du point de vue perceptif, définissant alors sept catégories mixant types et allures de véhicules. Leur modèle [78] permet de relier les caractéristiques des signaux à la gêne sonore exprimée par les sujets de leur étude. On rappelle les sept catégories :

- | | |
|---|---|
| Cat. 1 Deux-roues passant à vitesse constante, | Cat. 4 Deux-roues en décélération, |
| Cat. 2 Deux-roues en accélération, | Cat. 5 Bus, véhicules lourds et légers en décélération, |
| Cat. 3 Bus, véhicules lourds et légers passant à vitesse constante, | Cat. 6 Véhicules légers en accélération, |
| | Cat. 7 Bus et véhicules lourds en accélération. |

Dans l'objectif d'appliquer le modèle de gêne sonore de Morel *et al.*, il semble pertinent de mettre en place un outil de classification automatique des sources sonores (véhicules routiers) dans l'environnement urbain. On a vu en partie introductive que les méthodes actuelles d'estimation du trafic routier (tuyaux pneumatiques ou boucles électromagnétiques) ne permettent pas de différencier les deux deux-roues des véhicules légers. C'est pourquoi d'autres systèmes ont commencé à émerger dans les années 1990, comme les outils de classification par traitement d'images issues de caméras installées au-dessus de voies de circulation [60, 61, 103].

De nombreux travaux sur l'apprentissage machine (ou *machine learning* en anglais) ont complété cette approche pour la classification du trafic routier basé sur la vision par ordinateur. Ces approches ne suffisent pas pour appliquer le modèle de Morel *et al.* qui utilise la catégorie du véhicule mais aussi son signal audio pour estimer les indices psychoacoustiques et énergétiques. C'est pour cela que nous cherchons ici à classer le véhicule au passage principalement grâce à son signal audio.

L'objectif de ce chapitre est d'être capable de trouver la catégorie à laquelle appartient un véhicule à partir de la vidéo (donnant sa trajectoire) et du signal audio extrait par la méthode présentée au chapitre précédent.

Dans un premier temps, une revue des méthodes et des paramètres pertinents pour la classification est proposée. Puis la méthode est appliquée sur les mesures faites sur piste avec l'antenne "LFV-256".

3.1 Classification de signaux audio

3.1.1 Modèles et paramètres pertinents

En 1998, Couvreur *et al.* [33] ont proposé une méthode de classification de bruits environnementaux à partir des signaux issus de la base de données MADRAS. Parmi les différents signaux disponibles dans cette base de données, les auteurs proposent de se restreindre à la classification de cinq types de bruit : les véhicules légers, les véhicules lourds, les deux-roues motorisés, les avions et les trains. Pour cela, ils mettent en place un classifieur basé sur un automate de Markov à états cachés (en anglais *hidden Markov model* ou HMM) et l'utilisation de coefficients prédictifs linéaires (en anglais *Linear Predictive Coefficients* ou LPC) pour représenter les caractéristiques du signal.

Rabaoui *et al.* [89] apportent une comparaison de plusieurs descripteurs audio pour deux méthodes différentes de classification : un HMM et une machine à vecteur de support (SVM, pour *Support Vector Machine* en anglais). Ils mettent en avant de bons résultats de classification avec les LPCCs et les MFCC. Les LPCC (pour *Linear Prediction Cepstral Coefficients* en anglais) sont équivalents aux LPC calculés sur le cepstre et non sur le signal audio. Les coefficients cepstraux sur échelle Mel (ou MFCC pour *Mel Frequency Cepstral Coefficients* en anglais) permettent d'obtenir les meilleurs résultats en utilisant les SVM avec un taux de reconnaissance de 1% supérieur aux classifications par HMM atteignant 92,8% de bonne classification.

Le cepstre, introduit au paragraphe précédent, est le résultat de la transformée de Fourier inverse appliquée au logarithme naturel de la valeur absolue de transformée de la Fourier du signal temporel $s(t)$:

$$\mathcal{C}(\tau) = \mathcal{F}^{-1}(\ln(|\mathcal{F}(s(t))|)), \quad (3.1)$$

il s'exprime donc dans un domaine analogue au domaine temporel, le domaine des fréquences. Les MFCC sont des coefficients représentant le contenu spectral. Ils sont estimés classiquement sur des périodes de 25 ms et varient donc dans le temps. Ils sont sans dimension. Le détail du calcul est donné sur la figure 3.1. Comme on peut le voir, le spectre est filtré par un banc de filtres triangulaires répartis suivant l'échelle logarithmique des Mel.

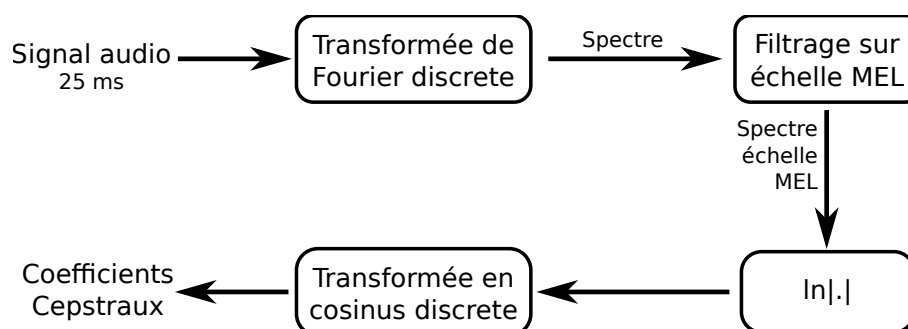


FIGURE 3.1 – Schéma explicatif du calcul des coefficients cepstraux sur échelle Mel (MFCC).

Valero *et al.* [108] proposent de comparer différents types de descripteurs pour quatre méthodes de classification : des arbres de décision (DT, pour *Decision Tree*), la méthode des k plus proches voisins (KNN, pour *K-Nearest Neighbour* en anglais), un réseau de neurones (NN, pour *Neural Network* en anglais) et un modèle de mélange gaussien (GMM, pour *Gaussian Mixture Model* en anglais). Ils montrent que de façon systématique, utiliser un arbre de décision revient à une erreur plus importante que les trois autres méthodes. Ils confirment l'intérêt de l'utilisation des MFCC, qui donnent de meilleurs résultats de classification que les LPCC ou que les descripteurs audio (encodant le timbre, le spectre, etc.) inclus dans le format MPEG-7. L'addition du MPEG-7 et des MFCC augmenterait de 1 ou 2 % le taux de bonne classification pour NN et KNN mais n'améliore pas la classification avec un GMM, qui est constant et maximum à 89,5 %. Valero *et al.* [108] comparent aussi leurs résultats avec des HMM. Ces derniers obtiennent de meilleurs résultats : 92,9 % de bonnes réponses. Or Rabaoui *et al.* [89] comparent des SVM avec des HMM et obtiennent de meilleurs résultats avec les SVM qu'avec les HMM : respectivement 92,8% et 91,7% de résultats exactes.

De plus, Giannoulis *et al.* [46] rapportent les travaux du Challenge DCASE 2013 pendant lequel onze groupes de chercheurs ont comparé l'efficacité de leurs algorithmes de classification de scènes sonores. Ils montrent alors que les SVM ont souvent été les classifieurs les plus performants. Cela est d'autant plus vrai quand le nombre d'échantillons est important. En effet, Justino *et al.* [57] ont montré qu'au-delà de 8 échantillons d'apprentissage, les SVM étaient plus performants que les HMM. Ce qui est expliqué par les auteurs par la capacité accrue des SVM à prendre en compte les variabilités intra-catégories.

Par ailleurs, Lundén *et al.* [64] ont montré une très bonne prédiction des réactions humaines à un environnement sonore grâce aux MFCC comme le caractère plaisant ou mouventé d'un son.

Les études citées précédemment semblent s'accorder sur l'intérêt des MFCC pour des tâches de classification de signaux audio et les SVM semblent être largement utilisés apportant une grande efficacité. Pour toutes ces raisons, nous proposons de mettre en place un outil de classification des bruits routiers utilisant des algorithmes de SVM et les MFCC extraits des signaux issus de la formation de voies.

3.2 Classification des véhicules par machine à vecteurs de support, ou *Support Vector Machines*

Dans cette section, le principe des machines à vecteurs de support comme outil de classification est exposé. Ensuite, cette méthode est appliquée à la base de données de véhicules isolés constituée à La Ferté-Vidame et présentée au chapitre 2. La maîtrise des allures et des types de véhicules sur cette base de données permet de chercher les meilleures performances de classification de la méthode SVM.

3.2.1 Principe des SVM

La figure 3.2 illustre le fonctionnement des SVM avec un exemple d'échantillons répartis dans deux catégories et ayant deux caractéristiques X_1 et X_2 . L'objectif des SVM est de trouver des hyperplans discriminant les n échantillons (dans notre cas le nombre de mesures de bruit de passage) répartis dans deux catégories linéairement séparables suivant les caractéristiques X_1 et X_2 . A chacun des échantillons de chaque catégorie une valeur de y est associé (ici $y = 1$ (points blancs) et $y = -1$ (points noirs)) L'objectif est de trouver l'hyperplan, défini par son vecteur normal \mathbf{w} , qui donne la plus grande marge entre les deux catégories de vecteurs. On note $\mathbf{X} = [\mathbf{x}_0, \dots, \mathbf{x}_n]$ l'ensemble des vecteurs $\mathbf{x}_i = (x_{1i}, x_{2i})$ pour $0 \leq i < n$. Cela revient donc à minimiser la norme de \mathbf{w} :

$$\begin{aligned} \min \|\mathbf{w}\| \\ \text{tel que } y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \\ \text{pour } 0 \leq i < n \end{aligned}$$

On peut généraliser à c clusters (ici $c = 7$: sept catégories perceptives) dont les données ne sont pas linéairement séparables en introduisant alors un paramètre de relaxation de la contrainte λ . \mathbf{x}_i est alors un vecteur de dimension p (le nombre de caractéristiques du signal, comme le nombre de MFCC par exemple). Les SVM sont alors généralisés en

$$\min_{\mathbf{w}} \left[\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i - b)) + \lambda \|\mathbf{w}\|^2 \right]. \quad (3.2)$$

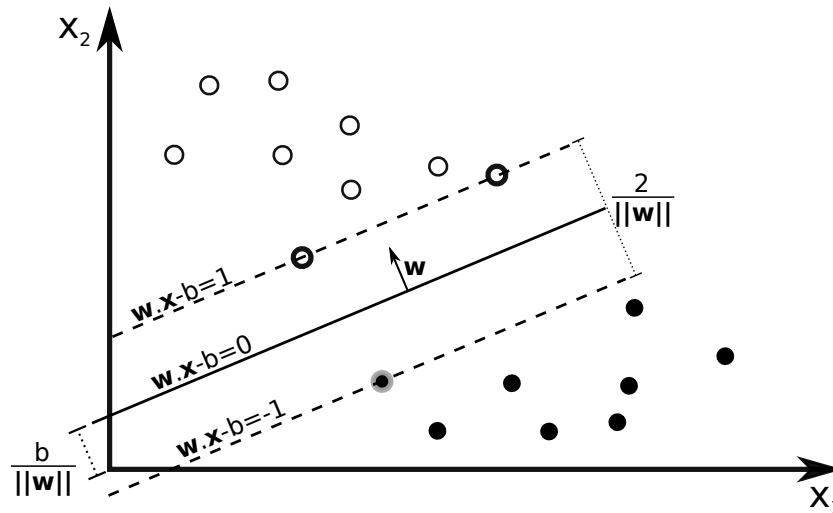


FIGURE 3.2 – Estimation du modèle de classification par SVM par la définition d'un hyperplan séparant deux catégories définies par deux caractéristiques X_1 et X_2 .

La fonction $\max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i - b))$ est égale à 0 si la donnée x_i est du bon côté de la marge $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) = 1$ ou, sinon, proportionnelle à la distance à la marge. On peut alors poser $\zeta_i = \max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b))$ et obtenons la formulation des SVM comme :

$$\min_{\mathbf{w}} C \sum_{i=1}^n \zeta_i + \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.3)$$

où C est une constante qui permet de contrôler le compromis entre le nombre d'erreurs de classement et la largeur de la marge.

L'équation (3.3) peut s'exprimer sous la forme d'un produit scalaire (calcul non présenté ici), la fonction ζ_i met alors en jeu ce qui est appelé un noyau (ou *kernel*) linéaire. Le noyau s'écrit alors $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$, avec $i, j \in [0, n[$. L'intérêt de la formulation en noyau est de pouvoir utiliser un classifieur linéaire pour résoudre un problème non-linéaire. En plus du noyau linéaire, des noyaux usuels employés avec les SVM sont

$$\text{le noyau polynomial : } K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^d \quad (3.4)$$

$$\text{le noyau gaussien : } K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{(2\sigma^2)}\right). \quad (3.5)$$

La figure 3.3 montre un exemple de comparaison de trois noyaux pour une série de données – issue de la base de données Iris incluse dans la librairie python `scikit-learn`¹ – ayant deux catégories qui se recouvrent et une troisième (points beiges) bien séparée. Comme on peut le voir sur cette figure, tous les noyaux sont capables de bien séparer les points beiges des autres. La frontière de décision entre les deux catégories se recouvrant est ensuite ce qui diffère entre les types de noyaux (linéaire, polynomial et gaussien).

1. scikit-learn.org/stable/modules/svm.html

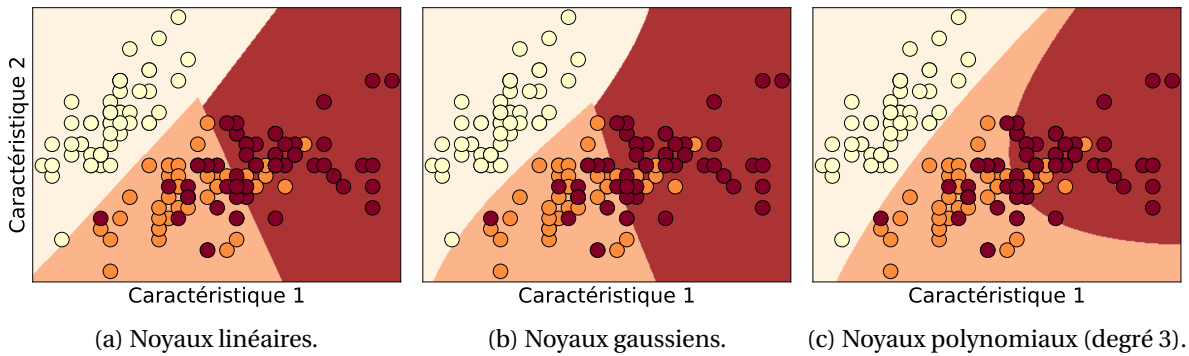


FIGURE 3.3 – Ensemble des valeurs issues de trois catégories représentées par deux caractéristiques ($p = 2$). Les zones de couleurs sont les zones estimées de chaque catégories. Données issues de la base de données Iris de `scikit-learn` – $C = 1$.

3.2.2 Application des SVM à la base de donnée de véhicules isolés

L'objectif de cette section est d'entraîner un classifieur sur les passages de véhicules isolés enregistrés sur la piste d'essais de La Ferté-Vidame, voir section 2.3. Introduite précédemment, la librairie `scikit-learn` sera utilisée pour mettre en place les machines à vecteurs de support (SVM). On se propose d'utiliser les MFCC extraits des signaux comme mode de description des signaux audio traités par le classifieur. Ils sont calculés grâce à la librairie python `python_speech_features`². Les valeurs par défaut sont utilisées, à savoir un nombre de filtres égal à 26 pour une bande fréquentielle allant de 0 à 25 kHz (car la fréquence d'échantillonnage du système d'acquisition est $F_e = 50$ kHz) et les 13 premiers MFCC sont utilisés. Dans la littérature différentes fréquences d'échantillonnage sont utilisées, ce qui fait varier les fréquences centrales des filtres. Dans notre cas, la fréquence centrale du dernier filtre triangulaire est de 3200 Hz, avec une fréquence de coupure de 3800 Hz, ce qui nous paraît cohérent avec le contenu fréquentiel du bruit urbain. Les MFCC sont calculés sur des trames temporelles de 100 ms sans recouvrement. On obtient donc 10 valeurs de MFCC par seconde. Maijala *et al.* [65] ont montré qu'au-delà de 8 MFCC utilisés, le gain sur l'erreur de classification était faible et ont donc trouvé suffisant d'utiliser 13 MFCC pour obtenir de bons résultats de classification (avec un GMM) tout en ayant un temps de calcul raisonnable. De plus, les auteurs obtiennent les meilleurs résultats de classification avec des trames temporelles de 100 ms. Ce qui conforte nos choix de paramètres.

Pour entraîner le classifieur, les véhicules électriques sont retirés du corpus de véhicules. En effet, Morel *et al.* [77] ne proposent pas de catégorie perceptive pour les véhicules électriques. Ainsi, en séparant les situations "feux rouges"³ en deux (pour les régimes de décélération et d'accélération), 68 passages sont conservés.

On se propose de tester, dans un premier temps, la classification sur le signal du microphone de référence pour chacun des passages. La robustesse de la classification sera améliorée via l'utilisation des signaux reconstruits au niveau du véhicule. Enfin, les paramètres

2. github.com/jameslyons/python_speech_features

3. présentées en page 52 : véhicule décélérant en deuxième rapport, stationnant quelques secondes au ralenti moteur au milieu de l'antenne et accélérant avec passage de rapport

permettant une classification optimale seront recherchés.

Signaux du microphone de référence

On se propose de faire les premiers essais sur les signaux du microphone de référence, placé au centre de l'antenne (voir description en section 2.3). La figure 3.4 montre les valeurs moyennées temporellement des treize MFCC pour chacun des passages sélectionnés. On peut remarquer de grande variation de ces valeurs, allant de -20 à 13.

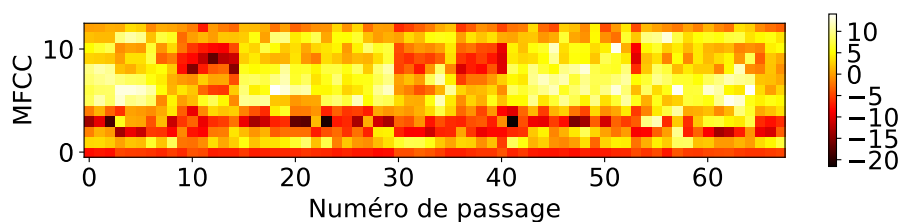


FIGURE 3.4 – Moyennes temporelles des 13 MFCC au cours de chacun des 68 passages retenus.

Un test de classification est réalisé en utilisant des machines à vecteur de support (SVM) et en utilisant les MFCC présentés en figure 3.4 comme données d'entrée. De sorte à quantifier l'erreur commise par le modèle, pour chaque passage nous faisons un test de "un contre tous". C'est-à-dire qu'un passage est retiré de la base d'apprentissage et est utilisé comme donnée de test et cela est répété sur l'ensemble des passages. L'erreur globale est alors estimée en moyennant les erreurs obtenues. Une étude préliminaire n'a montré qu'une faible influence du type noyau (linéaire, polynomial ou gaussien) sur le résultat de la classification pour nos données : c'est donc le noyau linéaire qui sera utilisé dans la suite de ce travail. Le coefficient C est laissé à sa valeur par défaut égale à 1.

Le résultat est présenté en tableau 3.1 et présente une erreur globale (aussi appelée taux d'échec) de 36,8%, ce qui est mauvais au vu de la littérature [89, 46, 108]. La colonne "Nombre" indique le nombre de passages par catégorie. Un tel tableau présente une somme à 100% suivant les lignes, les catégories attendues, aussi appelée vérité terrain (ou *ground truth*). En revanche, en fonction du modèle le pourcentage peut dépasser 100 suivant les colonnes. Le tableau 3.1 montre que certaines catégories comme les véhicules légers à vitesses constantes (Cat. 3) et les deux-roues à vitesses constantes (Cat. 1) obtiennent de bons résultats, respectivement 76,2% et 100% de bonnes classifications. On peut voir que les catégories 1, 3, 5 et 7 sont sur-représentées alors, qu'à l'inverse, les catégories 2, 4 et 6 ne sont pas suffisamment bien classifiées.

Les allures des véhicules ont été rajoutées au dictionnaire d'apprentissage sous forme binaire : chaque allure (constante, accélérée ou décélérée) vaut 0 ou 1. Cela permet de faire baisser le taux d'échec à 32,4%.

On peut améliorer significativement le résultat de classification en normalisant les MFCC. En effet, en connaissant l'information d'allure et en normalisant les treize MFCC par leurs

		Nombre	Estimée						
			Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	8	100	0	0	0	0	0	0
	Cat.2	9	22,2	66,7	0	11,1	0	0	0
	Cat.3	21	0	0	76,2	0	9,5	9,5	4,8
	Cat.4	5	0	0	0	60	20	0	20
	Cat.5	9	0	0	11,1	22,2	33,4	11,1	22,2
	Cat.6	12	0	0	25	0	25	41,7	8,3
	Cat.7	4	0	0	0	0	50	0	50

TABLE 3.1 – Matrice de confusion - Pourcentage de passages classifiés en fonction de la catégorie attendue et nombre de passage par catégorie - C = 1 - Erreur globale : 36,8%.

		Estimée						
		Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	100	0	0	0	0	0	0
	Cat.2	0	88,9	0	0	0	11,1	0
	Cat.3	0	0	100	0	0	0	0
	Cat.4	0	0	0	60	40	0	0
	Cat.5	0	0	0	0	100	0	0
	Cat.6	0	0	0	0	0	91,7	8,3
	Cat.7	0	0	0	0	0	50	50

TABLE 3.2 – Matrice de confusion - Pourcentage de passages classifiés en fonction de la catégorie attendue - C = 1 - MFCC normalisés et allures inclus à la base d'apprentissage - Erreur globale : 8,8%.

maximums sur les 68 passages on obtient alors 8,8% d'erreur et la matrice de confusion présentée en tableau 3.2. On peut voir qu'il reste quelques erreurs mais uniquement entre des allures équivalentes pour un coefficient C inchangé.

Nous avons remarqué, par ailleurs, que certains coefficients, et notamment le treizième MFCC, ne faisait que dégrader les performances de classification. En effet, en le supprimant on obtient 7,4% d'erreur et on supprime l'ambiguïté entre les catégories 2 et 6. Ainsi, le dictionnaire d'apprentissage se résumera aux 12 premiers MFCC normalisés et complétés par les trois informations binaires d'allure.

Signaux reconstruits au niveau des véhicules

Les signaux reconstruits au niveau des véhicules sont maintenant utilisés pour la classification. La figure 3.5 montre une augmentation de l'amplitude de variation des MFCC passant de 33 à 54. Cela est peut-être dû à la moyenne temporelle qui est ici réalisée sur un temps plus court (uniquement quand la voiture est visible par la webcam, soit quelques secondes) pendant lequel l'énergie acoustique est plus importante.

Un test de classification avec les données brutes, sans les normaliser et sans prendre en compte les allures, donne un taux d'échec de 27,9%, ce qui est mieux qu'avec le signal de référence (l'erreur était de 36,8%). L'ajout des allures et de la normalisation des MFCC par leurs valeurs maximums sur les 68 passages fait descendre à 14,7% d'erreur, qui est le double

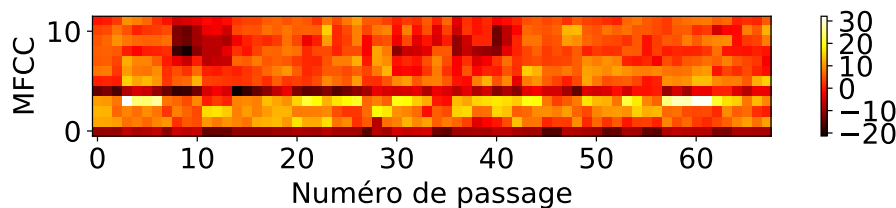


FIGURE 3.5 – Valeurs moyennes des douze premiers MFCC sur les 68 passages retenus - Signaux reconstruits au niveau du véhicule.

par rapport au cas précédent. Le tableau 3.3 récapitule les résultats de classification pour les différents cas testés jusqu'ici. On peut aussi y voir que la normalisation seule des données ne permet pas d'améliorer l'erreur mais que c'est bien dans les deux cas l'action cumulée de la normalisation et de l'ajout des allures qui permet un réel gain.

	Données Brutes	Normalisation	Allures	Normalisation + Allure
Signaux du microphone de référence	36,8%	37%	32,4%	8,8%
Signaux reconstruits au niveau du véhicule	27,9%	48,5%	23,5%	14,7%

TABLE 3.3 – Tableau récapitulatif des différents cas étudiés et des taux d'erreur associés

Influence du paramètre C

Pour améliorer les résultats de classification on se propose de regarder l'influence du paramètre C sur les données issues des signaux reconstruits au niveau des véhicules en mouvement. Il semble y avoir un optimum pour $C = 2,5$ pour lequel on obtient un taux d'échec de 4,41% et la matrice de confusion présentée en tableau 3.4. On peut voir ici qu'il y a encore des confusions de catégories pour des allures équivalentes : catégories 1 et 3, 2 et 6 et 4 et 5 respectivement à vitesses constantes, en accélération et en décélération.

	Estimée						
	Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Cat.1	87,5	0	12,5	0	0	0	0
Cat.2	0	88,89	0	0	0	11,11	0
Cat.3	0	0	100	0	0	0	0
Cat.4	0	0	0	80	20	0	0
Cat.5	0	0	0	0	100	0	0
Cat.6	0	0	0	0	0	100	0
Cat.7	0	0	0	0	0	0	100

TABLE 3.4 – Matrice de confusion - Pourcentage de passages classifiés en fonction de la catégorie attendue - $C = 2,5$ - MFCC normalisés et allures incluent à la base d'apprentissage - Signal reconstruit - Erreur globale : 4,41%.

Pondération du dictionnaire d'apprentissage

La pondération du dictionnaire d'apprentissage a été investiguée. En effet, nous avons remarqué que l'on pouvait améliorer encore les résultats de classification grâce à cela. En cherchant manuellement, nous avons pu mettre en évidence un certain nombre de vecteurs de pondération permettant d'obtenir 100% de bonnes prédictions, l'un d'eux est :

$$\mathbf{W} = \underbrace{[2, 1, 0, 4, 2, 1, 1, 0, 1, 1, 1, 1]}_{\text{MFCC}}, \underbrace{[1, 1, 1]}_{\text{Allures}}. \quad (3.6)$$

On peut voir que certains MFCC ne participent pas au modèle ici, à savoir les n° 3 et 8. Il en est de même pour d'autres vecteurs de pondération obtenus pour lesquels les MFCC n°3, 6, 8 et 11 peuvent même, dans certains cas, dégrader la classification. On se propose alors de réaliser une étude paramétrique afin de déterminer une pondération optimale des MFCC.

Une étude paramétrique a ainsi été menée pour identifier les vecteurs de pondération associés aux MFCC qui donnent les meilleurs résultats. La classification est testée pour chaque combinaison de MFCC en affectant à chacun d'eux un coefficient entre 0 (coefficient ignoré) et 5 pour 8 MFCC (les douze premiers auxquels ont été retirés les MFCC n°3, 6, 8 et 11). Ainsi, 1 679 616 combinaisons sont possibles, et 26 301 permettent d'obtenir un résultat de 100% de bonne classification sur ces données de véhicules isolés. Il pourra être intéressant dans le futur de compléter cette étude paramétrique en testant les combinaisons volontairement ignorées ici. En l'état, le temps de calcul dépasserait une semaine sur nos moyens de calcul (ce qui n'a pas été fait).

La figure 3.6 montre le dictionnaire d'apprentissage associé à ce vecteur de pondération \mathbf{W} . Les lignes numéro 12, 13 et 14 représentent, toujours de façon binaire, respectivement les vitesses stabilisées, les passages en accélération et ceux en décélération.

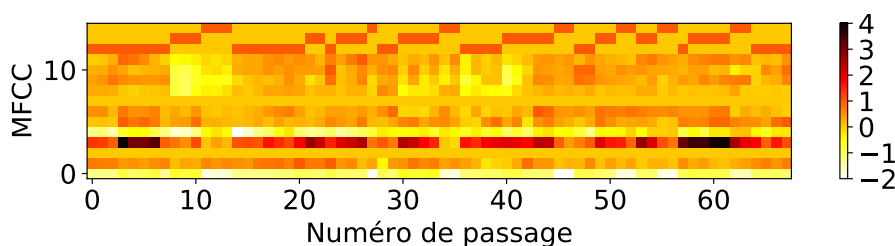


FIGURE 3.6 – Valeurs moyennes des douze MFCC et des trois allures sur les 68 passages retenus - Signaux reconstruits - Exemple de dictionnaire d'apprentissage pondéré par le vecteur \mathbf{W} permettant une classification sans erreur.

Ces tests permettent de valider la procédure d'apprentissage et de reconnaissance. Le nombre élevé de combinaisons de MFCC permettant une bonne classification est un élément positif qui apportera de la robustesse pour l'application dans un environnement urbain moins contrôlé. En effet, ces différents vecteurs de pondérations seront utilisés (en section 5.2.2) pour obtenir les meilleurs résultats de classification pour un corpus de signaux issus du milieu urbain.

Conclusion

Dans ce chapitre, il a été montré que dans des tâches de classification supervisée, le choix des caractéristiques représentant les signaux était important pour obtenir de bons résultats. Les coefficients cepstraux sur échelle Mel (MFCC) ont montré dans la littérature une très bonne efficacité dans cette tâche de classification pour des sons urbains comme pour des morceaux de musique. Mais en plus des caractéristiques des signaux, la littérature a montré que les algorithmes utilisés ont une grande influence sur les performances du classifieur. Chaque algorithme a ses spécificités mais les Machines à Vecteurs de Support (SVM) sont largement utilisées dans la classification des signaux audio.

Nous proposons donc de classifier les signaux à partir des MFCC (calculables à partir des signaux audio préalablement obtenus). La tâche de classification est faite à partir de Machines à Vecteur de Support (SVM). Cette méthode a été appliquée, dans un premier temps, sur les signaux issus du microphone de référence de la base de données acquise à La Ferté-Vidame et présentée en section 2.3. Il a été décidé d'utiliser les signaux des véhicules au passage obtenus grâce à la formation de voies sur la trajectoire du véhicule. En effet, lors de l'adaptation future de la méthode au milieu urbain le signal d'un microphone ne suffira pas à la classification car il enregistrera une scène sonore complexe composée de plusieurs source de bruit.

Ainsi suite à l'ajout de l'allure (extraite de la trajectoire du véhicule obtenue par traitement d'image de la caméra centrale) au dictionnaire ainsi qu'à la normalisation des MFCC, un résultat de 95,59% de bonne classification a été atteint. Enfin, une étude paramétrique a mis en évidence qu'une pondération des MFCC peut améliorer la classification jusqu'à atteindre les 100% de bonnes réponses et ce pour au moins 26 301 vecteurs de pondérations différents, soit 1,6% des vecteurs testés lors de l'étude paramétrique.

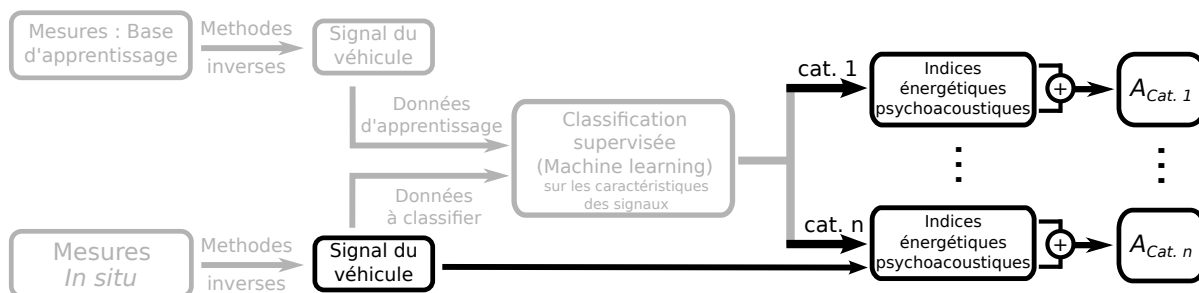
Cet ensemble de combinaisons de MFCC permettant une bonne classification sera utilisé pour chercher les meilleurs résultats de classification lors de l'application du système au milieu urbain.

Chapitre 4

Implémentation du modèle de gêne sonore

Sommaire

Introduction	81
4.1 Modèles de calcul d'indices psychoacoustiques	82
4.1.1 Calcul de la sonie	82
4.1.2 Calcul de la rugosité	84
4.1.3 Calcul de la force de fluctuation	88
4.2 Implémentation du modèle de Morel <i>et al.</i> et application à une mesure de référence	92
Conclusion	95



Introduction

Dans l'objectif final de notre travail de fournir un outil de mesure de la gêne associée au trafic urbain, la dernière étape consiste à évaluer cette gêne sur la base des signaux sources obtenus par *tracking* et formation de voies et des catégories perceptives associées identifiées comme cela a été décrit aux chapitres précédents. Pour réaliser cette tâche, parmi d'autres modèles existant, nous avons choisi de mettre en oeuvre celui de Morel *et al.* [78]. Ce modèle présente l'avantage de proposer un modèle de gêne sonore pour des catégories perceptivement cohérentes. Selon ce modèle, la gêne s'évalue différemment pour chacune des catégories perceptives proposées, comme une combinaison linéaire d'indices psychoacoustiques mesurés directement sur le signal audio de la source considérée, voir tableau 1.6.

La sonie, la rugosité et la force de fluctuation sont les trois indices psychoacoustiques à calculer pour mettre en place ce modèle. Ils font l'objet eux même de multiples modèles qui sont présentés et discutés dans un premier temps. Leurs performances sont analysées et des adaptations sont proposées le cas échéant. Dans une deuxième partie, le modèle de Morel *et al.* est appliqué aux passages de véhicules isolés observés sur la piste d'essai de La Ferté-Vidame. Finalement les résultats obtenus sont discutés dans la perspective du chapitre suivant qui applique la méthode à une situation de trafic réel.

4.1 Modèles de calcul d'indices psychoacoustiques

Les indices psychoacoustiques comme la sonie, la rugosité ou encore l'acuité sont des outils de description d'un son. Leur normalisation a débuté dans les années 1960 avec l'introduction de la recommandation ISO R 532 (1966) proposant deux méthodes de calcul de la sonie, l'une basée sur le modèle de Zwicker [121] et l'autre sur le modèle de Stevens [102]. Cette recommandation est devenue une norme en 1975 (ISO 532-1975).

Les autres indices psychoacoustiques n'ont fait l'objet d'aucune norme (nationale ou internationale). Les différentes méthodes d'estimation de la rugosité et de la force de fluctuation, notamment, ne produisent généralement pas des résultats semblables. Ils sont, de plus, intégrés à des logiciels propriétaires (Müller BBM¹, Head Acoustics² ou encore Brüel et Kjær³) dont les algorithmes ne sont pas accessibles, qui peuvent ne plus être distribués et qui sont difficilement interfaçables avec d'autres langages de programmation. C'est le cas du logiciel dBsonic de 01dB utilisé par Morel *et al.* [78].

Cette partie présente les modèles des indices psychoacoustiques utilisés et adaptés pour notre étude. Ils sont conçus pour permettre l'évaluation de la gêne perçue suivant les formules du tableau 1.6. Ces modèles sont validés par comparaison avec ceux de la littérature lorsqu'ils existent.

4.1.1 Calcul de la sonie

Pour leurs modèles, Morel *et al.* [78] ont utilisé la sonie calculée par le logiciel dBsonic selon la norme DIN 45631. De nombreuses autres implémentations sont disponibles en ligne dans différents langages (C, BASIC ou Matlab). Mais rares sont celles qui proposent d'estimer la sonie en fonction du temps.

La bibliothèque LoudnessToolbox développée par Genesis⁴ propose des codes mettant en œuvre deux modèles de sonie de sons instationnaire sous Matlab. Le premier est celui de Zwicker et Fastl (1999) [122]; le deuxième est celui de Glasberg et Moore (2002) [49]. Cette librairie propose aussi un modèle suivant la norme DIN 45631 (ou ISO 532B) qui ne propose

1. www.muellerbbm-vas.fr/produits/analyse-de-donnees/qualite-sonore/

2. www.head-acoustics.de/downloads/fr/Application_Notes/ArtemiS/Analyses_Psychoacoustiques_II_f.pdf

3. bksv.com/-/media/literature/Product-Data/bp1589.ashx

4. genesis-acoustics.com/sonie_en_ligne-32.html

qu'une sonie moyenne pour un signal audio. Les résultats de ce modèle sont comparés à celui instationnaire de Zwicker et Fastl.

Les résultats présentés figure 4.1 sont des courbes d'isophonie obtenues pour deux implémentations différentes. La sonie est calculée pour des signaux harmoniques dont la fréquence varie suivant une échelle logarithmique de 20 à 20 000 Hz (100 valeurs) et le niveau de -10 à 110 dB (60 valeurs).

La figure 4.1a montre le résultat suivant la norme DIN 45631 (ou ISO 532B). En comparant ces résultats avec les courbes d'isophonie telles que définies dans la norme ISO 226 présentées sur les courbes en pointillés oranges, on constate la simplification en hautes fréquences adoptée par cette norme et une sous-estimation de la sonie en basses fréquences. Cependant on peut noter un très bon accord global avec la référence, notamment à 1 kHz où le niveau sonore (en dB) équivaut au niveau de sonie (en phone).

La figure 4.1b montre le résultat suivant le modèle de sonie instationnaire de Zwicker et Fastl (1999) [122]. On peut voir un très bon accord avec les résultats suivant la norme DIN 45631 (figure 4.1a). Les résultats valident ce modèle de sonie instationnaire pour des signaux harmoniques. Il sera utilisé par la suite pour estimer la sonie des signaux extraits du trafic routier.

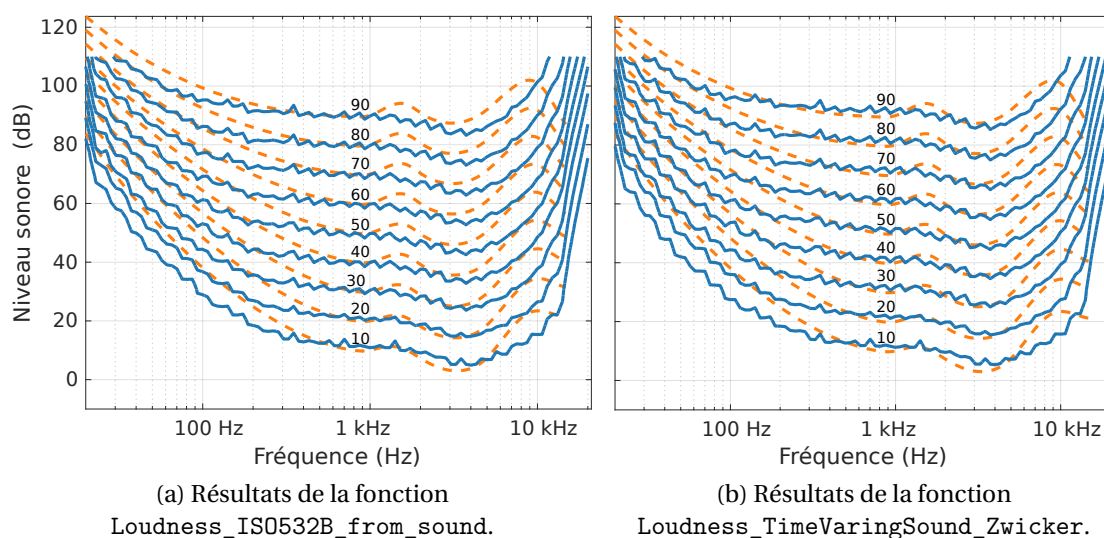


FIGURE 4.1 – Courbes d'isophonie (en phones) en bleu issus de la librairie `LoudnessToolbox` de Genesis. Valeurs de la norme ISO 226 en orange.

Il est à noter qu'en 2017, une nouvelle version de la norme ISO 532 a été publiée. L'ISO 532-1:2017 propose un modèle de sonie instationnaire basé sur le modèle de Zwicker et reposant sur la norme allemande DIN 45631/A1:2010. L'ISO 532-1:2017 propose également d'utiliser le modèle de sonie de sons stationnaires de Moore et Glasberg et non plus celui de Stevens qui était présent dans la version de 1975, il était peu utilisé. Bien que des implémentations de cette norme commencent à être disponibles, elle ne sera pas utilisée pour rester en accord avec le modèle de Morel *et al.*

4.1.2 Calcul de la rugosité

La rugosité, qui traduit la sensation de “grésillement” issue d’une fluctuation d’amplitude en haute fréquence, n’as pas été standardisé. Mais des modèles ont été proposés pour estimer cette sensation, telle que décrite par Zwicker et Fastl [38]. En 1985, Aures [12] a proposé un modèle pour estimer la rugosité. Il utilise l’enveloppe du signal filtré sur chaque bande critique i pour estimer le degré de modulation généralisé m_i^* , ainsi que la corrélation des enveloppes entre bandes critique pour calculer la rugosité spécifique dans chaque bande critiques. C’est ce modèle qui était mis en oeuvre dans dBsonic et utilisé par Morel *et al.* [78].

Lors de sa thèse, en 1993, Rolan Sottek [100] a cherché à étendre ces modèles en proposant un “modèle auditif” qui permet de modéliser la sonie, la rugosité et la force de fluctuation. Ce modèle [101] a été implémenté dans la suite logicielle Artémis de Head Acoustics⁵.

En 1997, Daniel et Weber [34] proposent un modèle de rugosité dont le schéma est présenté en figure 4.2. Il repose sur le modèle d’Aures [12] en y apportant quelques améliorations. Les deux principaux changements sont le passage à 47 filtres (qui ont toujours 1 Bark de largeur, mais sont espacés cette fois de 0,5 Bark, il y a donc un recouvrement à 50% des filtres) et l’estimation de la rugosité spécifique par la pondération du degré de modulation généralisé m_i^* par le produit des deux coefficients de corrélation avec les bandes adjacentes, et non par la moyenne des deux coefficients de corrélation dans le modèle d’Aures. Dans ce modèle également, ce produit est pondéré par un filtre dépendant de la bande critique (noté $g(z_i)$).

Une implémentation de ce modèle est disponible librement⁶. Le code a été mis à disposition par Rodrigo Garcia Leon [43] à l’issue de ses travaux à l’Université Technique d’Eindhoven (TUE) en 2015. L’objectif du travail était de développer un modèle de force de fluctuation basé sur le modèle de rugosité de Daniel et Weber [34]. On a pu voir en sections 1.1.8 et 1.1.9 que ces deux sensations sonores sont issues de la modulation de l’enveloppe du signal à des gammes de fréquences différentes. Utiliser un même cœur de modèle pour estimer ces deux sensations paraît cohérent.

La rugosité R est calculée, comme indiqué en figure 4.2, par la somme des rugosités spécifiques r_i tel que

$$R = \left| c_R \sum_{i=1}^{47} r_i \right|, \quad \text{avec} \quad r_i = (g(z_i) \cdot m_i^* \cdot (k_{i-2} \cdot k_i))^2 \quad (4.1)$$

où $c_R = 0,25$ est un coefficient de normalisation pour que le modèle produise une rugosité de 1 asper pour un signal harmonique à 1 kHz modulé à 100 % à 70 Hz et ayant un niveau équivalent de 60 dB et k_i, k_{i-2} sont les coefficients de corrélation entre les degrés de modulation m_i^* des bandes critiques supérieures (m_{i+2}^*) et inférieures (m_{i-2}^*). Dans cette implémentation, une adaptation est proposée en vue du calcul futur de la force de fluctuation.

5. www.head-acoustics.de/downloads/fr/Application_Notes/ArtemiS/Analyses_Psychoacoustiques_II_f.pdf

6. github.com/rodrigo-garcia-leon/MATLAB_shared/tree/master/Psychoacoustics/Roughness

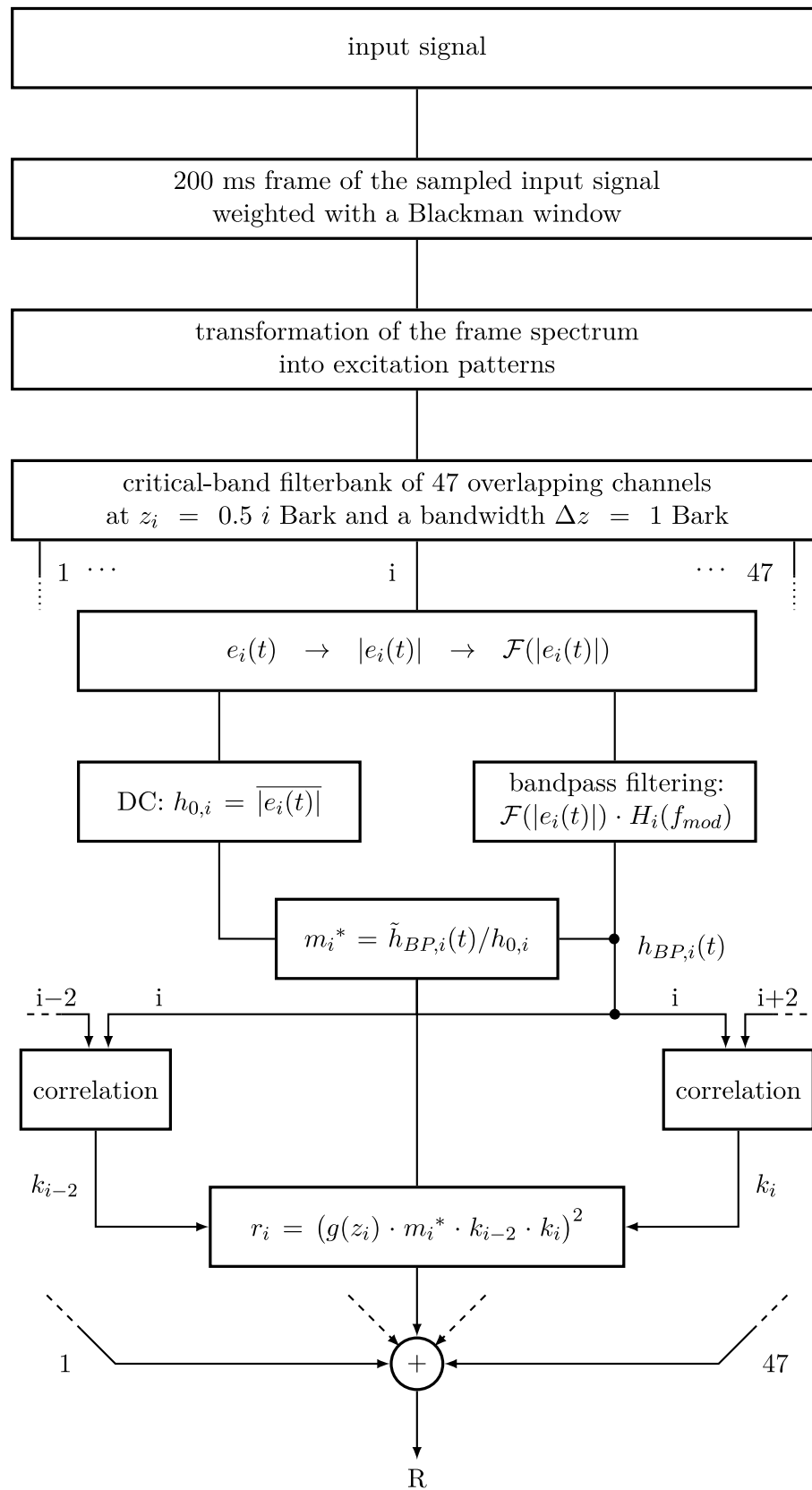


FIGURE 4.2 – Schéma de principe du modèle de rugosité de Daniel et Weber [34] - D'après [43].

Elle consiste à pondérer chaque élément permettant le calcul des r_i en introduisant p_g , p_m et p_k tels que :

$$r_i = g(z_i)^{p_g} \cdot m_i^{* p_m} \cdot (k_{i-2} \cdot k_i)^{p_k}. \quad (4.2)$$

En se plaçant dans les conditions du modèle de Daniel et Weber, en posant $p_g = p_m = p_k = 2$, on obtient les résultats présentés en figure 4.3. Les résultats sont comparés aux courbes de références de Zwicker et Fastl [38] pour différentes fréquences centrales (de 125 Hz à 1 kHz sur la figure 4.3a, de 1 à 8 kHz sur la figure 4.3b) et différents degrés de modulations m . Les signaux utilisés pour tester cette implémentation sont obtenus par

$$s(t) = \sin(2\pi f_c t) \cdot (1 - m \cos(2\pi f_{mod} t)), \quad (4.3)$$

avec f_c la fréquence centrale du signal, f_{mod} la fréquence de modulation et m le degré de modulation. Ce signal est ensuite normalisé à 60 dB pour se placer dans le cadre des courbes de références.

Les figures 4.3a et 4.3b comparent nos résultats aux courbes de référence (en tirets). On constate une allure semblable mais la normalisation n'est pas bonne. C'est aussi le cas du calcul de la rugosité en fonction du degré de modulation m (figure 4.3c). On peut noter alors une sous-estimation de la rugosité sur toute la plage de valeur.

Nous proposons une adaptation de ce modèle de sorte à obtenir des résultats plus proches des courbes de références. La rugosité a alors été cherchée sous la forme

$$R = \alpha \left| c_R \sum_{i=1}^{47} r_i \right|^{\beta}, \quad \text{avec} \quad r_i = g(z_i)^{p_g} \cdot m_i^{p_m} \cdot (k_{i-2} \cdot k_i)^{p_k}, \quad (4.4)$$

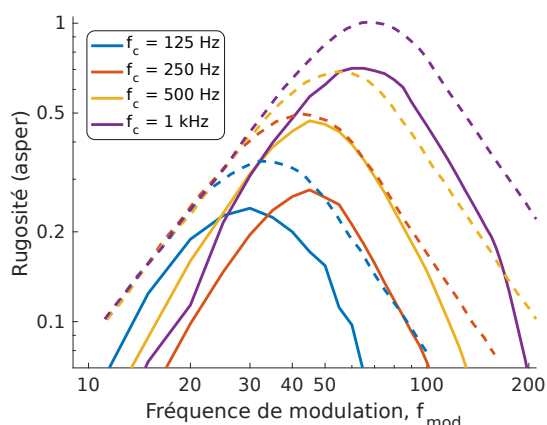
où α et β sont deux paramètres additionnels assurant un meilleur accord. Une minimisation au sens des moindres carrés entre les courbes issues du modèle de Garcia-Léon et les courbes de référence est mise en place grâce à l'algorithme de Levenberg-Marquardt [67, 75]. Il permet une optimisation rapide malgré un grand nombre de paramètres (ici cinq).

La fonction coût minimisée prend en compte l'erreur moyenne au sens des moindres carrés aux courbes de référence pour une évolution de f_c et f_{mod} (figures 4.3a et 4.3b) ainsi que celle en fonction du degré de modulation m (figure 4.3c) en ajoutant une contrainte supplémentaire sur les valeurs de rugosité maximale pour chaque fréquence centrale f_c et sur la valeur de rugosité pour $m = 1$.

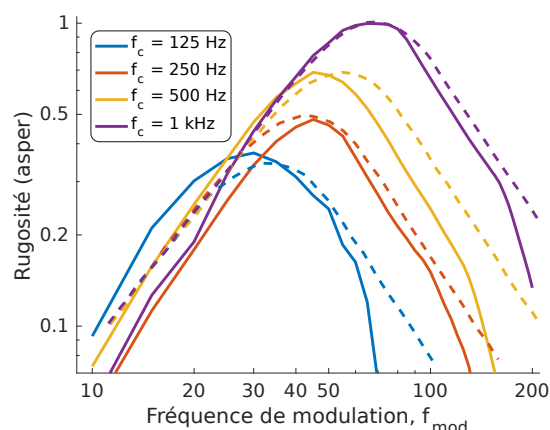
La figure 4.4 présente les résultats du modèle de Garcia-Léon modifié avec les paramètres suivant issus de l'algorithme d'optimisation :

$$p_g = 3,49, \quad p_m = 3,32, \quad p_k = 3,49, \quad \alpha = 1,47 \& \quad \beta = 0,55. \quad (4.5)$$

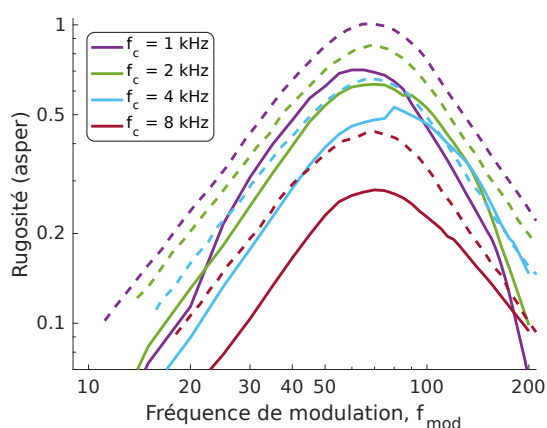
Pour certaines fréquences centrales basses ($f_c = 125$ et 500 Hz, voir figure 4.4a) le maximum de rugosité est obtenu pour une fréquence plus basse que la référence. En revanche on peut voir un assez bon accord dans les pentes des courbes et dans les maxima de rugosité en



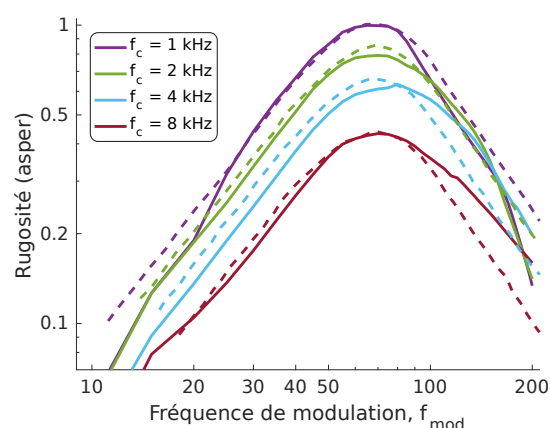
(a) Evolution de la rugosité en fonction de la fréquence de modulation et de la fréquence centrale du signal – De 125 Hz à 1 kHz - $m = 1$.



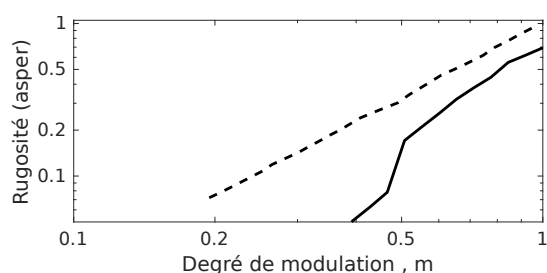
(a) Evolution de la rugosité en fonction de la fréquence de modulation et de la fréquence centrale du signal – De 125 Hz à 1 kHz - $m = 1$.



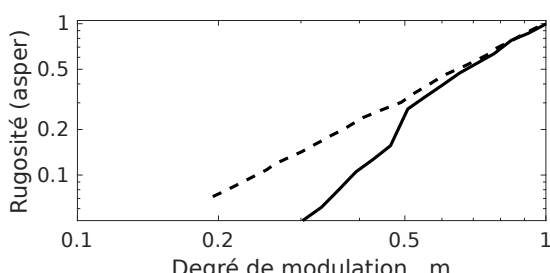
(b) Evolution de la rugosité en fonction de la fréquence de modulation et de la fréquence centrale du signal – De 1 à 8 kHz - $m = 1$.



(b) Evolution de la rugosité en fonction de la fréquence de modulation et de la fréquence centrale du signal – De 1 à 8 kHz - $m = 1$.



(c) Evolution de la rugosité en fonction du degré de modulation - $f_{mod} = 70$ Hz et $f_c = 1$ kHz.



(c) Evolution de la rugosité en fonction du degré de modulation - $f_{mod} = 70$ Hz et $f_c = 1$ kHz.

FIGURE 4.3 – Comparaison de la rugosité calculée par le modèle de Garcia-Léon (—) avec les valeurs de Fastl et Zwicker [38] (---).

FIGURE 4.4 – Comparaison de la rugosité calculée par le modèle de Garcia-Léon modifié (—) avec les valeurs de Fastl et Zwicker [38] (---).

fonction de la fréquence de modulation. L'évolution en fonction du degré de modulation est aussi améliorée. Toutefois, la rugosité est toujours sous-estimée pour un degré de modulation inférieur à 0,5.

On peut noter aussi la disponibilité d'une implémentation donnant aussi d'assez bons résultats dans le logiciel d'analyse sonore PsySound développé en Matlab et implémentant le modèle d'Aures. Cette implémentation ne sera pas retenue car son interfaçage avec le reste du projet (codé dans le langage Python) est plus délicat. C'est bien l'implémentation de Garcia-Léon modifiée avec nos paramètres qui sera utilisée dans la suite de nos travaux.

4.1.3 Calcul de la force de fluctuation

On rappelle que la sensation de force de fluctuation est issue, comme la rugosité, d'une modulation de l'enveloppe du son, mais dans un régime de plus basses fréquences. Cette sensation est maximale à 4 Hz d'après Zwicker et Fastl [38].

Les modèles pour estimer la sensation de force de fluctuation sont plus rares que pour la rugosité, peu formalisés (par exemple dans des mémoires de Master [99, 43]) ou non disponibles librement comme c'est le cas dans les suites logicielles citées précédemment.

Alejandro Osses Vecchi *et al.* [80] ont proposé un modèle et une implémentation disponible librement sur GitHub⁷. Il est issu d'une amélioration du travail de stage de Rodrigo García León et se base sur une adaptation du modèle de rugosité de Daniel et Weber [34] vu à la section précédente. Cette amélioration était nécessaire au vu des écarts importants entre le modèle et les courbes de référence à l'issue du travail de Master.

Dans leur article, Osses Vecchi *et al.* [80] comparent leur modèle à des courbes représentant les moyennes des résultats expérimentaux de Zwicker et Fastl [38, p. 248]. On peut voir un très bon accord de leur modèle par rapport aux courbes de référence pour les deux types de signaux (harmoniques ou larges bandes) en figure 4.5 malgré une légère surestimation de la force de fluctuation pour les bruits large bande modulés en amplitude (figure 4.5b). Il est important de noter que Osses Vecchi *et al.* modulent leurs signaux à 100% ($m = 1$) or les courbes de référence de Zwicker et Fastl sont obtenues pour une amplitude de modulation de 40 dB, soit $m = 1 - 10^{-40/20} = 0,99$.

Basé sur le modèle de rugosité de Daniel et Weber [34], les auteurs proposent de calculer la force de fluctuation F comme la somme des forces de fluctuation spécifiques à chacun des 47 filtres auditifs espacés de 0,5 Bark tel que

$$F = c_{FS} \sum_{i=1}^{47} f_i, \quad \text{avec,} \quad f_i = g(z_i)^{P_g} \cdot m_i^{P_m} \cdot (k_{i-2} \cdot k_i)^{P_k}, \quad (4.6)$$

avec $c_{FS} = 0,249$. La principale différence avec le modèle de rugosité vient de la variable

7. github.com/aosses-tue/mb/tree/master/FluctuationStrength_TUe

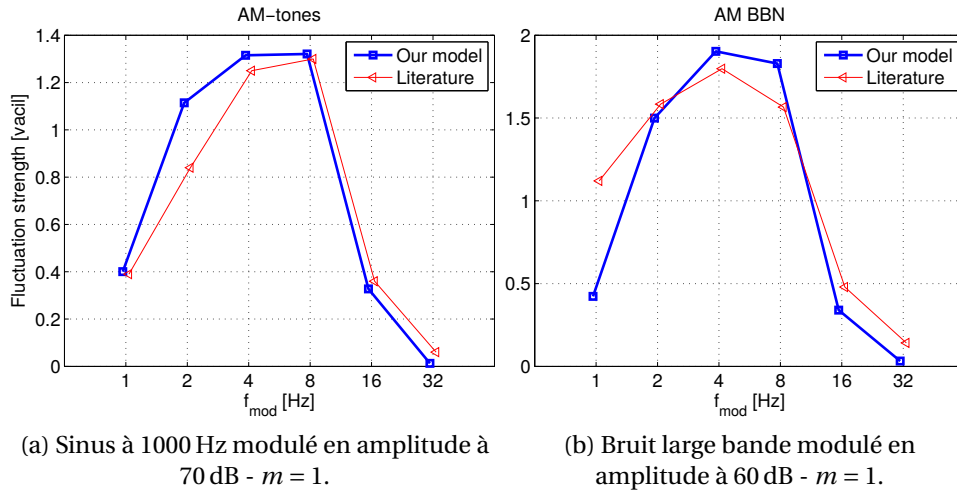


FIGURE 4.5 – Evolution de la force de fluctuation en fonction de la fréquence de modulation - La courbe bleue présente les résultats du modèle de Osses Vecchi *et al.*, la courbe rouge représente la moyennes des expériences présentées dans Zwicker et Fastl [38] - De Osses Vecchi *et al.* [80].

de pondération fréquentielle $g(z_i)$ et des valeurs des exposants p_g , p_m et p_k . La figure 4.6 montre la valeur de g pour chacun des filtres auditifs et pour les deux modèles. On peut voir que, pour le modèle de force de fluctuation (courbe bleue), il n'y a pas de pondération en dessous de $i = 27$ (donc en dessous de 14 Barks) et que celle-ci passe de 1 à 0,5 progressivement de 14 à 24 Barks. En revanche la pondération pour le modèle de rugosité est plus évoluée (courbe orange) et montre un maximum autour du 18^{ème} filtre auditif.

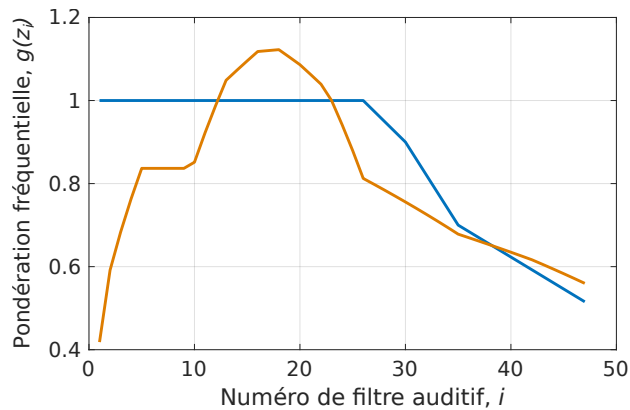


FIGURE 4.6 – Pondération fréquentielle $g(z_i)$ en fonction du numéro de filtre auditif i utilisé dans le modèle de rugosité de García León (en orange) et de Osses Vecchi *et al.* (en bleu).

Les auteurs indiquent que les exposants utilisés sont ici $p_g = 1$ et $p_m = p_k = 1,7$. Nous avons cherché à retrouver leurs résultats. Les courbes obtenues sont présentées en figure 4.7. Les signaux utilisés sont ceux décrits dans Zwicker et Fastl : les figures 4.7a et 4.7b sont obtenues pour des modulations d'amplitude de 40 dB. On peut voir un assez bon accord entre nos résultats et ceux présentés en figure 4.5a pour les signaux harmoniques. On note une légère sur-estimation dans notre cas. En revanche, en figure 4.7b, on voit un écart très important entre nos résultats et ceux publiés par Osses Vecchi *et al.* présentés en figure 4.5b. L'évolution de la force de fluctuation en fonction de la modulation du signal est présentée

en figure 4.7c. Les résultats montrent un assez bon accord avec la courbe de référence pour les signaux harmoniques modulés en amplitude.

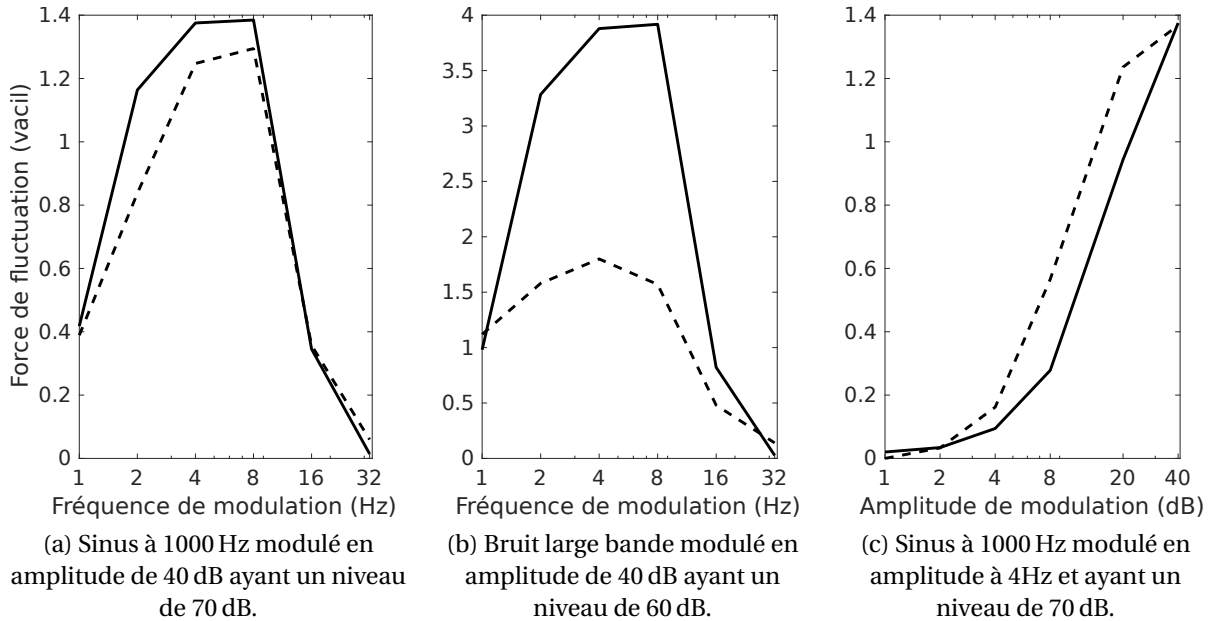


FIGURE 4.7 – Résultats du modèle de Osses Vecchi *et al.* [80](—) avec les valeurs de Fastl et Zwicker [38, p. 248] (--) [80].

Nous avons contacté les auteurs du modèle pour comprendre ensemble d'où provenaient les écarts importants entre les courbes publiées et celle que nous avons obtenues. Cet échange a mis en évidence une différence dans la génération des signaux au niveau de modulation d'amplitude. En effet, la figure 4.8 compare nos signaux théoriquement identiques : signal large bande modulé à 100% à 4 Hz et normalisé à 60 dB. Notre signal est présenté en orange et celui utilisé par Osses Vecchi *et al.* en bleu. On peut noter une différence de l'amplitude maximale mais surtout une forme de modulation différente. En effet, leur signal semble être modulé par la valeur absolue d'un sinus et non par un sinus. Pour vérifier notre hypothèse nous avons testé le modèle avec un signal de cette forme :

$$s(t) = \text{randn}(N_t) \cdot \left((1 - m) + m \left| \cos\left(2\pi \frac{f_{mod}}{2} t\right) \right| \right), \quad (4.7)$$

où $\text{randn}(N_t)$ génère un bruit blanc gaussien du nombre de points temporel N_t .

On obtient alors des résultats de rugosité sensiblement diminués qui ont la même allure que ceux présentés en figure 4.5b mais avec toujours une légère sur-estimation. Nous obtenons alors une force de fluctuation maximale à 4 Hz égale à 2,05 vacil, la référence étant à 1,8 vacil. On peut en déduire que le modèle a été construit sur la base de signaux ne respectant pas le cadre de Zwicker et Fastl, ce qui nous a été confirmé par les auteurs.

Une collaboration a été proposée pour améliorer le modèle et tenter de corriger cette erreur pour les sons large bande. En attendant, nous avons cherché, comme pour le modèle de rugosité, à optimiser les trois paramètres du modèle ainsi qu'un facteur de puissance α et

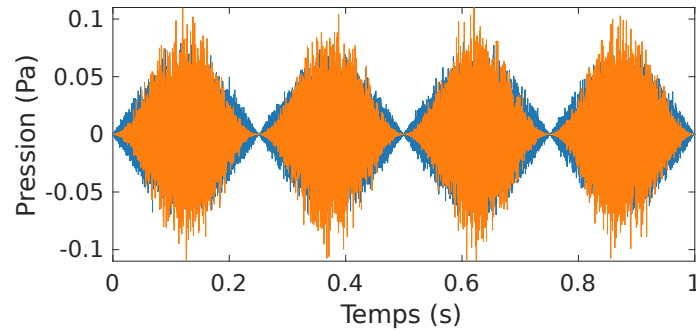


FIGURE 4.8 – Bruit large bande à 60 dB modulé à 100% en amplitude à une fréquence $f_{mod} = 4\text{ Hz}$ - Comparaison des signaux utilisés par Osses Vecchi *et al.* (en bleu) et ceux que nous synthétisons (en orange).

un facteur de multiplication β . La force de fluctuation a donc été formulée sous la forme :

$$F = \left(c_{FS} \sum_{i=1}^{47} f_i \right)^\beta, \quad \text{avec,} \quad f_i = g_{zi}^{p_g} \cdot m_i^{p_m} \cdot (k_{i-2} \cdot k_i)^{p_k}. \quad (4.8)$$

Le même algorithme de minimisation de l'erreur au sens des moindres carrés a été mis en place et des paramètres satisfaisants ont été trouvés pour des valeurs très proches de celles trouvées pour optimiser le modèle de rugosité, à savoir :

$$p_g = 3,5, \quad p_m = 3,3, \quad p_k = 3,5, \quad \alpha = 1,6 \quad \& \quad \beta = 0,3. \quad (4.9)$$

Les résultats du modèle pour ce jeu de paramètres sont présentés en figure 4.9. On distingue toujours une sur-estimation globale de la force de fluctuation pour les signaux harmoniques (figure 4.9a). En revanche, la figure 4.9b montre un très bon accord pour les signaux large bande où le seul écart aux valeurs de référence est à 8 Hz avec une sur-estimation de 0,2 vacil. La figure 4.9c montre une évolution en fonction de l'amplitude de modulation dont l'allure coïncide mieux avec la courbe de référence. Mais cela est couplé à une sur-estimation de cette sensation auditive pour des modulations de faible amplitude. Cela est aussi le cas pour la valeur de référence : un signal harmonique à 1 kHz modulé à 100% à 4 Hz ayant un niveau sonore de 60 dB est sensé faire 1 vacil, nous obtenons ici 1,36 vacil.

Pour réaliser ce calcul dans de bonnes conditions, le modèle nécessite deux secondes de signal et permet d'avoir des valeurs toutes les 200 ms avec un recouvrement à 90% des signaux.

Ce modèle modifié nous permet donc d'estimer la force de fluctuation avec une bonne précision. Il sera utilisé dans la suite de ce travail et intégré au calcul de l'estimation de la gêne sonore.

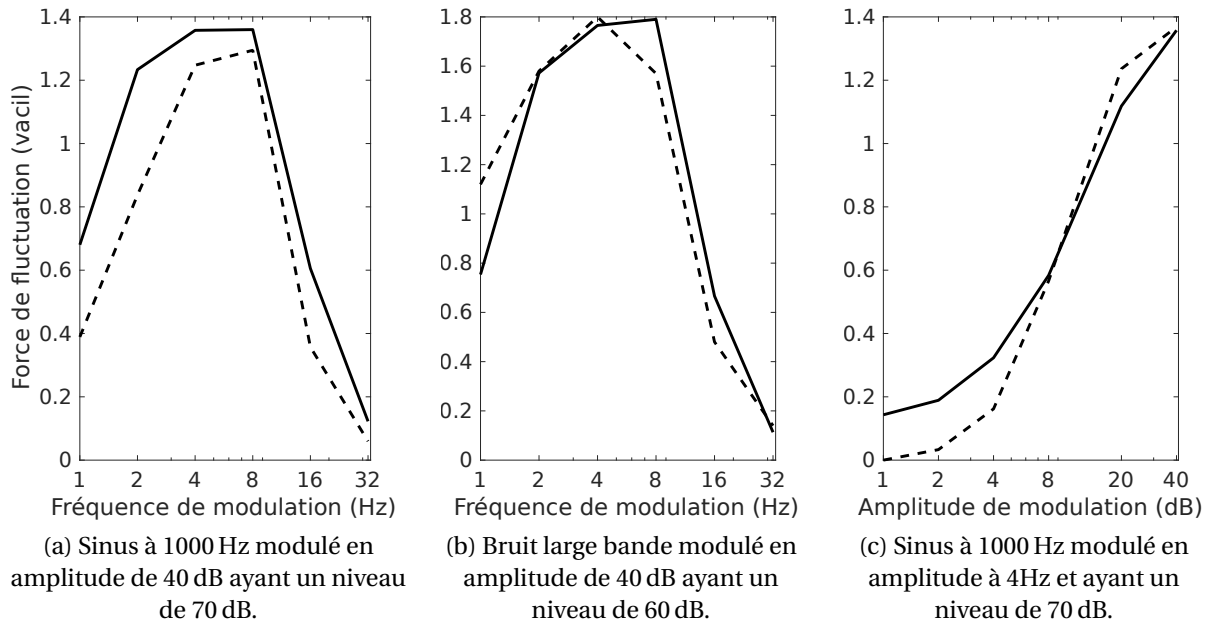


FIGURE 4.9 – Comparaison des résultats du modèle de Osses Vecchi *et al.* [80] modifié (—) avec les valeurs de Fastl et Zwicker [38, p. 248] (–) [80].

4.2 Implémentation du modèle de Morel *et al.* et application à une mesure de référence

Dans cette section, la gêne sonore spécifique est estimée sur les véhicules isolés lors des essais réalisés sur la piste d’essais de PSA à La Ferté-Vidame.

Comme présenté en section 2.4.2, les véhicules au passage sont suivis sur la vidéo, la formation de voies extrait le signal de la source le long de sa trajectoire et les indices psychoacoustiques sont calculés grâce aux modèles présentés dans la section 4.1.

La gêne sonore peut alors être estimée grâce au modèle de Morel *et al.* dont les équations ont été présentées dans le tableau 1.6 en section 1.2.2 et dépendent des indices rappelés ci-dessous qu’il est maintenant possible d’estimer :

- N : la moyenne temporelle de la sonie,
- N_{15-18} : la moyenne temporelle des sonies spécifiques intégrée entre 15 et 18 Barks,
- R_{max} : la rugosité maximale,
- F : la moyenne temporelle de la force de fluctuation,
- F_{max} : la force de fluctuation maximale,
- ΔN^- : le taux de diminution de la sonie dans le temps après le passage du véhicule,
- L_{MF} : le niveau sonore pondéré A calculé pour les bandes de tiers d’octave de fréquences moyennes (entre 315 et 1250 Hz).

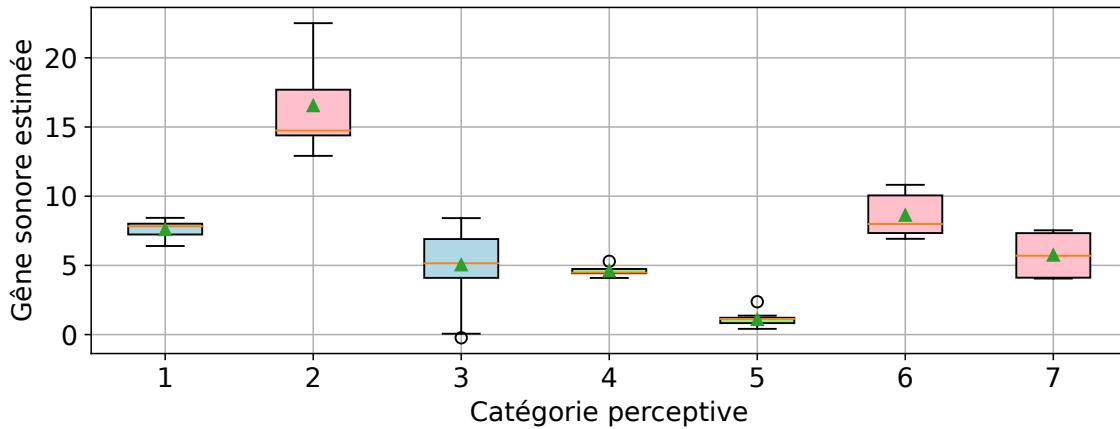
On rappelle que les catégories perceptives sont définies comme :

- Cat. 1 Deux-roues passant à vitesse constante,
- Cat. 2 Deux-roues en accélération,
- Cat. 3 Bus, véhicules lourds et légers passant à vitesse constante,
- Cat. 4 Deux-roues en décélération,
- Cat. 5 Bus, véhicules lourds et légers en décélération,
- Cat. 6 Véhicules légers en accélération,
- Cat. 7 Bus et véhicules lourds en accélération.

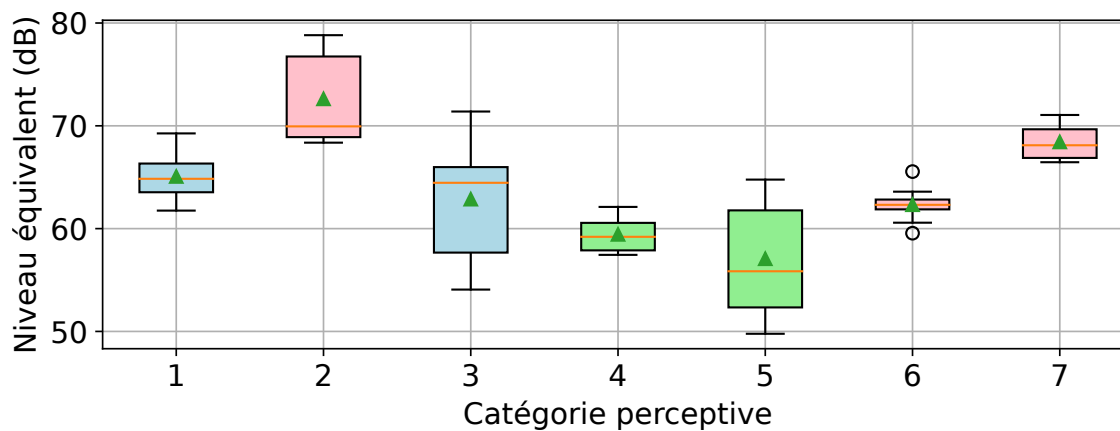
La figure 4.10a présente les résultats d'estimation réalisée sur les 68 passages utilisables en fonction de la catégorie perceptive du véhicule sous forme de boîtes à moustaches. La boîte représente l'étendu de 50% des valeurs, les moustaches représentent l'étendue des deux autres quartiles, la ligne orange est la médiane. La valeur moyenne de la gêne est représentée par le triangle vert et sera notée A_{mean} . On peut voir que les valeurs obtenues dépassent la borne maximale théorique de 10. Au regard de la valeur maximale du niveau équivalent utilisée dans l'étude de Morel *et al.* [78] ($L_{A,eq} = 62$ dB(A)), cela ne semble pas aberrant. En effet, les valeurs de gêne sonore élevées (autour de 22) de la catégorie 2 correspondent aux signaux dont le niveau sonore est mesuré entre 76 et 78 dB et la sonie (ainsi que les sonies partielles) est donc aussi augmentée. La sonie ou le niveau sonore étant utilisés pour le calcul de la gêne sonore pour toutes les catégories, on peut raisonnablement comparer nos résultats entre eux.

La figure 4.10a montre que les gênes ressenties comme les plus faibles sont celles issues des catégories 4 et 5, donc les véhicules en décélération, les véhicules légers et lourds étant moins gênants ($A_{mean} = 0,74$) que les deux-roues ($A_{mean} = 4,5$). Si l'on compare les résultats pour les allures stabilisées (catégories 1 et 3), on peut voir une valeur moyenne plus élevée pour les deux-roues (cat. 1, $A_{mean} = 7,5$) avec un écart type très faible, alors que les voitures et véhicules lourds présentent des valeurs de 0 à 8 avec $A_{mean} = 5$. La comparaison des véhicules en accélération (catégories 2, 6 et 7) montre une prédominance importante des deux-roues dont la valeur moyenne de gêne est de 16 (catégorie 2); suivent ensuite les véhicules légers et les véhicules lourds. On peut enfin noter que la gêne estimée pour les véhicules légers en accélération (cat.6) est plus importante que celle des deux-roues à vitesses constantes (cat.1).

La figure 4.10b présente les niveaux équivalents L_{eq} mesurés pour les mêmes essais, aussi présentés sous forme de boîtes à moustaches par catégorie perceptive. Par comparaison avec la figure 4.10a, on peut voir des écarts dans les tendances. On peut voir notamment qu'un niveau sonore important pour un véhicule lourd en accélération (catégorie 7) induit une gêne sonore relativement faible. De plus, alors que l'écart-type des valeurs de gêne sonore sont très faibles pour les véhicule en décélération sur la figure 4.10a, on peut voir ici une grande variabilité des niveaux équivalents mesurés, notamment pour la catégorie 5 (véhicules légers et lourds en décélération). On peut noter que ces résultats en niveaux équivalents, présentés par catégorie perceptive, constituent une avancée intéressante pour l'analyse fine du trafic routier.



(a) Gènes sonores estimées.



(b) Niveau sonore équivalent.

FIGURE 4.10 – Valeurs des gènes sonores estimées et du niveau équivalent représentés sous forme de boîtes à moustaches par catégories perceptives. La boîte représente l’étendu de 50% des valeurs, les moustaches représentent l’étendue des deux autres quartiles, la ligne orange est la médiane et le triangle vert la moyenne. Les ronds sont les valeurs exceptionnelles situées au-delà des “moustaches”. Les boîtes bleues correspondent aux vitesses stabilisées, les boîtes rouges aux catégories d’accélération et les boîtes vertes aux décélérations.

Ces résultats de gènes sonores estimées montrent des évolutions cohérentes avec la littérature [78, 22, 83]. Cependant, il semble nécessaire de procéder par la suite à un ajustement du modèle de gène sonore avec nos signaux pour obtenir des valeurs de gène s’échelonnant entre 0 (pas du tout gênant) et 10 (extrêmement gênant).

Par ailleurs, le modèle de Morel *et al.* appliqué ici montre un intérêt dans l’estimation de la gène sonore. En effet, il ne propose pas les mêmes tendances que le L_{eq} et réduit la variabilité des gènes estimées. Cela tend donc à penser qu’il explique une part plus importante de la gène sonore que le niveau équivalent.

Conclusion

Ce chapitre a présenté les modèles et implémentations nécessaires à la mise en place du modèle de Morel *et al.* [78]. Le logiciel utilisé par les auteurs (dBsonic) n'étant plus distribué par 01dB il a fallu trouver des alternatives. Pour chacun des indices psychoacoustiques utilisés un modèle issu de bibliothèques libres a pu être adapté et utilisé. Pour calculer la rugosité et la force de fluctuation, une implémentation du modèle de Daniel et Weber [34] a été développée à l'université d'Eindhoven[43]. Ces modèles ont été ajustés pour suivre au mieux les courbes de références de Zwicker et Fastl [38].

L'intégralité de notre outil d'estimation de la gêne sonore a pu être mis en place. Il part d'un suivi de cible qui donne sa trajectoire. On extrait le signal audio par formation de voies le long de cette trajectoire et on calcule les indices psychoacoustiques utiles à l'estimation de la gêne sonore associée à chaque véhicule. La méthode a été appliquée aux mesures de bruits au passage de véhicules isolés réalisés avec l'antenne "LFV-256".

Les résultats obtenus montrent qu'il est possible d'estimer la gêne de véhicules en mouvement grâce à l'utilisation conjointe de notre méthode de classification de sons urbains et d'un modèle de gêne sophistiqué. Ainsi l'évolution de la gêne sonore est cohérente par rapport à la catégorie perceptive du véhicule.

Ces résultats mettent aussi en avant des pistes d'amélioration futures. Par exemple, le poids lourd considéré dans cette étude est le plus gros utilitaire qu'il est possible de conduire avec le permis B, mais n'est pas représentatif de l'ensemble des poids lourds que l'on peut rencontrer dans un milieu urbain.

Par ailleurs, les dépassements importants de la valeur théorique maximum de 10 pour l'estimation des gênes spécifiques à chaque véhicule montrent un besoin d'enrichir le modèle proposé par Morel *et al.* Cela nécessite une nouvelle étude perceptive à mener dans les mêmes conditions expérimentales pour évaluer la gêne associée à nos stimuli. Le modèle pourrait alors être ajusté grâce à ces résultats. Un travail pourra aussi être réalisé pour améliorer les modèles de calcul de la rugosité et de la force de fluctuation.

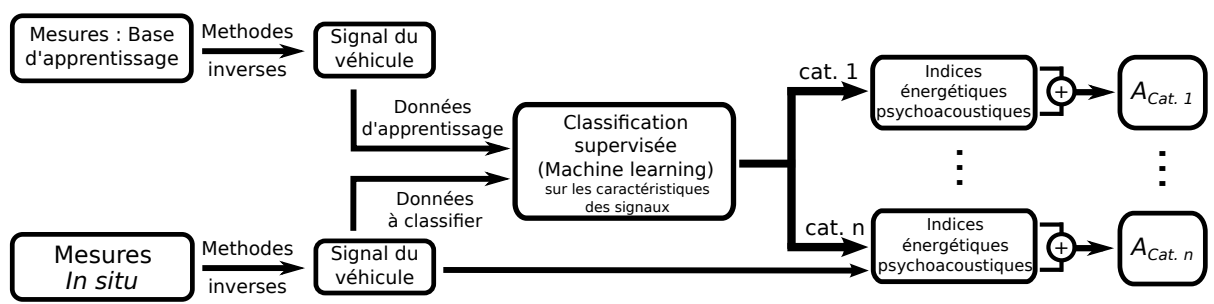
Cet outil de classification a des limites qui ont été présentées. Cependant elles sont restreintes et permettent d'envisager une application au milieu urbain. Cela sera fait dans le chapitre suivant.

Chapitre 5

Estimation *in situ* de la gêne spécifique de chaque véhicule du trafic routier

Sommaire

Introduction	97
5.1 Extraction du signal audio d'un véhicule dans la ville	98
5.1.1 Dispositif expérimental	98
5.1.2 Du pixel à la position 3D	99
5.1.3 Séparation de sources en milieu urbain	100
5.2 Classification <i>in situ</i>	103
5.2.1 Classification avec apprentissage sur véhicules isolés	104
5.2.2 Classification avec apprentissage sur données acquises dans la ville . .	105
5.2.3 Influence du noyau du classifieur	105
5.3 Estimation de la gêne sonore <i>in situ</i>	110
Conclusion	112



Introduction

Précédemment, la méthode proposée pour estimer la gêne sonore induite par un véhicule du trafic routier a été appliquée aux passages de véhicules réalisés sur la piste d'essais de La Ferté-Vidame. Cela a permis de constituer une base de données d'apprentissage qui doit permettre l'estimation de la gêne dans la ville. Dans ce chapitre nous présentons une application dans un contexte réel.

Cette application a mis en œuvre une antenne linéaire de 128 microphones le long d'une grande artère parisienne : le quai Saint-Bernard. Cette voie urbaine se caractérise par un trafic assez important (moyenne annuelle : entre 25 000 et 75 000 véhicules par jour mesuré en 2010 par l'Institut d'Aménagement et d'Urbanisme d'Île-de-France [26]). La configuration linéaire et régulière de l'antenne offre l'avantage de la simplicité d'installation particulièrement adaptée au contexte de la ville.

Cependant la méthode a nécessité quelques adaptations pour se montrer efficace en situation réelle. Il a fallu notamment rendre plus robuste l'étape de *tracking* pour tenir compte le mouvement des arbres longeant le QSB. Le dictionnaire d'apprentissage a par ailleurs été enrichi de signaux prélevés sur le QSB et classifié de façon supervisée.

Dans ce chapitre, nous présentons le dispositif expérimental utilisé pour ces mesures. Les adaptations nécessaires à l'étape de *tracking* et à l'amélioration de la robustesse de l'outil de classification seront ensuite exposées.

Finalement, les résultats de classification et d'estimation des gènes spécifiques sur une journée de mesure pour cette voie de circulation parisienne sont analysés.

5.1 Extraction du signal audio d'un véhicule dans la ville

Dans le contexte urbain du quai Saint-Bernard, le processus permettant d'obtenir la signature acoustique propre à chaque véhicule pris dans le flux du trafic est relativement complexe. Il s'agit dans un premier temps d'établir le lien géométrique entre l'image de la caméra et la position tridimensionnelle dans la scène. Il faut ensuite adapter la méthode de *tracking* en s'affranchissant de la présence d'arbres entre l'antenne et la voie de circulation. Enfin, le résultat de la séparation de sources est présenté.

5.1.1 Dispositif expérimental

L'antenne "LFV-256" utilisée pour les essais sur piste permettait d'avoir une bonne séparation des sources dans un plan qui lui est parallèle. En revanche, son installation en ville est impossible du fait de son encombrement.

Le système "QSB-128" déployé sur le quai Saint-Bernard est présenté en figure 5.1 (et précédemment en figures 2.9a et 2.9b). Il est constitué de 128 microphones MEMS (même modèle et fabricant que ceux présentés précédemment) disposés linéairement avec un pas constant $d = 17$ cm. En comparaison de l'antenne LFV-256, elle est plus légère, plus facile à transporter et très facile d'installation (sur pied ou attachée à une structure) ce qui la rend idéale pour une application dans la ville.

Cette antenne fait 21,6 m d'envergure, ce qui permet d'avoir une haute capacité de résolution en basses fréquences (voir section 2.1.2). Comme on peut le voir sur la figure 5.1, l'antenne est complétée d'un sonomètre de classe 1, d'un anémomètre sonique et d'une caméra enregistrant 8 images par seconde. Celle-ci est disposée au centre de l'antenne.

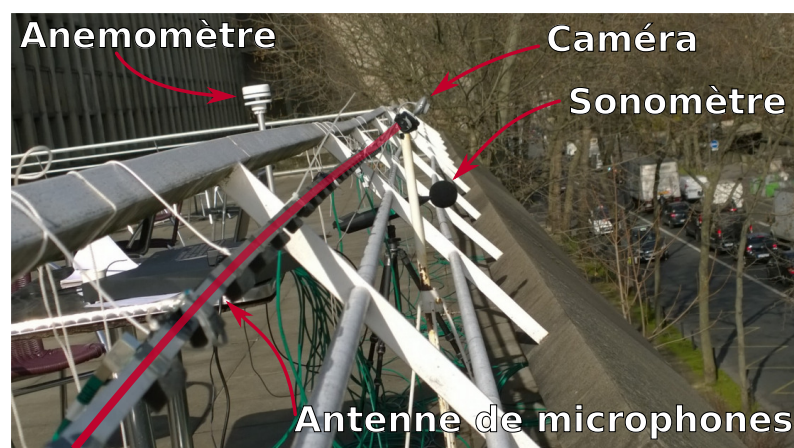


FIGURE 5.1 – Photographie du dispositif expérimental et description des instruments de mesure.

La figure 5.2 montre l'emplacement de l'antenne (représentée par un point ou une ligne rouge) par rapport aux voies de circulation. L'antenne (point rouge) est installée à 9 m au-dessus du Quai Saint-Bernard à Paris. On compte trois voies de circulation vers le centre de Paris et une voie (large, parfois utilisée comme deux voies) vers l'extérieur de Paris. La figure 5.2b situe l'antenne (ligne rouge) par rapport aux feux de signalisation. On constate la proximité avec une intersection. Cela aura un impact sur la distribution des allures des véhicules.

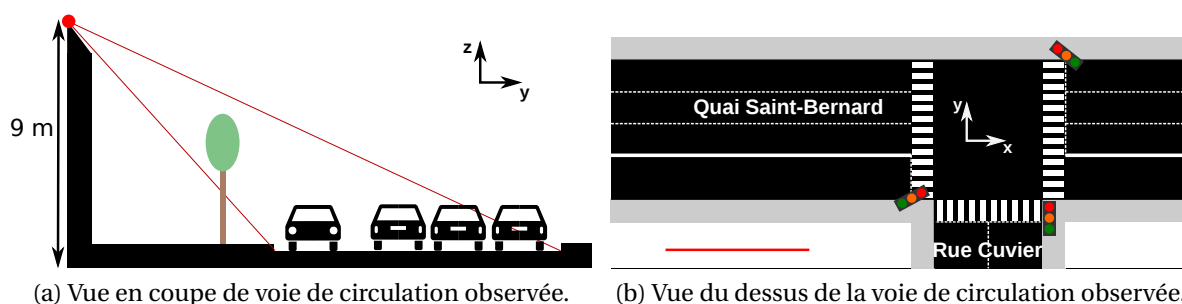


FIGURE 5.2 – Emplacement du système d'acquisition par rapport à la voie de circulation observée.

Comme on l'a vu figure 2.10, cette antenne permet une bonne séparation des sources suivant l'axe d'avancement des véhicules mais ne peut distinguer deux sources qui seraient sur un cercle autour de l'antenne, donc à la même distance radiale. Comme on peut le voir sur la figure 5.2a, la distance entre l'antenne et les deux bordures de la route n'est pas équivalente. Et nous utiliserons cette différence de profondeur pour situer acoustiquement chaque source sur sa voie de circulation. Si l'antenne était installée plus haut, voir complètement au-dessus des voies, il serait impossible de séparer les contributions des différentes voies de circulation.

5.1.2 Du pixel à la position 3D

Lors des essais sur pistes présentés précédemment (chapitre 2) la conversion entre la position en pixel en une position dans l'espace était aisée car les véhicules passaient à une seule et même distance de l'antenne. Grâce à des sources fixes placées au niveau de l'axe de

passage des véhicules et aux extrémités du plan sur lequel l'image acoustique était réalisée nous avons pu faire coïncider la position en pixel et la position dans l'espace par rapport à l'antenne.

Dans le cas présent, plusieurs voies de circulation sont observées dans la profondeur, un effet de perspective est donc naturellement présent et doit être pris en compte. Comme à La Ferté-Vidame, la table de passage entre les pixels de l'image et la position dans l'espace a été faite grâce à des sources fixes contrôlées. Un plan image horizontal a été cherché de sorte à relier les trois sources positionnées à environ 30 cm du sol. L'une était positionnée proche du centre de l'image sur le trottoir le plus proche et les deux autres étaient disposées de chaque côté de l'image sur le trottoir le plus éloigné.

Une formation de voie volumique a été mise en place de sorte à trouver les positions exactes des sources dans l'espace et ainsi définir un plan qui les relie. Les sources utilisées émettaient des signaux harmoniques à 1000 Hz. Une fois les positions des trois sources fixes connues, nous avons pu chercher le meilleur plan image qui permettait d'obtenir les meilleures images acoustiques et donc la meilleure table de passage pixels vers la position en mètres. La figure 5.3a montre la matrice de 60×40 pixels qui couvre la route et le bord des trottoirs. La figure 5.3b montre le résultat de cette recherche d'optimum. Les points bleus représentent les points associés à la matrice de pixel, les points rouges aux microphones de l'antenne. Les projections du plan image suivant un plan horizontal et vertical sont présentées respectivement en figures 5.3c et 5.3d.

Pour prendre en compte la perspective il faut donc avoir un plan trapézoïdal. Mais celui-ci n'est pas symétrique dans notre cas car la caméra n'était pas parfaitement perpendiculaire à l'antenne. Autre déformation à prendre en compte, les pixels représentent de plus en plus de surface au sol quand on s'éloigne de la caméra; ainsi on peut voir sur les figures 5.3b, 5.3c et 5.3d que l'écartement entre les points augmente en \sqrt{y} . Enfin, on peut voir sur la figure 5.3d qu'il a fallu prendre en compte la variation d'altitude de la route (dans la direction y), ce qui a été fait linéairement en pixel.

5.1.3 Séparation de sources en milieu urbain

La capacité de séparation de sources de notre dispositif expérimental est évaluée avec les trois sources fixes mono-fréquentielles présentées précédemment ainsi qu'avec des sources mobiles.

La figure 5.4 montre les densités spectrales de puissance du signal du microphone central de l'antenne et du signal reconstruit à l'emplacement de deux des trois sources fixes émettant un signal harmonique à 1 kHz (une sur chaque trottoir). On peut voir un effet de filtrage important du *beamforming* permettant la réduction de 25 dB autour de la fréquence de la source pour la source la plus proche (12,8 m, figure 5.4a) et 20 dB pour la source située de l'autre côté de la route (26,7 m, figure 5.4b). On remarque que le gain de la formation de voie diminue quand la fréquence baisse, ce qui est due à la perte en résolution avec l'augmentation de la longueur d'onde, intrinsèque à la méthode.

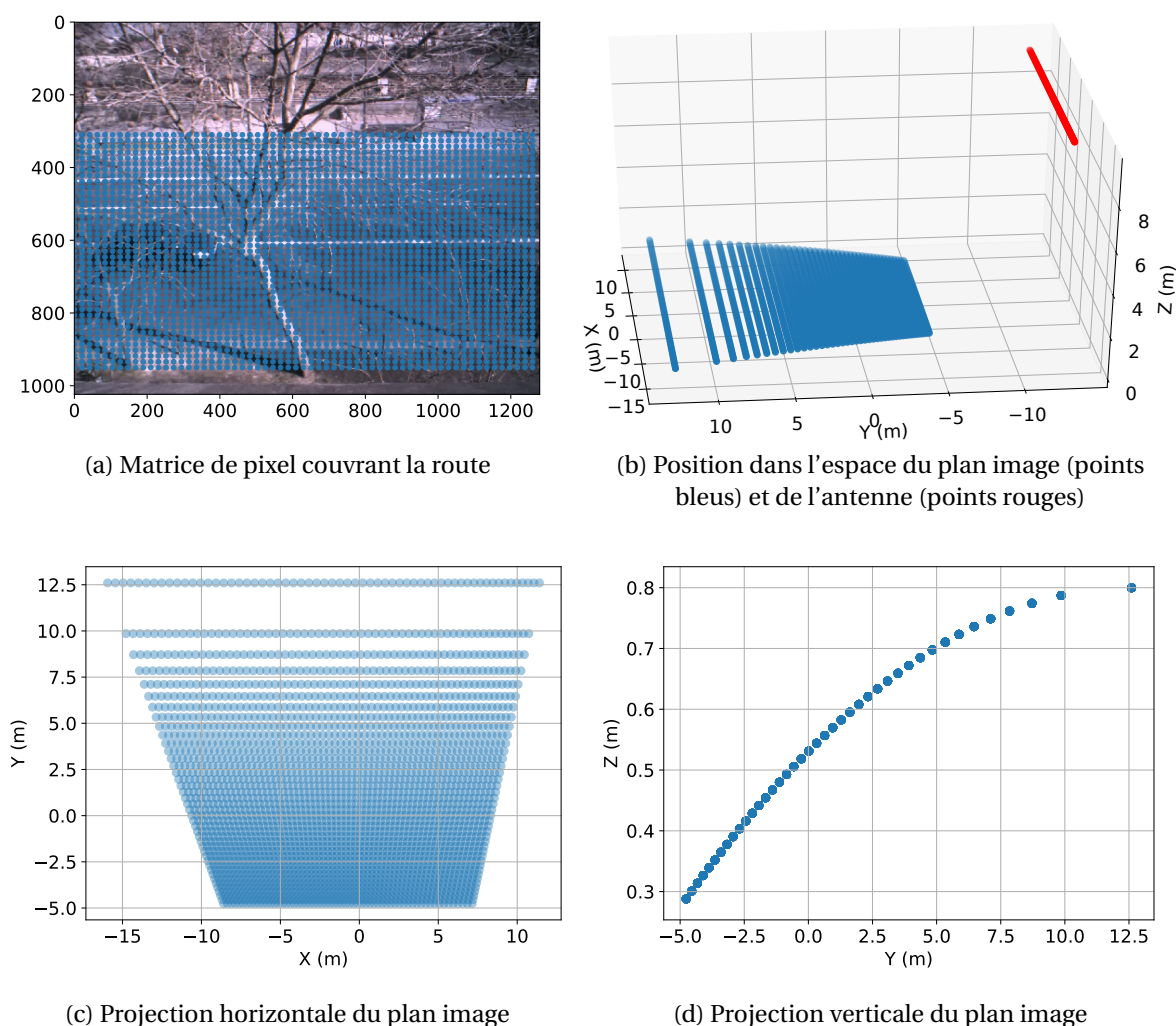


FIGURE 5.3 – Position dans l'espace du plan image pour une matrice de de 60×40 pixels (points bleus) couvrant la route et les trottoirs.

L'application aux sources mobiles nécessite une adaptation de l'algorithme de suivi de véhicule. En effet, la présence d'arbres entre la caméra et les véhicules a imposé de flouter uniquement suivant le sens d'avancement des véhicules avec un filtre gaussien (d'une largeur de 121 pixels) l'image servant de fond (sélectionnée quand aucun véhicule n'est présent) ainsi que l'image courante. La différence des deux images en niveau de gris a dû être de nouveau filtrée par un filtre gaussien de 101×101 pixels pour obtenir de bonnes détections.

Un exemple de suivi de véhicules est donné en figure 5.5. La figure 5.5a présente l'image des objets en mouvement détectés. On peut distinguer deux taches. L'objet en haut de l'image est un piéton, et celui en dessous représente en fait deux scooters. On peut ainsi voir un biais de notre méthode de suivi, deux véhicules proches ne peuvent être distingués et sont considérés comme un seul. La figure 5.5b présente la scène associée, le rectangle autour de l'objet en mouvement détecté et la trajectoire extraite en rouge. On peut voir que, malgré la présence des arbres, la trajectoire n'est pas rompue quand les véhicules passent derrière.

Par la suite, cette étape de *tracking* pourra être améliorée grâce notamment aux algorithmes discutés par Prajapati & Galiyawala [88]. En effet, actuellement la connectivité entre

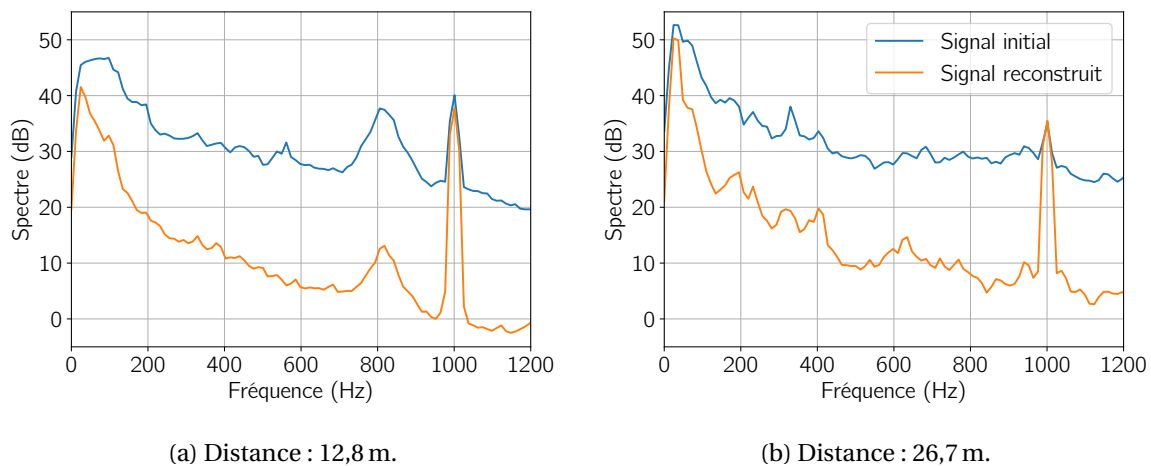
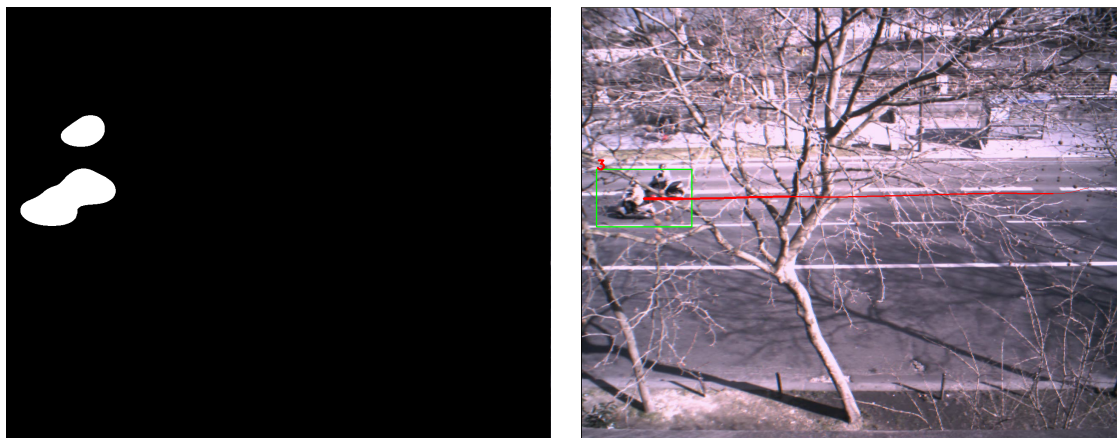


FIGURE 5.4 – Densité spectrale de puissance du signal du microphone central (bleu) et du signal reconstruit (orange) pour deux positions de haut-parleurs émettant un signal harmonique à 1 kHz. Les distances indiquées sont celles entre le microphone central et la source.

les centres des objets détectés entre chaque image n'est faite que par un critère de distance maximale entre deux images à respecter.



(a) Image seuil - Détection des objets en mouvement

(b) Véhicules détectés (rectangle vert) et trajectoire associé (ligne rouge)

FIGURE 5.5 – Exemple de suivi de véhicules sur le Quai Saint-Bernard à Paris

La séparation de sources est maintenant testée sur cette scène sonore composée de deux-roues face à l'antenne et suivis d'autres véhicules en accélération (en dehors du champ de vision de la caméra). La figure 5.6a montre le spectrogramme du signal du microphone central pendant le passage des deux scooters. On ne distingue pas ici de composante fréquentielle particulière mais plutôt un bruit très large bande. La figure 5.6b montre le spectrogramme du signal reconstruit au niveau des deux scooters, qui sont associés à un seul véhicule par l'algorithme de suivi de véhicule comme on l'a vu précédemment. On peut voir que le bruit large bande est sensiblement réduit et que le spectrogramme fait apparaître les composantes tonales des moteurs avec un grand rapport signal sur bruit (autour de 25 dB).

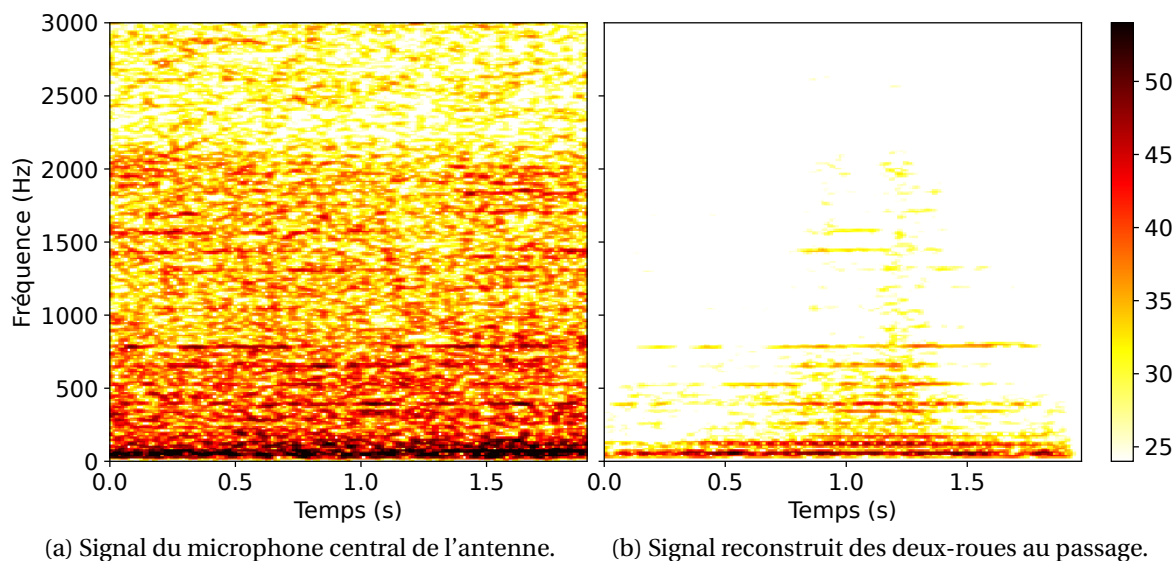


FIGURE 5.6 – Spectrogrammes de deux deux-roues passant détectés comme un seul véhicule – Dynamique : 30 dB.

Il est intéressant de voir que, bien que le critère d'échantillonnage spatial ne soit plus respecté au-dessus de 1000 Hz ($d > \lambda/2$), on arrive tout de même à filtrer les sources éloignées alors que des lobes de réseaux devraient apparaître dans le diagramme de rayonnement et ainsi "polluer" le signal extrait. En revanche on peut voir que la résolution de l'antenne est peut-être trop grande en hautes fréquences ce qui implique un filtrage spatial trop sélectif. Ainsi, si le centre du rectangle détecté ne coïncide pas avec les sources acoustiques, celles-ci sont filtrées.

Cette extraction pourra être améliorée lors de travaux futurs. En effet, Hafizovic *et al.* [52] proposent notamment de ne pas utiliser l'ensemble des microphones sur l'intégralité du spectre et ainsi de dégrader la résolution de l'antenne en hautes fréquences en "éteignant" des microphones pour garder une largeur de faisceau constante.

Nous avons vu dans cette section qu'il est possible d'utiliser un grand réseau de microphones en milieu urbain pour extraire le signal audio d'un véhicule dans le trafic. Pour ce faire, l'algorithme de suivi de véhicule permettant de connaître la trajectoire de l'objet en mouvement a été modifié pour obtenir la trajectoire de chaque véhicule malgré la présence d'arbres. Bien que perfectible, l'extraction du signal audio paraît satisfaisante pour le calcul des MFCCs et leur utilisation par le classifieur.

5.2 Classification *in situ*

L'algorithme de classification présenté au chapitre 3 est ici appliqué aux signaux issus des essais dans le milieu urbain. Les premiers résultats avec la base d'apprentissage issus des véhicules isolés sont présentés et l'amélioration de la robustesse est étudiée en complétant le dictionnaire d'apprentissage avec des données annotées manuellement et en étudiant les meilleurs vecteurs de pondérations possibles du dictionnaire constitué des MFCCs et des informations d'allure.

5.2.1 Classification avec apprentissage sur véhicules isolés

L'algorithme de classification est testé sur un segment de dix minutes enregistré à 11h50, traité pour affecter à chaque véhicule identifié une catégorie de façon supervisée. Pendant ces dix minutes d'acquisition, 180 véhicules ont été détectés et annotés manuellement. Les données (MFCC et allures) issues des signaux de chaque véhicule sont normalisées par les mêmes valeurs que pour l'établissement du dictionnaire d'apprentissage sur les données de La Ferté-Vidame (voir section 3.2.2). L'erreur d'identification pour les 180 véhicules détectés atteint 52,8% avec le vecteur de pondération \mathbf{W} présenté précédemment et rappelé ci-dessous :

$$\mathbf{W} = \underbrace{[2, 1, 0, 4, 2, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1]}_{\text{MFCCs}}, \underbrace{[1, 1, 1]}_{\text{Allures}} \quad (5.1)$$

La matrice de confusion présentant le détail des résultats de classification est donnée dans le tableau 5.1. On peut voir sur ce tableau que lors de ces dix minutes d'acquisition traitées, aucun deux-roues en décélération (catégorie 4) ou en accélération (catégorie 2) n'a pu être détectés. Pour la catégorie 4, cela est dû au fait que ces véhicules doublaient les véhicules déjà à l'arrêt. En effet, le floutage nécessaire pour le *tracking* ne distingue alors plus qu'un véhicule au lieu des deux et la trajectoire s'interrompt à chaque dépassement. Cela implique donc qu'il est très peu probable, pour l'instant, de détecter des véhicules de cette catégorie. L'absence de la deuxième catégorie s'explique par le fait que les deux-roues étaient déjà à vitesse constante en passant devant l'antenne. De plus, on peut noter la place prédominante de la catégorie 3.

Il faut remarquer qu'aucun des véhicules de la catégorie 7 n'est classifié correctement et que beaucoup d'erreurs sont commises au sein d'un même type de véhicule ou au sein d'une même allure.

		Estimée							
		Nombre	Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	13	30,8	7,7	238	0	0	38,5	0
	Cat.2	0	–	–	–	–	–	–	–
	Cat.3	121	26,5	0	47,1	0	0	26,4	0
	Cat.4	0	–	–	–	–	–	–	–
	Cat.5	23	17,3	4,4	4,4	13	56,5	4,4	0
	Cat.6	18	0	22,2	5,6	0	11,1	61,1	0
	Cat.7	5	0	0	0	0	0	100	0

TABLE 5.1 – Matrice de confusion en pourcentages de véhicules par catégories perceptives - C = 2,5 - Erreur globale : 52,8% - Pas de données des catégories 2 et 4 disponibles.

Il s'avère donc nécessaire d'enrichir la base de données d'apprentissage par des données mesurées en situations réelles et annotées. Les dix minutes traitées permettent de constituer 180 échantillons supplémentaires d'apprentissage qui vont venir compléter le dictionnaire d'apprentissage initialement composé de 68 passages de véhicules isolés.

5.2.2 Classification avec apprentissage sur données acquises dans la ville

Les 180 échantillons annotés sont ajoutés aux données d'apprentissages constituant donc un dictionnaire de 248 échantillons. La classification est alors testée sur des signaux issus de cinq minutes extraites de l'acquisition de 13h25 réalisée la même journée, ce qui représente 62 signaux annotés. Il est à noter que, ici encore, la catégorie 4 n'est pas représentée dans cet ensemble de signaux, mais trois véhicules de la catégorie 2 sont présents. Le taux d'échec de classification est alors (toujours avec le même vecteur de pondération et la même valeur du coefficient $C = 2,5$) de 43,6 %.

Il est donc important de chercher à améliorer le résultat de la classification automatique. Pour cela, nous allons utiliser les résultats de l'étude paramétrique précédemment présentée (section 3.2.2) qui a permis de montrer que 26 301 combinaisons permettent d'obtenir un résultat de 100% de bonne classification sur un apprentissage et un test sur les données de véhicules isolés.

La classification des 62 passages dans la ville avec une base d'apprentissage de 248 signaux est donc testée sur ces 26 301 combinaisons de coefficients de pondération pour $C = 2,5$. L'erreur globale est réduite et s'établit à 21% pour sept combinaisons différentes. Si le coefficient C prend les valeurs entre 1 et 0,2 l'erreur est légèrement réduite à 19,4% (une combinaison possible pour $C = 1$ et deux combinaisons possibles pour $C = 0,2$). Les deux valeurs de pondérations donnant ce résultat pour $C = 0,2$ sont les suivantes :

$$\mathbf{W} = [2, 5, 0, 2, 2, 0, 2, 0, 4, 2, 0, 1, 1, 1, 1] \quad (5.2)$$

$$= \underbrace{[2, 5, 0, 3, 3, 0, 3, 0, 5, 2, 0, 0]}_{\text{MFCCs}}, \underbrace{[1, 1, 1]}_{\text{Allures}} \quad (5.3)$$

Ces résultats ne sont toujours pas satisfaisants. En effet, au-delà de la valeur importante de l'erreur globale, la matrice de confusion (non-présentée ici) met en évidence une incapacité totale à reconnaître les deux-roues ainsi que les véhicules lourds en accélération.

5.2.3 Influence du noyau du classifieur

La modification du noyau de l'algorithme de classification n'avait présenté aucun intérêt lors d'une étude préliminaire sur les données issues de la campagne de mesure à La Ferté-Vidame (section 3.2.2). Le noyau linéaire était même celui qui proposait les meilleurs résultats. Nous nous proposons ici de confirmer ou d'infirmer que le noyau linéaire est le plus adapté à notre étude en testant les deux autres noyaux classiques à notre cas présent : dictionnaire d'apprentissage étendu aux données *in situ*.

Noyau gaussien Le noyau gaussien (appelé RBF pour *radial basis function kernel*) est ici utilisé. Dans la librairie `scikit-learn`, le noyau est posé légèrement différemment que dans

l'équation (3.5) :

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \text{avec} \quad \gamma = \frac{1}{2\sigma^2}. \quad (5.4)$$

La classification est réalisée sur les mêmes données : 248 échantillons d'apprentissage (mixant données de véhicules isolés et signaux extraits dans le milieu urbain) et les 62 signaux de test issus des mesures sur le quai Saint-Bernard à Paris présentés à la section précédente. Le meilleur résultat de classification est alors obtenu avec $\gamma = 0,1$ et $C = 11$. L'erreur globale n'est pas abaissée et est de 19,4%. Celle-ci a été obtenue pour six vecteurs de pondérations sur les 26 301 testés, l'un d'entre eux vaut

$$\mathbf{W} = \underbrace{0, 1, 0, 3, 3, 0, 2, 0, 4, 1, 0, 0}_{\text{MFCCs}}, \underbrace{1, 1, 1}_{\text{Allures}}]. \quad (5.5)$$

On peut voir que les premiers MFCCs interviennent peu dans cette classification.

Le détail des confusions entre les catégories estimées et attendues est présenté en tableau 5.2. Les confusions sont assez nombreuses. En effet, tous les deux-roues passant à vitesse constante (Cat. 1) sont considérés comme des véhicules légers ou lourds à la même allure (Cat. 3), les deux-roues en accélération (Cat. 2) sont pour 66 % confondus avec des véhicules légers en accélération (Cat. 6) et 33 % sont confondus avec la catégorie 3. Le système ne classe bien les poids lourds et bus en accélération (Cat. 7) que 25% du temps. Ces véhicules se répartissent aussi sur deux autres catégories représentant différents types de véhicules et différentes allures.

		Estimée						
		Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	0	0	100	0	0	0	0
	Cat.2	0	0	33,3	0	0	66,7	0
	Cat.3	0	0	95,1	0	2,4	2,4	0
	Cat.4	-	-	-	-	-	-	-
	Cat.5	0	0	20	0	80	0	0
	Cat.6	0	0	14,3	0	0	85,7	0
	Cat.7	0	0	0	0	25	50	25

TABLE 5.2 – Matrice de confusion en pourcentage de véhicules par catégories perceptives - Noyau gaussien avec $C = 11$ et $\gamma = 0,1$ - Erreur globale : 19,4% - Pas de données de la catégorie 4 disponibles.

Noyau polynomial Le noyau polynomial est ici testé avec les mêmes données d'apprentissage et de test que dans le paragraphe précédent. Conformément à la formulation du noyau (vu à l'équation (3.4)), les paramètres à optimiser sont d (l'exposant du noyau) et C (contrôlant toujours le compromis entre le nombre d'erreurs et la largeur des marges).

Le meilleur résultat obtenu est pour $d = 4$ et $C = 3$, avec une erreur qui égale celle du noyau gaussien (19,4%). Mais les confusions entre catégories sont légèrement différentes comme on peut le voir dans le tableau 5.3. On peut voir que l'on perd 14,3% de bonne clas-

sification sur la catégorie 6 ainsi que 25% sur la catégorie 7. La seule différence intéressante étant la classification parfaite de la catégorie 3.

	Estimée							
	Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7	
Attendue	Cat.1	0	0	100	0	0	0	0
	Cat.2	0	0	33,3	0	0	66,7	0
	Cat.3	0	0	100	0	0	0	0
	Cat.4	–	–	–	–	–	–	–
	Cat.5	0	0	20	0	80	0	0
	Cat.6	0	0	28,6	0	0	71,4	0
	Cat.7	0	0	0	0	25	75	0

TABLE 5.3 – Matrice de confusion en pourcentage de véhicules par catégories perceptives - Noyau polynomial avec $C = 3$ et $d = 4$ - Erreur globale : 19,4% - Pas de données de la catégorie 4 disponibles.

Ces résultats ne sont donc pas satisfaisants au regard de la littérature pour cette méthode [108], ni au regard de notre objectif. En effet, tous les véhicules sont catégorisés comme des véhicules légers ou lourds.

Optimisation finale Pour rendre plus robuste la classification, il est proposé de créer un nouveau dictionnaire d'apprentissage avec les données des deux séquences annotées (acquisitions à 11h50 et 13h25) pour avoir une meilleure diversité des scènes sonores. Ainsi, comme résumé en figure 5.7, 60 signaux sont isolés à la fin de la mesure faite à 11h50 (qui en contient 180) pour être utilisés comme données de test et les 120 autres signaux sont ajoutés à ceux issus de la mesure de 13h25 ainsi que ceux des véhicules isolés sur piste d'essais (LFV) pour constituer le dictionnaire d'apprentissage. Les signaux ont été sélectionnés de manière contiguë, de sorte à avoir au moins un représentant de chaque catégorie présent dans ce jeu de données (dont les catégories 2 et 4 sont absents). Les différents noyaux avec différentes valeurs de leurs paramètres seront testés sur toutes les pondérations \mathbf{W} disponibles.

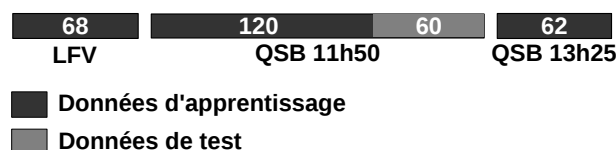


FIGURE 5.7 – Répartition des Signaux entre les données d'apprentissage et de test pour les mesures sur piste d'essais à La Ferté-Vidame (noté LFV) et les données en milieu urbain (notés QSB). Les nombres indiquant le nombre de signaux utilisés

Les meilleures performances de classifications pour un noyau gaussien sont obtenues avec $\gamma = 0,1$ et $C = 15$, l'erreur diminue alors à 13,3 %, ce qui est satisfaisant. Cela est obtenu pour six pondérations dont voici un exemple :

$$\mathbf{W} = \underbrace{[5, 5, 0, 4, 4, 0, 0, 0, 2, 0, 0, 1, 1, 1, 1]}_{\text{MFCCs}}, \underbrace{[1, 1, 1, 1]}_{\text{Allures}}. \quad (5.6)$$

Contrairement aux vecteurs de pondération présentés précédemment, ce sont ici les premiers MFCCs, représentant le contenu en basses fréquences, qui sont les plus importants (hormis ceux volontairement fixés à zéro, voir section 3.2.2) pour l'algorithme SVM.

La matrice de confusion est présentée en tableau 5.4. On peut voir que les bus et véhicules lourds (Cat. 7) sont bien mieux estimés. On peut aussi noter que les véhicules légers et lourds en décélération (Cat. 5) sont estimés à 40% comme passant à vitesse constante (Cat.3). Enfin, on peut voir que certains deux-roues roulant à vitesses constantes (Cat. 1) sont détectés à 50% comme des véhicules légers ou lourds.

		Estimée						
		Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	50	0	50	0	0	0	0
	Cat.2	–	–	–	–	–	–	–
	Cat.3	0	0	95,4	0	0	4,6	0
	Cat.4	–	–	–	–	–	–	–
	Cat.5	0	0	40	0	60	0	0
	Cat.6	0	0	0	0	0	100	0
	Cat.7	0	0	0	0	0	0	100

TABLE 5.4 – Matrice de confusion en pourcentage de véhicules par catégories perceptives - Noyau gaussien avec $\gamma = 0,1$ et $C = 15$ - Erreur globale : 13,3% - Pas de données pour les catégories 2 et 4.

Le noyau polynomial est ensuite appliqué à la même tâche de classification. Il permet d'obtenir au mieux une classification avec erreur globale de 16,7% pour les mêmes paramètres que ceux présentés précédemment : $C = 3$ et $d = 4$. Cette erreur de classification a été obtenue pour 263 vecteurs de pondération différents avec des valeurs très variées des coefficients. Le tableau 5.5 présente la matrice de confusion associée. On peut voir une perte de qualité du classifieur pour les catégories 3 et 7 et une relative amélioration des performances de classification pour les catégories 1 et 5.

		Estimée						
		Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Attendue	Cat.1	75	0	25	0	0	0	0
	Cat.2	–	–	–	–	–	–	–
	Cat.3	4,7	0	86	0	0	7	2,3
	Cat.4	–	–	–	–	–	–	–
	Cat.5	10	0	10	0	70	10	0
	Cat.6	0	0	0	0	0	100	0
	Cat.7	0	0	100	0	0	0	0

TABLE 5.5 – Matrice de confusion en pourcentage de véhicules par catégories perceptives - Noyau polynomial avec $C = 3$ et $d = 4$ - Erreur globale : 16,7% - Pas de données pour les catégories 2 et 4.

Cette étude montre l'importance de fournir un nombre important de données d'apprentissage et de diversité des scènes sonores aux différentes périodes de la journée. De plus, le noyau gaussien est apparu comme le plus efficace. En effet, l'erreur de classification associée est proche des résultats rapportés dans la littérature pour cette méthode [108]. Ce sera donc

celui utilisé dans la suite de cette étude.

Par ailleurs, il faut garder à l'esprit que de nombreux facteurs jouent dans les erreurs de classification que nous avons observées. On peut citer par exemple l'estimation de l'allure faite de façon automatique grâce à l'algorithme de suivi de véhicule. Ce dernier n'étant pas parfait, la vitesse estimée est parfois très bruitée et notre algorithme ne détermine alors pas la bonne allure, ce qui n'aide pas à la bonne classification. Une autre source d'erreur a pu intervenir dans l'annotation manuelle des véhicules. En effet, bien qu'une attention particulière ait été portée à cette tâche, l'erreur humaine n'est pas impossible, notamment sur l'estimation de l'allure du véhicule qui, là encore, s'est basée en partie sur l'évolution de la vitesse obtenue par le suivi d'objet.

De plus, il faut rappeler que nous ne pouvons connaître le taux d'échec de bonne classification pour les catégories 2 et 4, celles-ci n'étant pas présentes dans les acquisitions étudiées.

Comme cela a été rappelé, le *tracking* est un point clé du traitement. C'est d'autant plus le cas dans notre situation. En effet, comme indiqué en section 5.1.3, la présence d'arbres entre les véhicules et la caméra a nécessité un floutage qui permet la continuité des trajectoires mais qui induit aussi de nombreuses fois le groupement de plusieurs véhicules au sein d'un même objet en mouvement détecté. Cela induit forcément un signal audio extrait de moins bonne qualité, d'autant plus s'il y a des véhicules de différents types.

Néanmoins, nous avons montré dans cette section que, bien que perfectible, cet outil était capable d'estimer la catégorie perceptive des véhicules composant le trafic routier sur le tronçon de voie de circulation observé avec des résultats tout à fait satisfaisants. Nous pouvons maintenant chercher à estimer la gêne sonore induite par chacun de ces véhicules routiers.

5.3 Estimation de la gêne sonore *in situ*

Dans cette section, nous appliquons l'outil dans sa globalité pour l'estimation de la gêne spécifique à chaque véhicule passant dans l'artère parisienne observée : le quai Saint-Bernard. Les adaptations apportées dans les sections précédentes de ce chapitre sont utilisées et la gêne est estimée grâce au modèle de Morel *et al.* [78] et aux indices psychoacoustiques dont le mode de calcul a été présenté au chapitre 4.

L'outil est donc testé sur six séquences de dix minutes réparties de 11h50 à 15h15 lors de mesures réalisées le 6 mars 2015. La figure 5.8 présente les résultats d'estimation de la gêne spécifique obtenus pour trois acquisitions en fonction de l'heure de passage du véhicule et de sa catégorie perceptive estimée. On peut noter sur toutes les séquences une prédominance de la troisième catégorie rassemblant les véhicules légers et lourds passant à vitesse constante. De plus, on peut noter une certaine périodicité dans la détection de cette catégorie. En effet, il semble y avoir des absences à rythme régulier (environ toutes les minutes). On peut noter que sur ces trois exemples, les deux-roues en accélération et en décélération (respectivement catégories 2 et 4) ne sont pas présents ou pas détectés. Il n'y a qu'à 12h27 qu'un véhicule est classé en catégorie 4 (au début de la séquence). L'échelle des couleurs symbolisant la gêne spécifique estimée est fixée entre 0 et 10. On peut voir que certaines catégories dépassent systématiquement la valeur maximale du modèle (non-borné) de Morel *et al.*, c'est le cas notamment des véhicules légers en accélération (catégorie 6) dont l'estimation de la gêne ne dépend que du niveau sonore en moyennes fréquences (voir tableau 1.6). Les expériences utilisées pour construire le modèle de Morel *et al.* se sont limitées à un niveau maximum de 62 dB(A). Or la moyenne des niveaux sonore mesurés à 11h50 pour les véhicules de la catégorie est de 66,5 dB(A) avec un écart-type des valeurs de 2,9 dB(A), ce qui explique cette sur-estimation.

L'analyse dans le détail fait apparaître sur les figures 5.8a et 5.8c, et dans une moindre mesure sur la figure 5.8b, des motifs temporels qui se répètent. Un exemple de ce type de motif est mis en évidence sur la figure 5.8a par une séquence de chiffres romains. Cela est dû à la proximité d'un feu de circulation (voir figure 5.2). D'abord, on observe des véhicules légers et lourds passant à vitesse constante (phase I). Ensuite, ces mêmes véhicules décélèrent (phase II). Puis, on détecte les véhicules légers en accélération (phase III). Ces trois phases traduisent le comportement des véhicules sur une voie de circulation. La phase IV correspond à la présence de deux-roues passant sur la voie opposée qui ont fini d'accélérer en arrivant devant l'antenne.

L'évolution des gênes spécifiques au cours de la journée est présentée en figure 5.9 via les valeurs moyennes et les écarts-type sur chaque acquisition de dix minutes. On peut voir en figure 5.9a que les catégories 2 et 4 sont détectées lors de certaines acquisitions mais toujours en nombre très faible : une fois par acquisition, d'où les écarts-type nuls associés.

L'analyse des figures 5.9a et 5.9b montre que la seule valeur disponible de deux-roues en accélération (catégorie 4 à 15h00) induit la plus forte gêne. La seconde catégorie la plus gênante est la sixième (véhicules légers en accélération). Les valeurs obtenues ont un écart-type assez faible, hormis à 14h45, et sont relativement constantes le long de la journée. Les

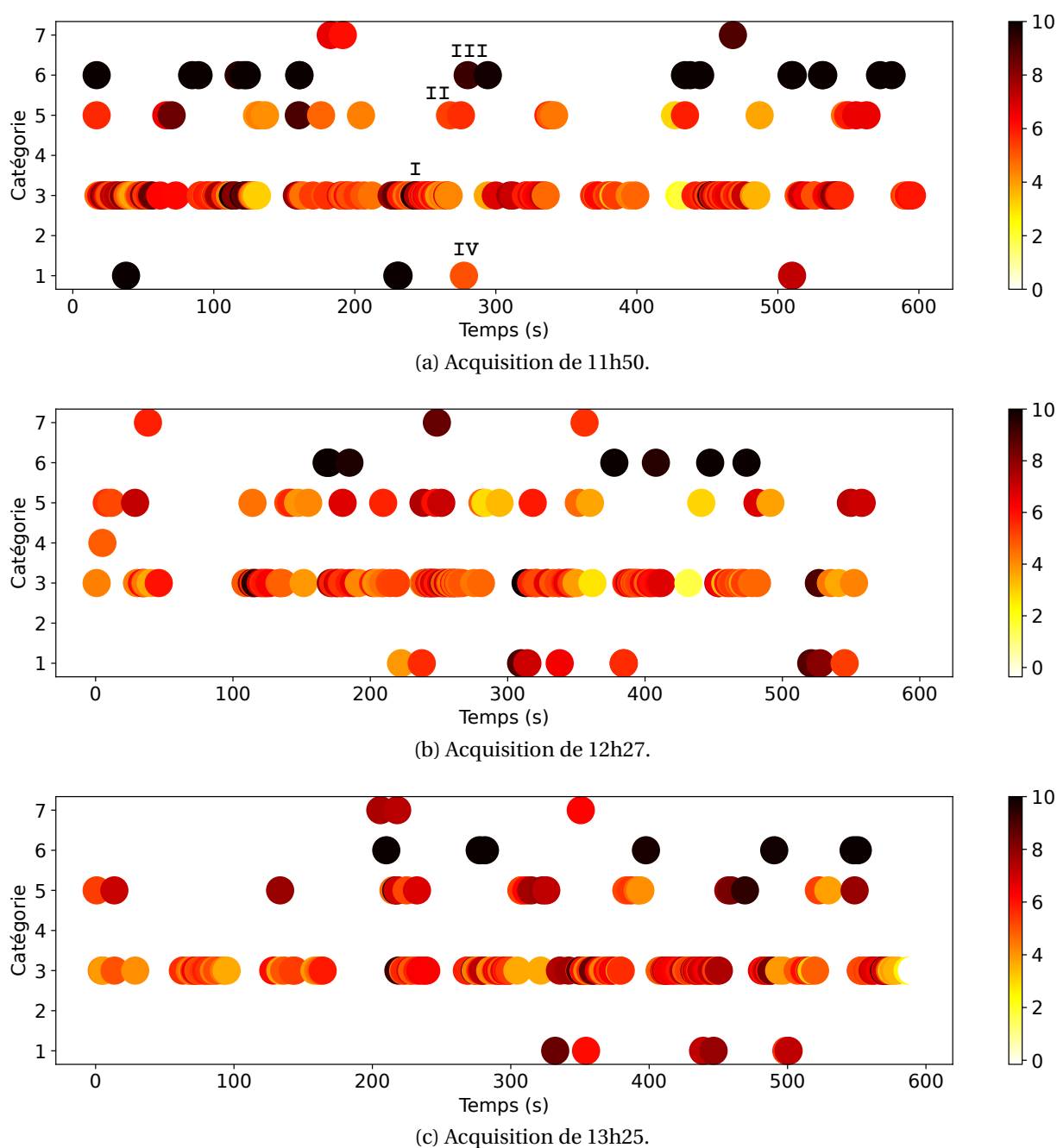
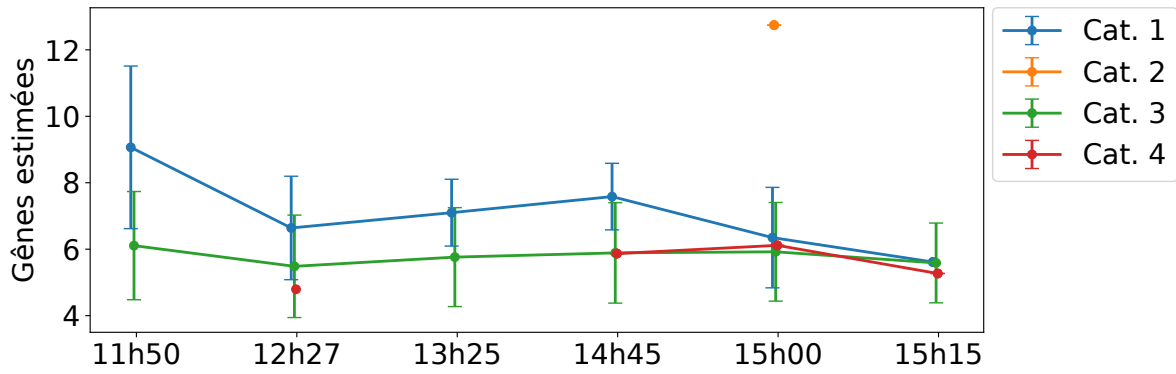
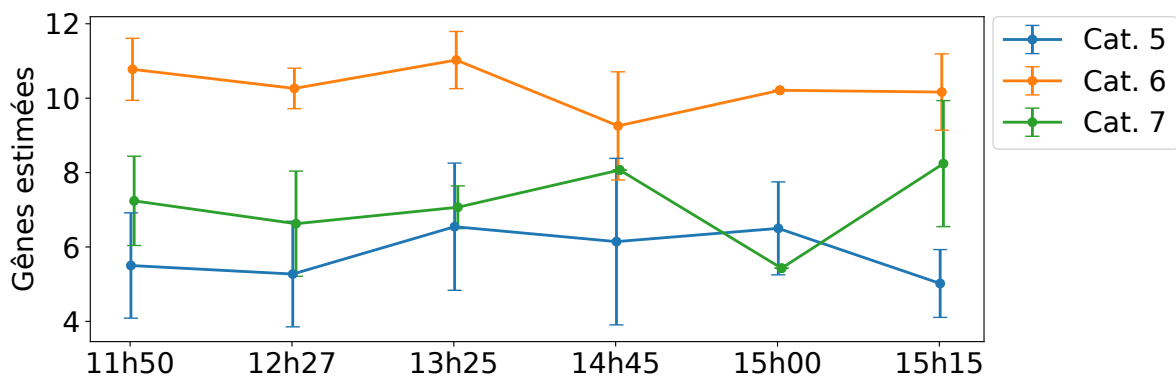


FIGURE 5.8 – Estimation de la catégorie perceptive et de la gène spécifique de chaque véhicule du trafic routier pendant trois acquisitions de dix minutes.

catégories 1 et 7 ont des valeurs de gènes spécifiques assez proches. On peut noter l'écart-type important, s'élevant à 2,4. On distingue une baisse de la gène due à la catégorie 7 à 15h, qui est obtenue pour un seul véhicule : un bus dont l'échappement était masqué. Enfin on peut voir que les catégories 3 et 5 sont les catégories pour lesquelles la gène estimée est la plus faible : en moyenne 6/10. On peut noter cependant une grande variabilité des gènes pour la catégorie 5. Ce qui peut s'expliquer par la présence des bus, véhicules lourds et légers dans cette catégorie, or le niveau sonore de la source principale (le moteur) diffère beaucoup entre ces trois types de véhicules.



(a) Catégories perceptives 1 à 4.



(b) Catégories perceptives 5 à 7.

FIGURE 5.9 – Moyennes et écarts-type des gènes spécifiques estimées pour chaque catégorie en fonction de la période de la journée.

Conclusion

Nous avons montré dans ce chapitre une application de l’outil d’estimation des gènes sonores spécifiques de véhicules routiers au milieu urbain. Cela a été fait sur des acquisitions de dix minutes réalisées en mars 2015 sur le Quai Saint-Bernard, une grande artère parisienne.

Dans un premier temps, la validité de l’approche a été confirmée par la réalisation de la formation de voies sur des sources fixes de part et d’autre des voies de circulation. On a alors montré une réduction du bruit de fond de 25 dB en hautes fréquences, avec une efficacité décroissante en basses fréquences pour atteindre 10 dB de réduction à 100 Hz.

La robustesse du *tracking* a été obtenue en floutant les images utilisées pour la détection de contours et l’efficacité de formation de voies sur sources en mouvement a été montrée sur l’exemple d’un deux-roues passant à vitesse constante. Les composantes tonales dues au moteur étaient alors clairement visibles sur le spectrogramme alors qu’elles étaient noyées dans le bruit de fond sur le spectrogramme d’un signal microphonique seul.

L’étape de classification a été ensuite testée. Une robustesse accrue a été nécessaire pour obtenir des résultats satisfaisants. Cela a été obtenu en ajoutant, à notre base de données

d'apprentissage issue des expériences de La Ferté-Vidame, des données annotées manuellement issues de deux acquisitions de la base de données du Quai Saint-Bernard. Les vecteurs de pondération du dictionnaire d'apprentissage permettant d'obtenir les meilleurs résultats de classification ont été obtenus parmi ceux issus de l'étude paramétrique présentée en section 3.2.2. En testant le classifieur sur des signaux annotés manuellement n'ayant pas servi à l'apprentissage, une erreur de 13,3% a pu être obtenue avec l'utilisation d'un noyau gaussien pour six pondérations différentes.

Cet outil de classification a ensuite pu être appliqué à l'ensemble des données disponibles pour estimer la gêne sonore spécifique à chaque véhicule tout au long de la journée. Les résultats obtenus montrent des valeurs de gêne sonore dépassant 10, valeur théorique maximale du modèle (non-borné) signifiant "extrêmement gênant". Cela étant probablement dû au fait que le signal audio est reconstruit au niveau de la source, compensant alors de la perte d'énergie lors de la propagation. Cela pourra être modifié dans des travaux futurs en modifiant le coefficient de normalisation de la formation de voies pour obtenir le signal audio de la source évalué au point de mesure.

Les limites du système ont été énoncées. Elles résident principalement dans l'algorithme de suivi de véhicule actuellement utilisé et dans l'estimation des paramètres psycho-acoustiques du signal audio. En effet, le *tracking* ne permet pas d'obtenir la trajectoire de tous les véhicules passant sur la voie de circulation. De plus, il peut regrouper au sein d'un même objet en mouvement plusieurs véhicules circulant côte-à-côte, ce qui sous-estime le nombre de véhicules et dégrade la classification si différents types de véhicules sont regroupés.

A l'issue de cette dernière partie de la thèse, on peut considérer disposer d'un outil opérationnel de mesure de la gêne sonore liée au trafic en milieu urbain. Cela valide la méthode dont nous voulions faire la preuve de concept dans cette thèse.

Cet outil présente encore des pistes d'amélioration et permet de tracer quelques perspectives d'un travail futur impliquant notamment plus fortement les considérations psychoacoustiques.

Conclusion

Le bruit est cité par de nombreuses études comme étant une source de préoccupation majeure, aussi bien en ville qu'en province. Dans de nombreuses études, une grande part de la population française se déclare gênée par le bruit aussi bien à son domicile qu'au travail. Jusqu'à présent les réponses des pouvoirs publics ne se basent que sur une quantification énergétique de l'exposition sonore, souvent via la mesure ou l'estimation du L_A ou du L_{den} , et des prises de décisions relatives à la diminution du niveau sonore. Or des études psychoacoustiques ont montré que le niveau sonore n'expliquait qu'une faible part de la gêne sonore ressentie.

Les travaux présentés dans ce document cherchent donc à proposer un outil pour estimer la gêne sonore perçue dans le milieu urbain via l'assimilation de différentes approches et techniques. Ce travail repose sur l'utilisation de deux modèles psycho-acoustique proposés par Morel *et al.* en 2016. Ils sont issus de plusieurs expériences perceptives. La première a abouti à une catégorisation perceptive des véhicules alors que la deuxième a mené à la proposition d'un modèle de gêne sonore pour chacune de ces catégories. Les sept modèles se basent sur des sommes pondérées de différents indices énergétiques ou psychoacoustiques. Ces derniers permettent d'avoir une information plus complète sur la source de bruit en modélisant des sensations sonores. Notre travail a consisté notamment à établir le lien entre les sources immergées dans le trafic urbain et ce modèle pour constituer un outil complet (encore perfectible) de mesure de la gêne sonore liée au trafic routier en milieu urbain.

Estimation de la gêne. L'estimation de la gêne a nécessité le calcul d'indices psychoacoustiques tels que la rugosité ou la sonie. Pour ce dernier, une norme internationale existe et permet d'avoir des résultats satisfaisants avec des implémentations disponibles librement. En revanche pour les autres indices, aucune norme n'existe et les modèles disponibles sont le plus souvent intégrés dans des logiciels propriétaires. Nous avons cependant trouvé et utilisé des modèles développés récemment et disponibles librement. Cependant, ils ont nécessité d'être modifiés pour suivre les bonnes tendances pour les signaux caractéristiques. Les résultats obtenus sont satisfaisants et permettent d'estimer la gêne sonore.

Classification de véhicule au passage. Pour utiliser ce modèle de gêne sonore dans des situations réelles, il faut connaître le type de véhicule et son allure de sorte à lui associer une catégorie perceptive. Pour réaliser cela, nous avons proposé une méthode basée sur l'utilisation du signal audio du véhicule et sur du traitement d'images. L'intérêt est de disposer d'un système adaptable à différentes situations, que ce soit avec une voie très peu passante

où seul un micro suffira, ou sur une grande artère où une séparation de sources est alors nécessaire pour filtrer spatialement les sources observées.

Le signal audio du véhicule et l'information d'allure obtenue par le traitement des images issus d'une caméra filmant la scène sonore sont utilisés pour réaliser cette classification. Nous avons utilisé les *Mel Frequency Cepstral Coefficients* (MFCC) pour extraire les caractéristiques du signal ainsi que l'information d'allure (vitesse constante, accélération ou décélération). Ces informations ont alimenté un classifieur basé sur une machine à vecteurs de support (SVM) qui a donné un bon taux de classification pour des véhicules isolés sur piste d'essai.

Séparation de sources. Pour utiliser le modèle psychoacoustique précédent dans le milieu urbain de façon automatique, il a fallu concevoir un outil de séparation des signaux spécifiques des différentes sources de bruit, donc des différents véhicules routiers dans notre cas. Pour réaliser cela, des antennes étendues à grand nombre de microphones basées sur le projet *Mégamicros* ont été conçues et une méthode de formation de voies a été adaptée.

Cette dernière est issue des méthodes de formation de voies fréquentielles parallélisées sur carte graphique développées dans la thèse de Charles Vanwynsberghe [110]. Grâce à cette méthode de calcul, il est possible d'estimer le spectre complexe provenant d'un point de l'espace : le signal audio est reconstruit par la somme des transformées de Fourier inverses des spectres complexes obtenus par formation de voies sur des trames de 50 ms recouvertes à 75%.

Les antennes, couplées avec la technique de formation de voies, ont permis de filtrer spatialement chaque véhicule et d'extraire son signal audio avec une dynamique moyenne de 20 dB. Ce système a montré son efficacité en milieu urbain où le signal microphonique ne permet pas de distinguer les différentes sources de bruit alors que le signal reconstruit au niveau du véhicule filtre efficacement les sources environnantes pour ne garder que le son venant du véhicule.

La gêne spécifique à chaque véhicule a alors pu être estimée de façon automatique sur les données obtenues en milieu urbain pendant cette thèse. Cela permet ainsi d'avoir des informations bien plus détaillées sur le paysage sonore en un lieu donné que celles fournies par les cartes de bruit réglementaires.

Cependant, nos résultats n'ont pas pu être confrontés à des expériences perceptives. De même, faute d'expérience perceptive et d'accord dans la littérature sur des modèles de gêne totale, nous n'avons pas proposé d'estimation de la gêne globale perçue à partir de la connaissance des gênes spécifiques.

Perspectives

Nous avons pu dessiner dans ce manuscrit certaines limites de notre outil de diagnostic de la gêne sonore. Dans la conception de l'outil lui-même, nous avons cité tout au long du manuscrit des pistes d'amélioration. Une des plus importantes est la question de la fiabilité du *tracking* (ou suivi de véhicule). En effet, cette étape est fondamentale dans la bonne es-

timation de la trajectoire du véhicule mais aussi dans l'estimation du nombre de véhicules. Une grande communauté de chercheurs travaillant sur ces thématiques, une amélioration de la robustesse devrait pouvoir être faite.

La formation de voies pourra aussi être améliorée sur plusieurs points. Le premier étant l'optimisation de la fonction de Green constituant le noyau de rétropropagation des signaux acoustiques. En effet, dans les différentes situations, la fonction de Green en champ libre a été utilisée. Or, la prise en compte de la géométrie de la rue doit permettre d'avoir une meilleure séparation des sources entre les voies de circulation. Par ailleurs, la résolution de l'antenne pourra être maintenue constante en fonction de la fréquence de sorte à garder tout le contenu fréquentiel de la source étudiée. Cela permettra d'avoir un signal audio encore plus fidèle à celui du véhicule, mais devrait aussi apporter des informations complémentaires à l'algorithme de classification et augmenter ses performances.

L'algorithme d'apprentissage machine aussi devrait pouvoir être amélioré. En plus de l'idée précédemment citée, l'apprentissage sur plus de données devrait rendre plus robuste la classification. De même, la classification du type de véhicule est étudiée depuis la fin des années 1990 dans la communauté du traitement d'image. L'image du véhicule ajoutée à la base de données d'apprentissage pourra éventuellement améliorer les performances du classifieur.

Par ailleurs, la prise de contact avec les auteurs des modèles de rugosité et de force de fluctuation a mis en évidence une volonté de travailler ensemble sur leur amélioration. Cela permettra sans doute une estimation de la gêne sonore plus proche de celle prévue par Morel *et al.*

Enfin, il est prévu, à la suite de cette thèse, une expérience perceptive qui aura plusieurs objectifs en utilisant les différentes bases de données obtenues durant cette thèse. Le premier sera de confirmer le modèle de calcul des gênes spécifiques en utilisant les données binaurales obtenues lors des mesures à La Ferté-Vidame avec nos modes de calculs des indices psychoacoustiques. Un deuxième objectif sera de proposer un modèle mixant les approches de Morel *et al.* [78] et de Klein *et al.* [59]. En effet, Klein *et al.* introduisent des indices intéressants qui semblent être davantage corrélés avec la gêne sonore ressentie, mais ils ne proposent qu'un seul modèle de gêne sonore globale alors qu'il serait possible de séparer suivant les catégories perceptives de Morel *et al.* pour affiner la précision du modèle. Enfin, la question du processus de sommation des gênes spécifiques pourra être investiguée en cherchant le meilleur modèle applicable à nos données. Cela pourra être fait grâce aux acquisitions de 10 minutes obtenues sur le quai Saint-Bernard à Paris, temps suffisant pour avoir un échantillon représentatif du paysage sonore [106, 36, 24].

D'autres expériences comme celle réalisée sur le quai Saint-Bernard pourront aussi être réalisées dans d'autres endroits de la ville. Cela permettra de diversifier la composition du trafic observée de sorte à alimenter le classifieur avec de nouvelles données, notamment pour les catégories 2 et 4. L'amélioration de la qualité du *tracking* pourra alors aussi être investiguée.

Vu le nombre croissant de véhicules électriques, leur influence sur les cartes de bruit commence à être questionnée [28]. Il sera intéressant d'intégrer ces véhicules électriques que nous avons à La Ferté-Vidame dans nos expériences perceptives de sorte à proposer une extension la catégorisation perceptive de Morel *et al.* D'autant plus que, pour un véhicule léger passant à vitesse constante supérieure à 35 km/h, la source dominante de bruit est le contact pneu/chaussée : le véhicule électrique émet donc un bruit de niveau sonore comparable au véhicule thermique (voir section 2.3.1). Mais qu'en est-il de la gêne perçue?

Bibliographie

- [1] Directive européenne 70/157/cee relatives au niveau sonore admissible et au dispositif d'échappement des véhicules à moteur, 1969. (Cité page 2)
- [2] Directive européenne 2002/49/ce relative à l'évaluation et à la gestion du bruit dans l'environnement, 2002. (Cité pages 3, 14 et 17)
- [3] Réseau permanent de mesure du bruit aérien – projet survol – bilan 2015. Rap. tech., Bruitparif, Dec. 2016. (Cité page 36)
- [4] Rapport d'activité 2016. Rap. tech., Bruitparif, juin 2017. (Cité page 36)
- [5] J. E. ADCOCK : *Optimal filtering and speech recognition with microphone arrays*. Thèse de doctorat, Brown University, 2001. (Cité page 60)
- [6] U. AHRLIN et R. RYLANDER : Annoyance caused by different environmental noises. *Journal of Sound and Vibration*, 66:459–462, 1979. (Cité page 24)
- [7] J. B. ALLEN et D. A. BERKLEY : Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979. (Cité page 40)
- [8] N. ALT, K. WOLFF, G. EISELE et F. PICHOT : Fahrzeugaußengeräuschsimulation (vehicle exterior noise simulation). *ATZ - Automobiltechnische Zeitschrift*, 108(10):832–836, 2006. (Cité pages 53 et 54)
- [9] H. AUGOYARD, Jean-François; Torgue : *A l'écoute de l'environnement : Répertoire des effets sonores*. Editions Parenthèses, 1995. (Cité page 9)
- [10] P. AUMOND, A. CAN, B. DE COENSEL, D. BOTTELDOOREN, C. RIBEIRO et C. LAVANDIER : Modeling soundscape pleasantness using perceptual assessments and acoustic measurements along paths in urban context. *Acta Acustica united with Acustica*, 103(3):430–443, 2017. (Cité page 5)
- [11] W. AURES : Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale. *Acta Acustica united with Acustica*, 59(2):130–141, 1985. (Cité page 31)
- [12] W. AURES : Ein berechnungsverfahren der rauigkeit (a procedure for calculating auditory roughness). *ACUSTICA*, 58:268–281, October 1985. (Cité pages 31 et 84)
- [13] W. BABISCH : Traffic noise and cardiovascular disease : Epidemiological review and synthesis. *Noise and Health*, 2(8):9–32, 2000. (Cité page 2)
- [14] W. BABISCH : Updated exposure-response relationship between road traffic noise and coronary heart diseases: a meta-analysis. *Noise and Health*, 16(68):1–9, 2014. (Cité page 2)

- [15] W. BABISCH, H. ISING, J. GALLACHER, P. ELWOOD, P. SWEETNAM, J. YARNELL, D. BAINTON et I. BAKER : Traffic noise and work noise and cardiovascular risk and factors and the caerphilly and speedwell collaborative and heart disease and studies. *Envirosmnt Intemationat*, vol. 16:pp. 425–435, 1990. (Cité page 24)
- [16] W. BABISCH, G. WÖLKE, J. HEINRICH et W. STRAFF : Road traffic noise and hypertension – accounting for the location of rooms. *Environmental Research*, 133:380–387, aug 2014. (Cité page 24)
- [17] J. A. BALLESTEROS, E. SARRADJ, M. D. FERNÁNDEZ, T. GEYER et M. J. BALLESTEROS : Noise source identification with beamforming in the pass-by of a car. vol. 93, p. 106–119. Elsevier BV, Jun 2015. (Cité pages 37 et 52)
- [18] J. BARKER, R. MARXER, E. VINCENT et S. WATANABE : The third ‘chime’ speech separation and recognition challenge: Analysis and outcomes. *Computer Speech & Language*, 46:605–626, 2017. (Cité page 38)
- [19] B. BERGLUND : Community noise in a public health perspective. *Dans INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 1998, p. 1683–1688. Institute of Noise Control Engineering, 1998. (Cité page 10)
- [20] B. BERGLUND, U. BERGLUND, M. GOLDSTEIN et T. LINDVALL : Loudness (or annoyance) summation of combined community noises. *The Journal of the Acoustical Society of America*, 70(6):1628–1634, Dec. 1981. (Cité pages 24 et 25)
- [21] B. BERGLUND, U. BERGLUND et T. LINDVALL : Scaling loudness, noisiness, and annoyance of aircraft noise. *The Journal of the Acoustical Society of America*, 57(4):930, 1975. (Cité page 24)
- [22] B. BERGLUND et M. NILSSON : Total annoyance and perceptually discernible noise sources. *Dans InterNoise2000*, 2000. (Cité pages 25 et 94)
- [23] J. L. BESNERAIS, V. LANFRANCHI, M. HECQUET, S. RÉCORBET, J. SAPENA et P. BROCHET : Characterization of the audible magnetic noise emitted by traction motors in railway rolling stock. *Dans InterNoise 2008*, Shanghai, China, 2008. (Cité page 141)
- [24] D. BOTTELDOOREN, C. LAVANDIER, A. PREIS, D. DUBOIS, I. ASPURU, C. GUASTAVINO, L. BROWN, M. NILSSON et T. C. ANDRINGA : Understanding urban and natural soundscapes. *Dans Forum Acusticum 2011 - Aalborg - Denmark*, 2011. (Cité page 117)
- [25] D. BOTTELDOOREN et A. VERKEYN : Fuzzy models for accumulation of reported community noise annoyance from combined sources. *J. Acoust. Soc. Am.*, 112(4):1496, 2002. (Cité page 25)
- [26] M. BOULEAU : La circulation routière en Île-de-france en 2010. Rap. tech., IAU Île-de-France, 2013. (Cité page 98)
- [27] T. F. BROOKS et W. M. HUMPHREYS : A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294(4-5):856–879, Jul 2006. (Cité page 43)
- [28] H. CAMPELLO-VICENTE, R. PERAL-ORTS, N. CAMPILLO-DAVO et E. VELASCO-SANCHEZ : The effect of electric vehicles on urban noise maps. *Applied Acoustics*, 116:59–64, jan 2017. (Cité page 118)

- [29] A. CAN, G. GUILLAUME et B. GAUVREAU : Noise indicators to diagnose urban sound environments at multiple spatial scales. *Acta Acustica united with Acustica*, 101(5): 964–974, Sep 2015. (Cité pages 5 et 6)
- [30] J. CAPON : High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, Aug 1969. (Cité page 43)
- [31] R. CHATAGNON et J. LELONG : Evaluation des émissions sonores de deux-roues motorisés en conditions urbaines de fonctionnement. *Acoustique et Techniques*, p. 15–20, 2014. (Cité page 59)
- [32] B. D. COENSEL et D. BOTTELDOOREN : A model of saliency-based auditory attention to environmental sound. *Dans International Congress on Acoustics, ICA*, Sydney, Australia, 2010. (Cité page 30)
- [33] C. COUVREUR, V. FONTAINE, P. GAUNARDA et C. GINETTE MUBIKANGIEYA : Automatic classification of environmental noise events by hidden markov models. *Applied Acoustics*, 54(3):187–206, July 1998. (Cité page 70)
- [34] P. DANIEL et R. WEBER : Psychoacoustical roughness and implementation of an optimized and model. *ACUSTICA*, 83(1):113–123, Jan/Feb 1997. (Cité pages 84, 85, 88 et 95)
- [35] B. DE COENSEL, D. BOTTELDOOREN, T. DE MUER, B. BERGLUND, M. E. NILSSON et P. LERCHER : A model for the perception of environmental sound based on notice-events. *J. Acoust. Soc. Am.*, 126(2):656–665, 2009. (Cité pages 28 et 30)
- [36] P. DELAITRE, C. LAVANDIER, E. D'HONDT, E. G. BOIX, K. KAMBONA, M. BASILE, L. CAZEAUX et K. I. C : Evaluation de l'agrément sonore en milieu urbain à l'aide de téléphones mobiles. *Dans Congrès Français d'Acoustique (CFA)*, 2014. (Cité page 117)
- [37] E. den BOER et A. SCHROTEN : Traffic noise reduction in europe : Health effects, social costs and technical and policy options to reduce road and rail traffic noise. Rap. tech. to, CE Delft, 2007. (Cité page 2)
- [38] H. FASTL et E. ZWICKER : *Psychoacoustics - Facts and Models*. Berlin : Springer Verlag, 3rd éd., 2007. (Cité pages 20, 21, 22, 23, 84, 86, 87, 88, 89, 90, 92 et 95)
- [39] S. FIDELL, D. S. BARBER et T. J. SCHULTZ : Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise. *J. Acoust. Soc. Am.*, 89(1):221–233, January 1991. (Cité page 24)
- [40] I. FLINDELL et P. J. STALLEN : Non-acoustical factors in environmental noise. *Noise and Health*, 1(3):11–16, 1999. (Cité page 26)
- [41] M. FLORENTINE : *Loudness*, vol. 37 de *Springer Handbook of Auditory Research*. Springer-Verlag New York, 2011. (Cité page 21)
- [42] L. FRITSCHI, A. L. BROWN, R. KIM, D. SCHWELA et S. KEPHALOPOULOS, éd. *Burden of disease from environmental noise. Quantification of healthy life years lost in Europe*. Num. support. WHO Regional Office for Europe, 2011. (Cité pages 2 et 24)
- [43] R. GARCÍA LEÓN : Modeling the sensation of fluctuation strength. Mémoire de Master, Eindhoven University of Technology, 2015. (Cité pages 84, 85, 88 et 95)
- [44] K. GENUIT et A. FIEBIG : Psychoacoustics and its benefit and for the and soundscape approach. *Acta Acustica United with Acustica*, 92:1–7, 2006. (Cité page 19)

- [45] E. GEORGE et M. SMITH : Speech analysis/synthesis modification using an analysis-by-synthesis/overlap-add. *IEEE T Speech Audi P*, 5(5):389–406, Sept. 1997. (Cité page 62)
- [46] D. GIANNOULIS, E. BENETOS, D. STOWELL, M. ROSSIGNOL, M. LAGRANGE et M. D. PLUMBLEY : Detection and classification of acoustic scenes and events: An ieee aasp challenge. *Dans 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, p. 1–4, Oct 2013. (Cité pages 71 et 75)
- [47] B. M. GIBBS et D. K. JONES : A simple image method for calculating the distribution of sound pressure levels within an enclosure. *Acta Acustica united with Acustica*, 26(1): 24–32, 1972. (Cité page 40)
- [48] L.-A. GILLE, C. MARQUIS-FAVRE et J. MOREL : Testing of the european union exposure-response relationships and annoyance equivalents model for annoyance due to transportation noises: The need of revised exposure-response relationships and annoyance equivalents model. *Environment International*, 94:83 – 94, 2016. (Cité page 24)
- [49] B. R. GLASBERG et B. C. J. MOORE : A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc*, 50(5):331–342, 2002. (Cité page 82)
- [50] I. GRIFFITHS et F. LANGDON : Subjective response and to road and traffic noise. *Journal of Sound and Vibration*, 8(1):16–32, 1968. (Cité page 19)
- [51] R. GUSKI, U. FELSCHER-SUHR et R. SCHUEMER : The concept of noise annoyance: How international experts see it. *Journal of Sound and Vibration*, 223(4):513–527, 1999. (Cité page 14)
- [52] I. HAFIZOVIC, C.-I. C. NILSEN, M. KJØLERBAKKEN et V. JAHR : Design and implementation of a MEMS microphone array system for real-time speech acquisition. *Applied Acoustics*, 73(2):132–143, Feb 2012. (Cité pages 38, 44, 60, 65 et 103)
- [53] K. HIRAMATSU : Soundscape - the concept and its significance in acoustics. *Dans ICA*, num. concept, Kyoto, 2004. (Cité page 7)
- [54] J. Y. HONG et J. Y. JEON : Soundscape mapping in urban contexts using gis techniques. *Dans InterNoise*, 2014. (Cité pages 7 et 8)
- [55] ISO : ISO 226:2003 - Acoustique - Lignes isosoniques normales. (Cité page 15)
- [56] B. JAKOVLJEVIC, K. PAUNOVIC et G. BELOJEVIC : Road-traffic noise and factors influencing noise annoyance in an urban population. *Environment International*, 35(3): 552–556, Apr 2009. (Cité page 24)
- [57] E. J. JUSTINO, F. BORTOLOZZI et R. SABOURIN : A comparison of svm and hmm classifiers in the off-line signature verification. *Pattern Recognition Letters*, 26(9):1377 – 1385, 2005. (Cité page 71)
- [58] S. KEPHALOPOULOS, M. PAVIOTTI et F. ANFOSSO-LÉDÉE : *Common noise assessment methods in Europe (CNOSSOS-EU)*. PUBLICATIONS OFFICE OF THE EUROPEAN UNION, 2012. (Cité page 56)
- [59] A. KLEIN, C. MARQUIS-FAVRE, R. WEBER et A. TROLLÉ : Spectral and modulation indices for annoyance-relevant features of urban road single-vehicle pass-by noises. *J. Acoust. Soc. Am.*, 137(3):1238–1250, Mar 2015. (Cité pages 24, 32, 33, 34, 69 et 117)

- [60] D. KOLLER, DANIILIDIS et H. H. NAGEL : Model-based object tracking in monocular image sequences of road traffic scenes. *Int. J. Comput. Vision*, 10(3):257–281, June 1993. (Cité page 70)
- [61] D. KOLLER, J. WEBER et J. MALIK : Robust multiple car tracking with occlusion reasoning. *Dans Computer Vision — ECCV'94*, p. 186–196, 1994. (Cité page 70)
- [62] P. LERCHER : Deviant dose-response curves for traffic noise in 'sensitive areas'? *Dans INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 1998, p. 710–713. Institute of Noise Control Engineering, 1998. (Cité page 10)
- [63] P. LERCHER, B. D. COENSEL, L. DEKONINK et D. BOTTELDOOREN : Community response to multiple sound sources: Integrating acoustic and contextual approaches in the analysis. *Int. J. Environ. Res. Public Health*, 14(6):663, 2017. (Cité page 24)
- [64] P. LUNDÉN, Östen AXELSSON et M. HURTIG : On urban soundscape mapping: A computer can predict the outcome of soundscape assessments. *Dans INTER-NOISE 2016*, Hambourg, 2016. (Cité page 72)
- [65] P. MAIJALA, Z. SHUYANG, T. HEITTOLA et T. VIRTANEN : Environmental noise monitoring using source classification in sensors. *Applied Acoustics*, 129:258 – 267, 2018. (Cité page 74)
- [66] P. MARMAROLI, J.-M. ODOBEZ, X. FALOURD et H. LISSEK : Pass-by noise acoustic sensing for estimating speed and wheelbase length of two-axle vehicles. *Dans ICA*, Montreal, 2013. Acoustical Society of America (ASA). (Cité page 36)
- [67] D. W. MARQUARDT : An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. (Cité page 86)
- [68] C. MARQUIS-FAVRE et J. MOREL : A simulated environment experiment on annoyance due to combined road traffic and industrial noises. *International Journal of Environmental Research and Public Health*, 12(7):8413–8433, 2015. (Cité page 25)
- [69] C. MARQUIS-FAVRE, E. PREMAT et D. AUBRÉE : Noise and its Effects – A Review on Qualitative Aspects of Sound. Part II: Noise and Annoyance. *Acta Acustica united with Acustica*, 91, 2005. (Cité page 25)
- [70] C. MARQUIS-FAVRE, E. PREMAT, D. AUBRÉE et M. VALLET : Noise and its Effects – A Review on Qualitative Aspects of Sound. Part I : Notions and Acoustic and Ratings. *Acta Acustica united with Acustica*, 91:613 – 625, 2005. (Cité page 19)
- [71] H. M. MIEDEMA et C. G. OUDSHOORN : Annoyance from transportation and noise: Relationships and with exposure and metrics and dnl and denl and their confidence and intervals. *Environmental Health Perspectives*, 109(4):409–416, April 2001. (Cité page 24)
- [72] H. M. E. MIEDEMA et H. VOS : Noise sensitivity and reactions to noise and other environmental conditions. *The Journal of the Acoustical Society of America*, 113(3):1492–1504, 2003. (Cité page 27)
- [73] F. MIETLICKI, C. MIETLICKI, C. RIBEIRO, P. GAUDIBERT, B. VINCENT et V. GISSINGER : www.noiseineu.eu: New tools to inform the public about environmental noise in cities and to assist decision-making. *Dans EuroNoise*, 2015. (Cité pages 27, 28 et 29)

- [74] C. MONCHATRE : Étude du trafic des deux-roues motorisés en Île-de-france. Rap. tech., CETE - Île-de-France, 2012. (Cité page 4)
- [75] J. J. MORÉ : The levenberg-marquardt algorithm: implementation and theory. Dans *Numerical analysis*, p. 105–116. Springer, 1978. (Cité page 86)
- [76] J. MOREL : *Caractérisation physique et perceptive du bruit routier urbain pour la définition d'indicateurs de gêne sonore en situation de mono-exposition et de multi-exposition en présence de bruit industriel*. Thèse de doctorat, Ecole Nationale des Travaux Publics de l'Etat, 2012. (Cité pages 25 et 26)
- [77] J. MOREL, C. MARQUIS-FAVRE, D. DUBOIS et M. PIERRETTE : Road traffic in urban areas: A perceptual and cognitive typology of pass-by noises. *Acta Acustica united with Acustica*, 98(1):166–178, Jan 2012. (Cité pages 10, 30, 31, 32, 34, 69 et 74)
- [78] J. MOREL, C. MARQUIS-FAVRE et L.-A. GILLE : Noise annoyance assessment of various urban road vehicle pass-by noises in isolation and combined with industrial noise: A laboratory study. *Applied Acoustics*, 101:47–57, Jan 2016. (Cité pages 24, 25, 31, 32, 33, 34, 69, 81, 82, 84, 93, 94, 95, 110 et 117)
- [79] J. B. OLLERHEAD : Predicting public reaction to noise from mixed sources. Dans *Inter-Noise 1978*, p. 579–584, 1978. (Cité page 25)
- [80] A. OSSES VECCHI, R. GARCÍA LEÓN et A. KOHLRAUSCH : Modelling the sensation of fluctuation strength. Dans *ICA*, 2016. (Cité pages 88, 89, 90 et 92)
- [81] M. PALLAS, R. CHATAGNON et J. LELONG : Noise emission assessment of a hybrid electric mid-size truck. *Applied Acoustics*, 76:280–290, Feb 2014. (Cité pages 56 et 58)
- [82] M. PAVIOTTI et K. VOGIATZIS : On the outdoor annoyance from scooter and motorbike noise in the urban environment. *Science of The Total Environment*, 430:223–230, Jul 2012. (Cité pages 31, 33, 34 et 94)
- [83] Y. PENE : *Antennerie numérique pour la caractérisation de sources aéroacoustiques en milieu complexe*. Thèse de doctorat, Université Pierre et Marie Curie, 2015. (Cité page 40)
- [84] R. PEPPIN et C. RODMAN : *Community Noise*. American Society for Testing and Materials, 1978. (Cité page 17)
- [85] J. POLACK, J. BEAUMONT, B. ROBIN, S. LESAUX, C. PRONELLO, C. CAMUSSO, C. ARRAS, D. BOZZETO et L. DROIN : Pertinence des descripteurs d'ambiance sonore urbaine : application aux bruits des transports pendant les périodes sensibles. Rap. tech., Programme "Bruit et nuisances sonores", 2007. (Cité page 10)
- [86] C. A. POWELL : A summation and inhibition model of annoyance response to multiple community noise sources. Technical Paper 1479, NASA, Sep 1979. (Cité page 25)
- [87] D. PRAJAPATI et H. J. GALIYAWALA : A review on moving object detection and tracking. *International Journal of Computer Application*, 5(3):168–175, April 2015. (Cité page 101)
- [88] A. RABAOUI, M. DAVY, S. ROSSIGNOL et N. ELOUZE : Using one-class svms and wavelets for audio surveillance. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 3(4), Dec. 2008. (Cité pages 70, 71 et 75)

- [89] P. RICCIARDI, P. DELAITRE, C. LAVANDIER, F. TORCHIA et P. AUMOND : Sound quality indicators for urban places in paris cross-validated by milan data. *The Journal of the Acoustical Society of America*, 138(4):2337–2348, oct 2015. (Cité page 5)
- [90] D. W. ROBINSON : The concept of noise pollution level. *Journal of Occupational Medicine*, 13(12):602, Dec 1971. (Cité page 19)
- [91] U. SANDBERG : Tyre/road noise – myths and realities. *Dans InterNoise 2001*, 2001. (Cité pages 55 et 56)
- [92] U. SANDBERG et J. A. EJSMONT : *Tyre/road noise reference book*. Informex (Kisa), 2002. (Cité page 53)
- [93] R. M. SCHAFER : *Le Paysage Sonore*. Wildproject, 4ème éd., 2010. (Cité pages 6, 7, 8, 10 et 14)
- [94] R. SCHMIDT : Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, Mar 1986. (Cité page 43)
- [95] W. E. SHOLES et J. W. SARGENT : Designing against noise from road traffic. *Applied Acoustics*, 4(3):203–234, July 1971. (Cité page 19)
- [96] B. SCHULTE-FORTKAMP : Soundscape- focusing on resources. *Dans ICA 2013*. Acoustical Society of America (ASA), 2013. (Cité pages 6 et 7)
- [97] T. J. SCHULTZ : Synthesis of social surveys on noise annoyance. *J. Acoust Soc. Am.*, 64(2):377–405, Aug. 1978. (Cité page 24)
- [98] A. SONTACCHI : Entwicklung eines modulkonzeptes für die psychoakustische geräuschanalyse unter matlab (Développement d'un concept modulaire pour l'analyse psychoacoustique du bruit sous MatLab). Mémoire de Master, Technischen Universität Graz, 1998. (Cité page 88)
- [99] R. SOTTEK : *Modelle zur Signalverarbeitung im menschlichen Gehör (Signal Processing Model of the Human Auditory System)*. Thèse de doctorat, 1993. Aachen, Techn. Hochsch. (Cité page 84)
- [100] R. SOTTEK et K. GENUIT : Models of signal processing in human hearing. *AEU - International Journal of Electronics and Communications*, 59(3):157 – 165, 2005. (Cité page 84)
- [101] S. S. STEVENS : Calculation of the loudness of complex noise. *The Journal of the Acoustical Society of America*, 28(5):807–832, 1956. (Cité page 82)
- [102] T. N. TAN, D. SULLIVAN et D. BAKER : Pose determination and recognition of vehicles in traffic scenes. *Dans Computer Vision — ECCV'94*, p. 501–506, 1994. (Cité page 70)
- [103] S. TAYLOR : A comparison of models to predict annoyance reactions to noise from mixed sources. *Journal of Sound and Vibration*, 81(1):123–138, 1982. (Cité page 25)
- [104] J. TERROIR, B. DE COENSEL, D. BOTTELDOOREN et C. LAVANDIER : Activity interference caused by traffic noise: Experimental determination and modeling of the number of noticed sound events. *Acta Acustica united with Acustica*, 99(3):389–398, 2013. (Cité page 30)
- [105] A. J. TORIJA, D. P. RUIZ, B. D. COENSEL, D. BOTTELDOOREN, B. BERGLUND et A. RAMOS-RIDAO : Relationship between road and railway noise annoyance and overall indoor

- sound exposure. *Transportation Research Part D: Transport and Environment*, 16(1): 15–22, Jan 2011. (Cité page 117)
- [106] A. TROLLÉ, C. MARQUIS-FAVRE et A. KLEIN : Short-term annoyance due to tramway noise: Determination of an acoustical indicator of annoyance via multilevel regression analysis. *Acta Acustica united with Acustica*, 2014. (Cité pages 31 et 32)
- [107] X. VALERO et F. ALIAS : Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys. *Archives of Acoustics*, 37(4), jan 2012. (Cité pages 71, 75, 107 et 108)
- [108] M. VALLET, J.-M. GAGNEUX, V. BLANCHET, B. FAVRE et G. LABIALE : Long term sleep disturbance due to traffic noise. *Journal of Sound and Vibration*, 90(2):173–191, Sep. 1983. (Cité page 19)
- [109] C. VANWYNSBERGHE : Réseaux à grand nombre de microphones : applicabilité et mise en œuvre. Thèse de doctorat, Université Paris VI, 2016. (Cité pages 46 et 116)
- [110] C. VANWYNSBERGHE, P. CHALLANDE, J. MARCHAL, R. MARCHIANO et F. OLLIVIER : A robust and passive method for geometric calibration of large arrays. *The Journal of the Acoustical Society of America*, 139(3):1252–1263, mar 2016. (Cité page 60)
- [111] C. VANWYNSBERGHE, R. MARCHIANO, F. OLLIVIER, P. CHALLANDE, H. MOINGEON et J. MARCHAL : Design and implementation of a multi-octave-band audio camera for realtime diagnosis. *Applied Acoustics*, 89:281–287, Mar 2015. (Cité pages 45 et 52)
- [112] D. VIENNEAU, C. SCHINDLER, L. PEREZ, N. PROBST-HENSCH et M. RÖÖSLI : The relationship between transportation noise exposure and ischemic heart disease: A meta-analysis. *Environmental Research*, 138:372–380, apr 2015. (Cité page 2)
- [113] E. VINCENT : Séparation de sources: Quand l’acoustique rencontre le machine learning. *Dans 13e Congrès Français d’Acoustique*, 2016. (Cité page 38)
- [114] M. VORLÄNDER : Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America*, 86(1):172–178, 1989. (Cité page 40)
- [115] J. VOS : Annoyance caused by simultaneous impulse, road-traffic, and aircraft sounds: a quantitative model. *The Journal of the Acoustical Society of America*, 91(6): 3330–3345, 1992. (Cité pages 24 et 25)
- [116] J. VOS : On total annoyance caused by different environmental sounds: A review and suggestions for additional research. *Dans InterNoise 2000*, 2000. (Cité page 24)
- [117] J. VOS : Noise annoyance caused by mopeds and other traffic sources. *Dans INTER-NOISE*, 2006. (Cité page 31)
- [118] E. WEINSTEIN, K. STEELE, A. AGARWAL, et J. GLASS : Loud: A 1020-node modular microphone array and beamformer and for intelligent. 2004. (Cité pages 37 et 44)
- [119] E. WEINSTEIN, K. STEELE, A. AGARWAL et J. GLASS : Loud: A 1020-node microphone array and acoustic beamformer. *Dans ICSV14 Cairns • Australia 9-12 July*, 2007. (Cité page 60)
- [120] E. ZWICKER : Ein verfahren zur berechnung der lautstärke (une procédure pour calculer la sonie). *Acustica*, 10:304–308, 1960. (Cité page 82)

-
- [121] E. ZWICKER et H. FASTL : *Psychoacoustics: Facts and Models*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999. (Cité pages [82](#) et [83](#))

Annexe

Localisation de sources sur deux-roues et véhicules lourds

Sommaire

Véhicules lourds	129
Vitesses constantes	129
Accélération et décélération	132
Deux-roues motorisés	134
Vitesses constantes	134
Accélération et décélération	137
Deux-roues électrique	140

Dans cette annexe, les résultats d'imagerie obtenus lors des essais à La Ferté-Vidame (piste d'essais de PSA Peugeot-Citroën) (section 2.3.2) sont détaillés pour les deux-roues (DR-Es-1, DR-Es-2 et DR-E1) et le véhicule lourd (VL-D). L'expérience est décrite en section 2.3 (page 50).

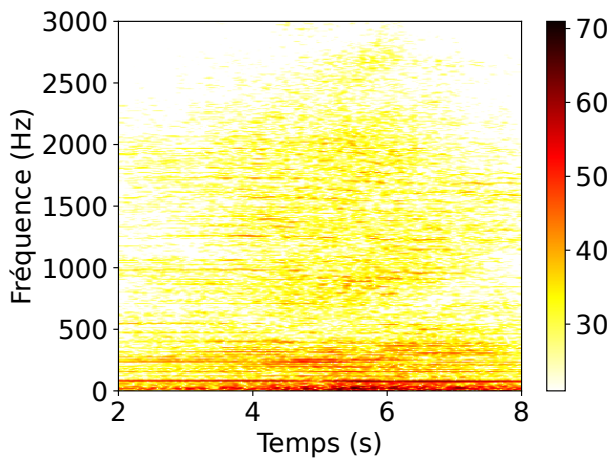
Véhicules lourds

On étudie dans cette section les sources principales de bruit générées par le véhicule utilitaire (considéré comme un véhicule lourd). Il est à noter que, sur ce véhicule, le pot d'échappement est situé juste derrière la cabine de pilotage.

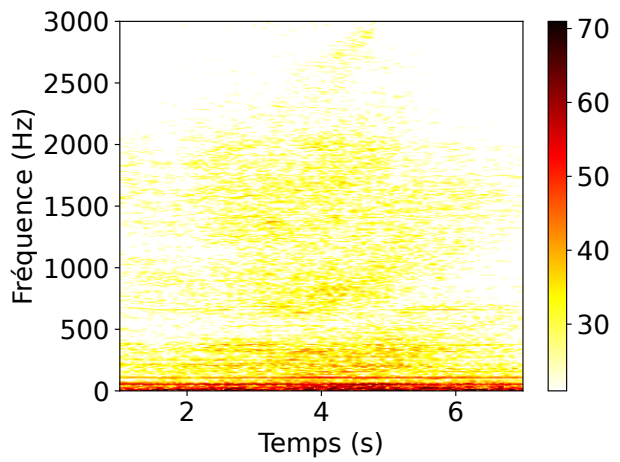
Vitesses constantes

Passage à 25 km/h. Les passages des véhicules légers (section 2.3) ont été enregistrés en deuxième rapport et en régime moteur normal. Or à cette vitesse, l'utilitaire est en surrégime dans le même rapport. Les résultats obtenus par formation de voie en deuxième et troisième rapport sont présentés figure 10 et 11 respectivement.

Sur les spectrogrammes des signaux de référence (figures 10a et 11a) on constate une nette diminution du niveau des composantes tonale entre le deuxième et le troisième rapport. On peut aussi voir que l'énergie concentrée sur les très basses fréquences (inférieures à 200 Hz) est équivalente. Cela est confirmé par les figures 10b et 11b qui montrent que le niveau sonore de la source moteur est conservé. Pour l'octave de 250 Hz (figure 10c et 11c) une différence de 6 dB du niveau de la source maximale est à noter. Cela est dû au fait que



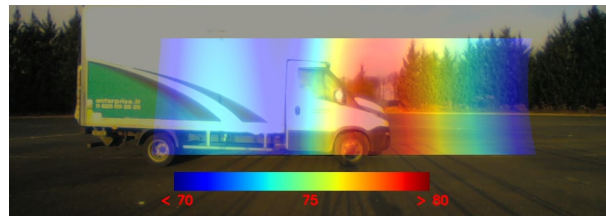
(a) Spectrogramme du microphone de référence - Dynamique : 50 dB.



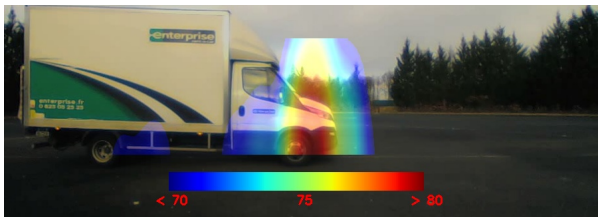
(a) Spectrogramme du microphone de référence - Dynamique : 50 dB.



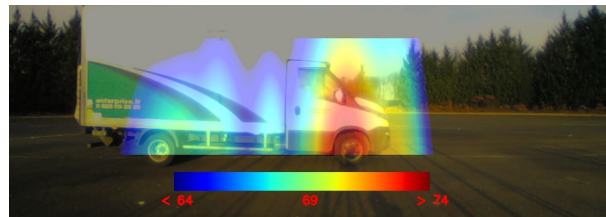
(b) Image acoustique - Octave de 125 Hz



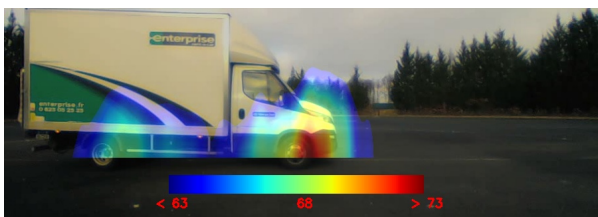
(b) Image acoustique - Octave de 125 Hz



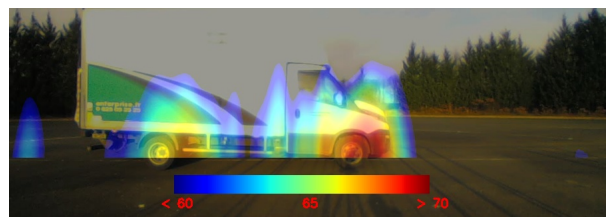
(c) Image acoustique - Octave de 250 Hz



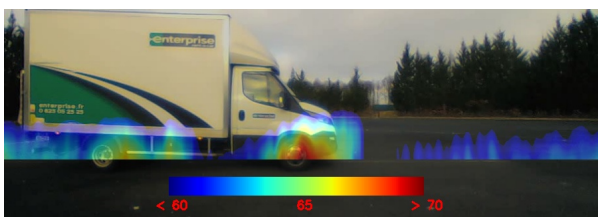
(c) Image acoustique - Octave de 250 Hz



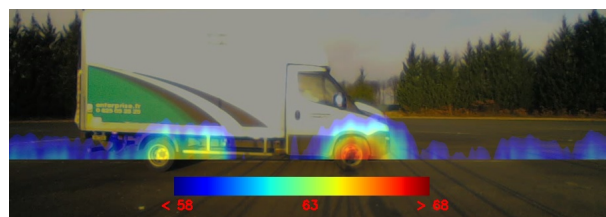
(d) Image acoustique - Octave de 500 Hz



(d) Image acoustique - Octave de 500 Hz



(e) Image acoustique - Octave de 1 kHz



(e) Image acoustique - Octave de 1 kHz

FIGURE 10 – VL-D - Passage à 25 km/h - 2nd rapport

FIGURE 11 – VL-D - Passage à 25 km/h - 3^{ème} rapport

la source moteur est plus faible en 3^{ème} rapport et que le contact pneu/chaussée est alors la source dominante de bruit. Son niveau maximum à 74 dB est mesuré à l'essieu avant. On peut noter aussi l'apparition en troisième rapport d'une source au niveau de l'échappement (à l'arrière de la cabine) mesurée à -5 dB de la source principale à 250 et 500 Hz (figures 11c et 11d). En hautes fréquences les sources dans les deux rapports diffèrent peu et sont dominées par les contacts pneu/chaussée.

Passage à 50 km/h. A cette vitesse, le spectrogramme (figure 12a) met en évidence un contenu large bande avec quelques composantes tonales en basses fréquences (inférieures à 300 Hz). Le véhicule lourd passe ici en 4^{ème} rapport (les sources de bruit des véhicules légers ont été étudiées en 3^{ème} rapport), ce qui induit un régime moteur bas et donc un niveau de la source moteur en basses fréquences équivalent au passage à 25 km/h (80 dB sur l'octave de 125 Hz, voir figure 12d).

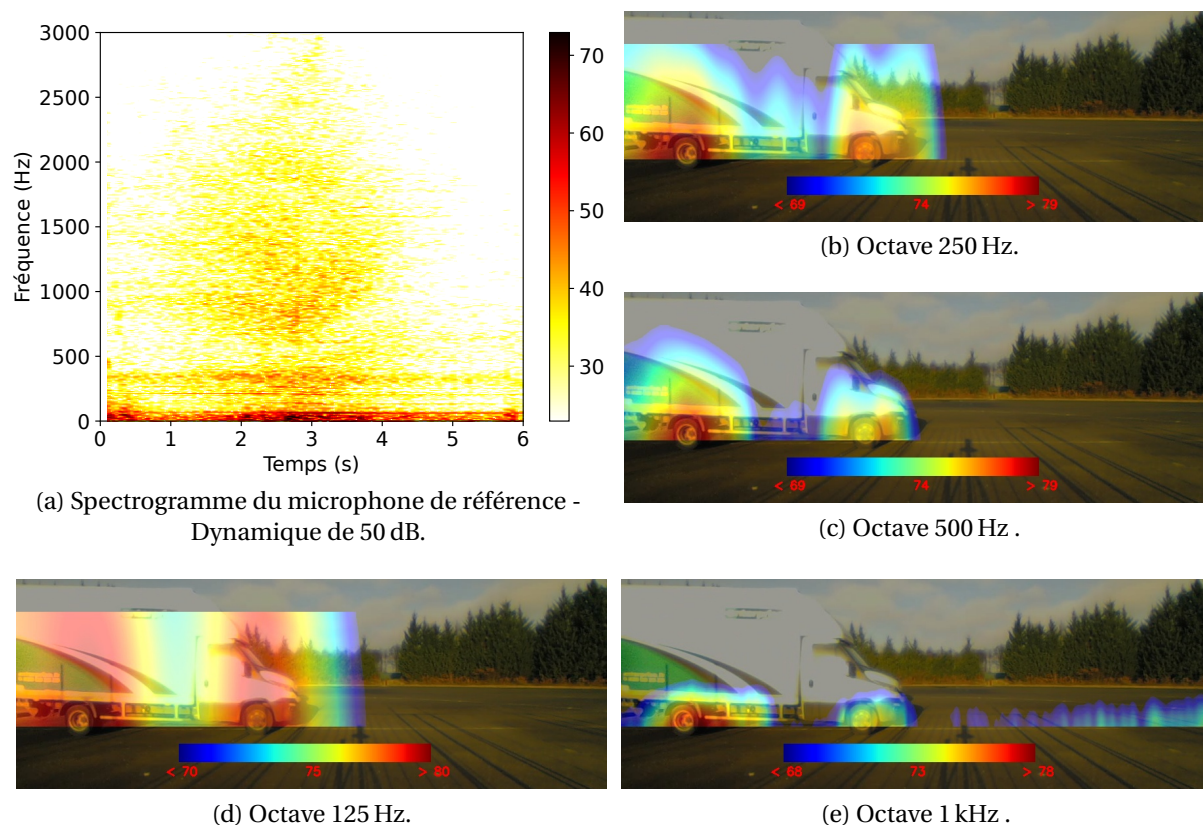


FIGURE 12 – Spectrogramme et images acoustiques lors d'un passage à 50 km/h du VL-D - 4^{ème} rapport.

On peut voir sur toute la plage fréquentielle étudiée une nette augmentation du bruit de contact pneu/chaussée. En effet, cette source croît de 74 à 80 dB à 125 Hz lorsque la vitesse est doublée ou encore de 65 à 78 dB pour la roue arrière à 1 kHz. On confirme aussi grâce aux figures 12c et 12e que la source moteur n'est pas la source principale de bruit à cette allure. De même, l'échappement qui rayonne à 74 dB en basses fréquences, est en deçà de 10 dB des sources pneumatiques en hautes fréquences et n'apparaît donc pas figure 12e.

Accélération et décélération

On s'intéresse ici au rayonnement acoustique du VL-D pour les deux autres allures étudiées : l'accélération et la décélération. La figure 13 présente les spectrogrammes du microphone de référence situé au centre de l'antenne. On remarque lors de l'accélération (figure 13a) la présence de composantes tonales dont les fréquences évoluent avec le régime moteur, principalement en basses fréquences. A plus hautes fréquences, ces composantes sont moins présentes et se confondent à un bruit large bande.

Lors de la décélération (spectrogramme en figure 13b), on peut noter que l'énergie acoustique se concentre sur les basses fréquences (inférieures à 500 Hz). On voit aussi une baisse en fréquence des composantes tonales jusqu'au débrayage qui intervient à $t = 6,5$ s (le véhicule étant à l'arrêt à $t = 8$ s).

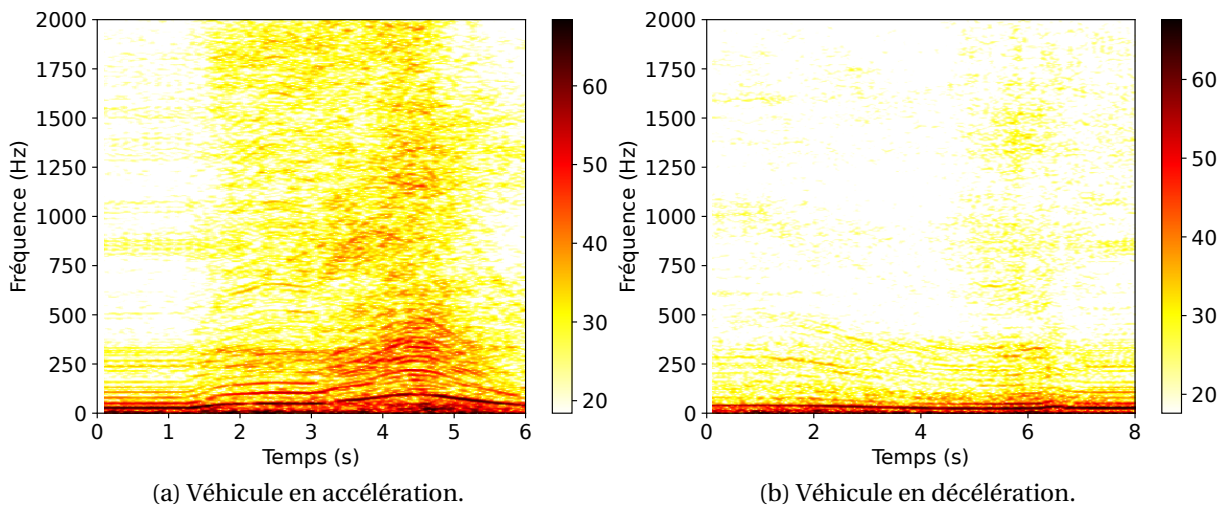


FIGURE 13 – Spectrogramme du microphone de référence - Yveco Daily en accélération et décélération -Dynamique : 50 dB.

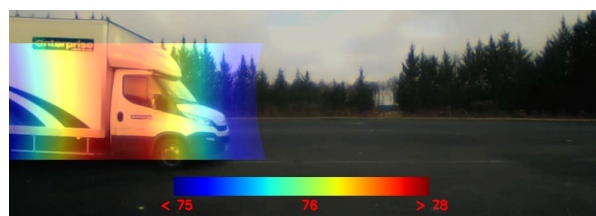
Accélération. La figure 14 présente les résultats pour le véhicule en accélération à $t = 2,5$ s (voir spectre en figure 13a). On peut voir que, sur la bande d'octave de 62,5 Hz (figure 14a), la source principale de bruit est le pot d'échappement mesurée à 89 dB. A plus hautes fréquences (125 Hz et 250 Hz, respectivement figures 14b et 14c) on note que le moteur est la source principale de bruit. On peut voir qu'à 500 Hz et 1 kHz (respectivement figures 14d et 14e) la source moteur est toujours présente mais le contact pneu/chaussée sur l'essieu avant produit un bruit d'un niveau équivalent.

Décélération. La figure 15 concerne le véhicule lourd en décélération à $t = 7$ s, donc après le débrayage qui a été déclenché tôt par rapport au point d'arrêt lors de cet essai. On note un niveau important en très basses fréquences (octave de 62,5 Hz, figure 15a) dont la source est localisée au niveau de l'échappement. Sur les octaves plus hautes fréquence on constate

la présence toujours importante de la source moteur, dont le niveau diminue lorsque la fréquence augmente. Les sources pneumatiques sont aussi présentes mais ne sont jamais dominantes, sauf sur la bande d'octave de 1 kHz (figure 15e).



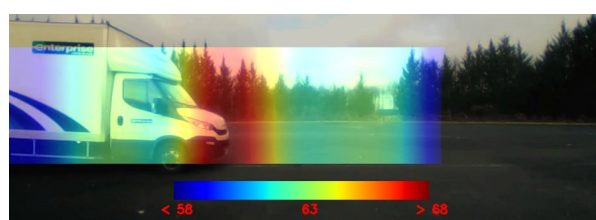
(a) Octave de 62,5 Hz - Dynamique : 3 dB.



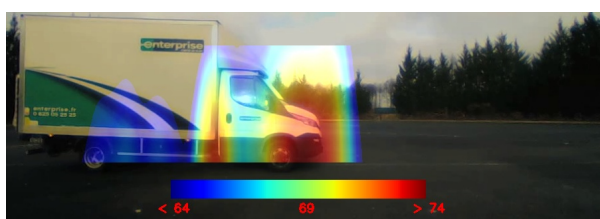
(a) Octave de 62,5 Hz - Dynamique : 3 dB.



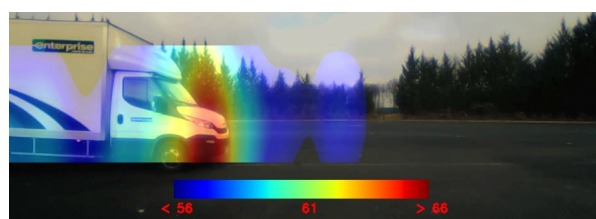
(b) Octave de 125 Hz - Dynamique : 10 dB.



(b) Octave de 125 Hz - Dynamique : 10 dB.



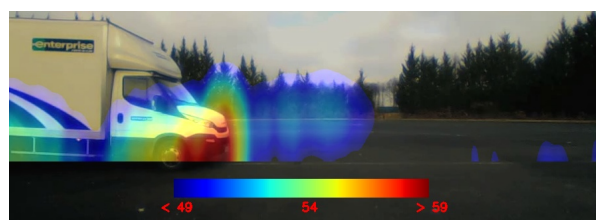
(c) Octave de 250 Hz - Dynamique : 10 dB.



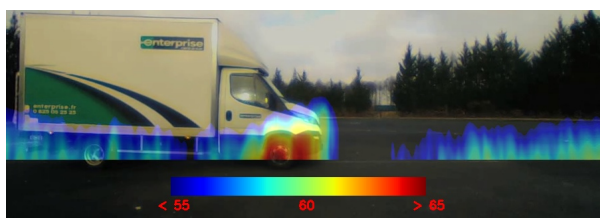
(c) Octave de 250 Hz - Dynamique : 10 dB.



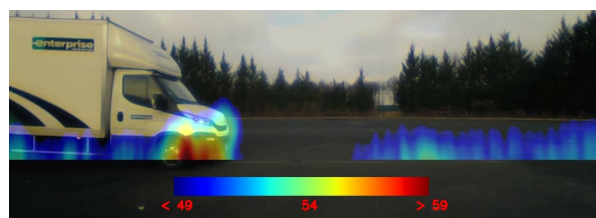
(d) Octave de 500 Hz - Dynamique : 10 dB.



(d) Octave de 500 Hz - Dynamique : 10 dB.



(e) Octave de 1 kHz - Dynamique : 10 dB.



(e) Octave de 1 kHz - Dynamique : 10 dB.

FIGURE 14 – Images acoustiques sur bandes d'octave - VL-D en accélération - $t = 2,5$ s.

FIGURE 15 – Images acoustiques sur bandes d'octave - VL-D en décélération - $t = 7$ s.

Deux-roues motorisés

Dans cette partie nous présentons les résultats pour les deux-roues disponibles pour l'expérience : un scooter thermique de 50 cm³ (DR-Es-1), un scooter thermique de 400 cm³ (DR-Es-2) et un scooter électrique équivalent 50 cm³ (DR-E1). Les images acoustiques sont ici présentées sur une dynamique de 3, 6 ou 10 dB en fonction des situations pour localiser la source principale de bruit pour chacun des véhicules aux différentes allures avec précision. Les deux-roues thermiques sont analysés pour les différentes allures dans un premier temps et les résultats pour le DR-E1 sont ensuite brièvement présentés.

Vitesses constantes

L'analyse des sources de bruit est faite dans un premier temps sur le DR-Es-2. La comparaison est ensuite faite avec le DR-Es-1 (50 cm³).

Passages à 25 km/h. Le spectrogramme du microphone de référence pour le passage du deux-roues de grosse cylindrée (figure 16a) montre un nombre important de composantes tonales jusqu'à 1000 Hz et un contenu fréquentiel très large bande lorsque le véhicule passe devant le microphone.

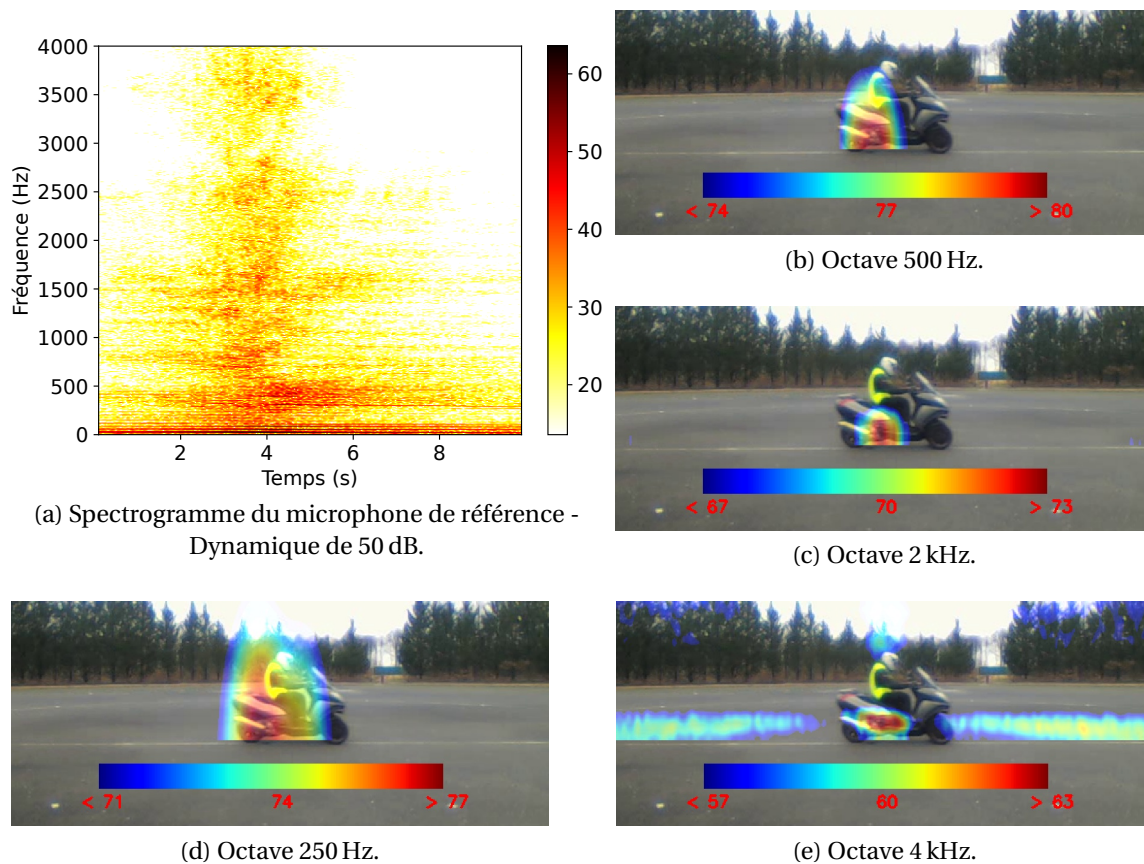


FIGURE 16 – Spectrogramme et images acoustiques du DR-Es-2 passant à 25 km/h - Dynamique : 6 dB.

Les images acoustiques, présentées sur les bandes d'octaves de 250 Hz à 4 kHz (figure 16b à 16e) montrent une seule et même source de bruit principale : le moteur. On mesure un niveau maximum (80 dB) sur la bande de 500 Hz et un niveau encore élevé à 4 kHz. Dans cette dernière bande fréquentielle le niveau est mesuré à 63 dB. En comparaison, le V1-Es (véhicule léger) a un niveau maximum de 39 dB sur cette même plage fréquentielle.

Il est à noter que sur la figure 16e, les “sources” à l’avant et à l’arrière du véhicule sont des artefacts de reconstruction dus à un sous-échantillonnage spatial du champ de pression acoustique par l’antenne.

Passages à 50 km/h. Le spectrogramme de la figure 17a à un même niveau inférieur (10 dB) que celui de la figure 16a mais avec un niveau maximum augmenté à 75 dB. On peut y voir que l’ensemble des composantes spectrales sont augmentées (aussi bien tonales que large bande). Sur les images acoustiques présentées en figures 17b à 17e, on peut remarquer la présence de deux sources sur tout le spectre : le pot d’échappement et le moteur. On peut aussi noter que le niveau est assez constant jusqu’à la bande d’octave de 1 kHz.

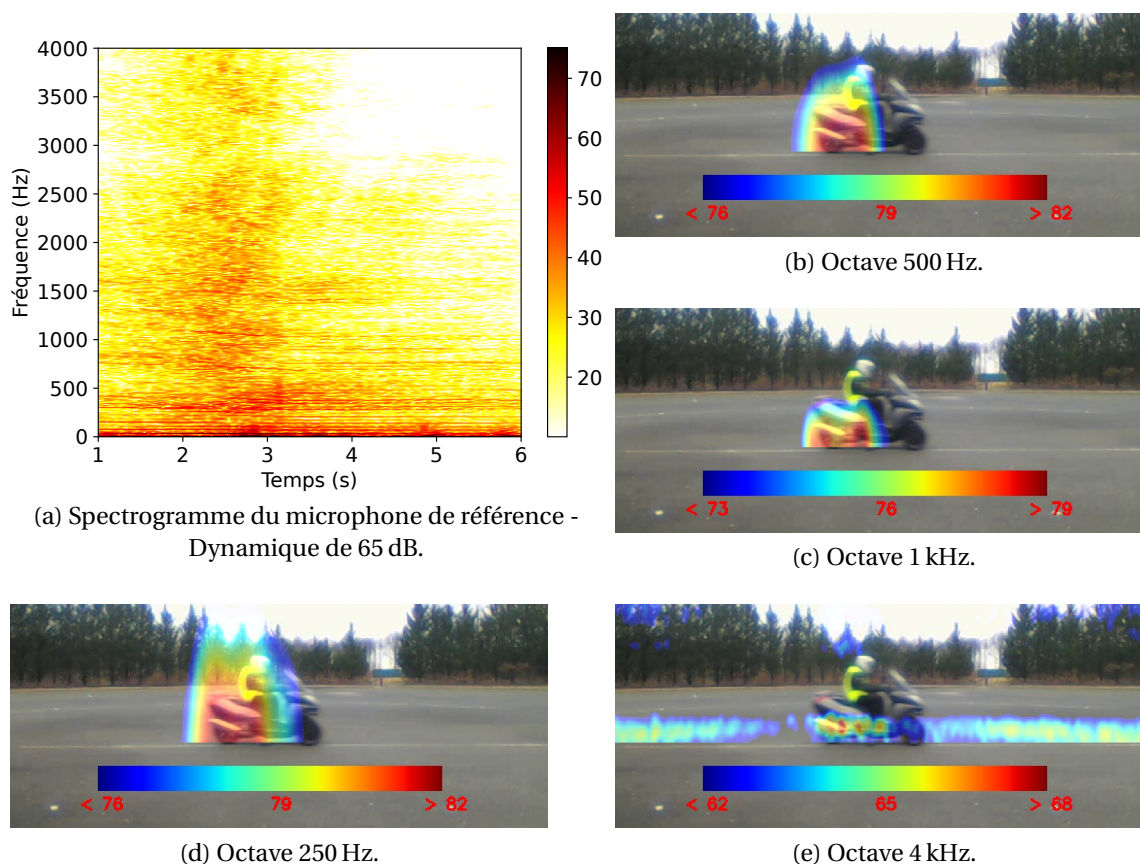


FIGURE 17 – Spectrogramme et images acoustiques du DR-Es-2 passant à 50 km/h - Dynamique : 6 dB.

Petite cylindrée. Le rayonnement acoustique du DR-Es-1 est ici étudié pour les mesures acquises aux deux vitesses constantes. La figure 18 montre que ce scooter a un contenu spectral beaucoup plus riche en composantes tonales que celui du DR-Es-2 (figures 16a et 17a). On constate une augmentation de 3 dB du niveau maximum entre les spectrogrammes des deux vitesses testées. Par ailleurs, le bruit large bande est plus important à 50 km/h quand le véhicule est proche du centre de l'antenne.

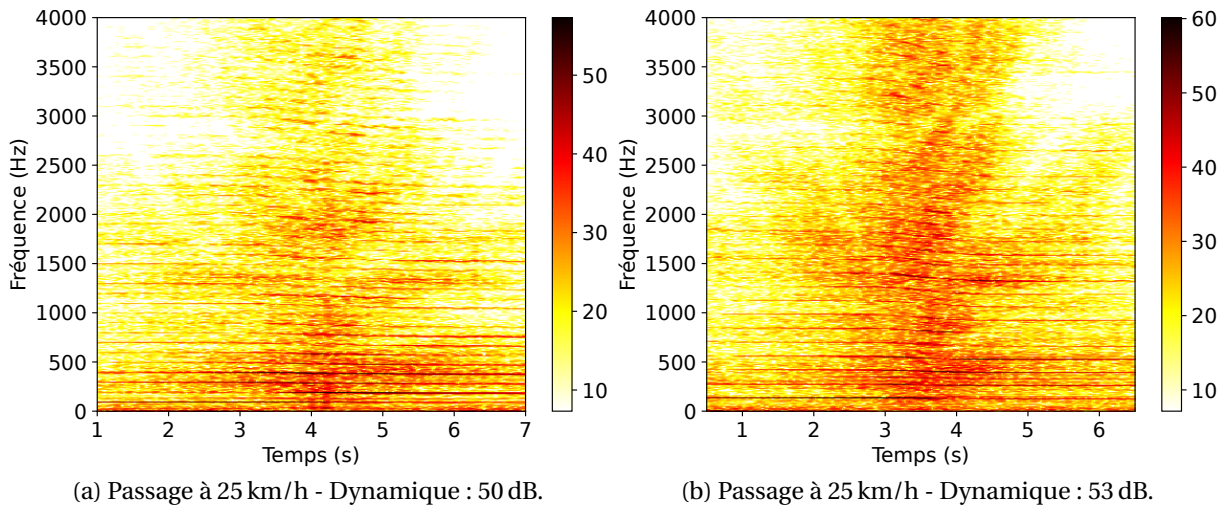


FIGURE 18 – Spectrogramme du microphone de référence - DR-Es-1 (50 cm³) à 25 et 50 km/h .

La figure 19 présente l'évolution fréquentielle de la répartition spatiale des sources de bruit à 25 km/h. Le pot d'échappement est la source principale sur l'ensemble du spectre. Pour les octaves de 250 et 500 Hz ce véhicule est plus bruyant (respectivement 83 et 84 dB mesurés) que le DR-Es-2. En revanche en hautes fréquences l'inverse est observé.

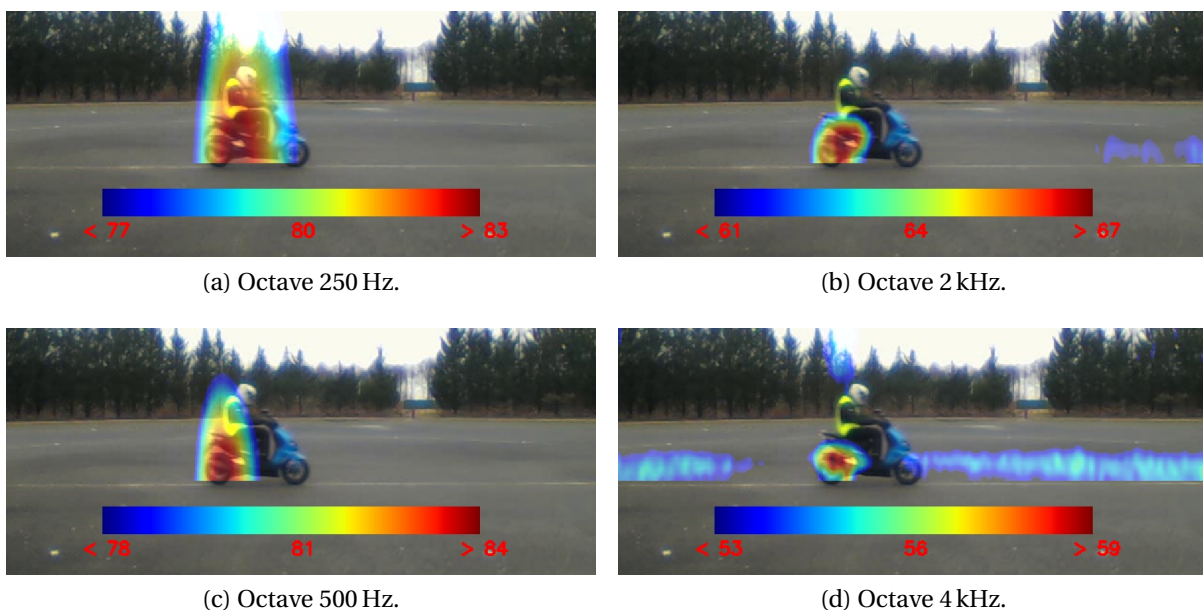


FIGURE 19 – Images acoustiques du DR-Es-1 (50 cm³) passant à 25 km/h - Dynamique : 6 dB.

En augmentant la vitesse à 50 km/h on peut voir (sur la figure 20) que la source moteur est présente en plus de l'échappement. L'absence de bruit pneumatique peut s'expliquer par le poids du véhicule et la largeur des pneus, tout deux plus faibles que sur un véhicule léger ou lourd. Sur la figure 17a, on peut aussi noter l'absence de composantes tonales en dessous de 100 Hz, ce qui se traduit par une distribution homogène de l'énergie acoustique, donc sans source de bruit localisée, sur l'octave de 62,5 Hz.

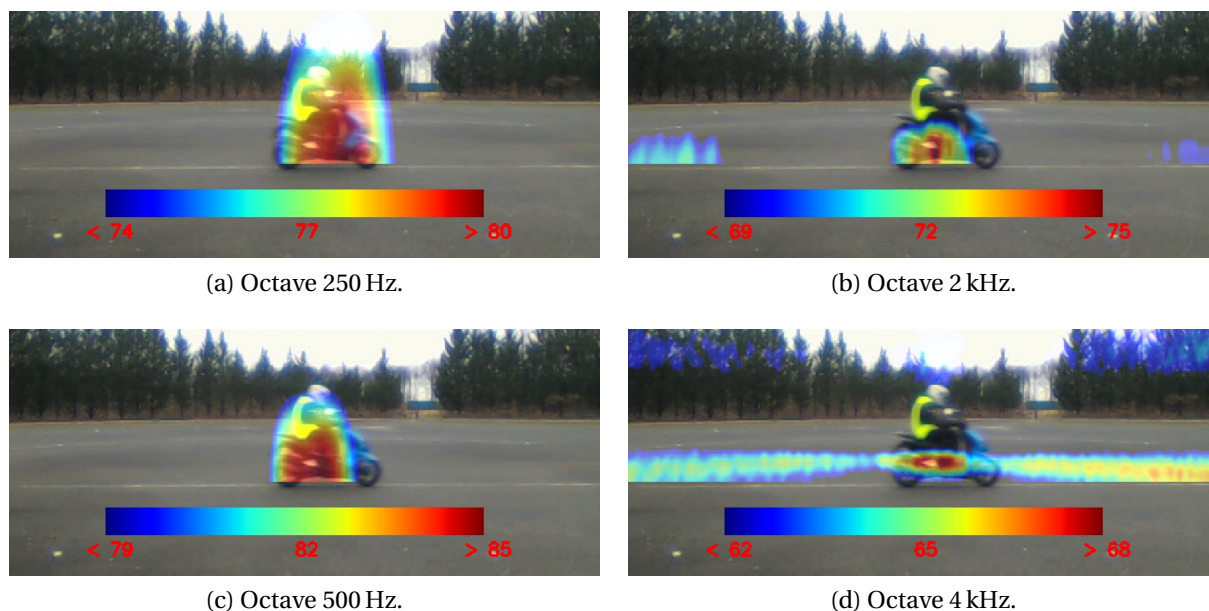


FIGURE 20 – Images acoustiques du DR-Es-1 (50 cm³) passant à 50 km/h - Dynamique : 6 dB.

Accélération et décélération

L'évolution du contenu spectral lors des accélérations et décélérations du DR-Es-2 est étudiée figure 21. On peut voir sur la figure 21a, où le véhicule accélère à partir de 12 s, des composantes tonales très présentes jusqu'à 2 kHz. Il y a aussi de l'énergie plus haut en fréquence mais sans avoir de fréquence émergente.

Lors de la décélération (figure 21b), ce deux-roues rayonne très peu d'énergie en hautes fréquences, le bruit émis se concentre principalement dans les très basses fréquences (inférieures à 500 Hz).

Accélération. La figure 22 présente les sources de bruit mises en évidence pendant la phase d'accélération (à $t=13,2$ s) sur 3 dB de dynamique. On peut donc voir que la source de bruit dominante est, sur de nombreuses bandes fréquentielles, le pot d'échappement. Cependant, sur les octaves de 250, 1000 et 2000 Hz (respectivement sur les figures 22e, 22b et 22d) il semble que la source principale soit localisée à l'avant, au niveau du bloc moteur. La capacité résolutive de l'antenne augmentant en fréquence, on peut voir avec précision les sources présentes sur les octaves de 4 et 8 kHz (figures 22f et 22h) : la sortie du pot d'échappement et le bloc moteur. On voit ici que l'énergie acoustique est encore très importante en hautes fréquences avec un niveau mesuré à 73 dB à 8 kHz.

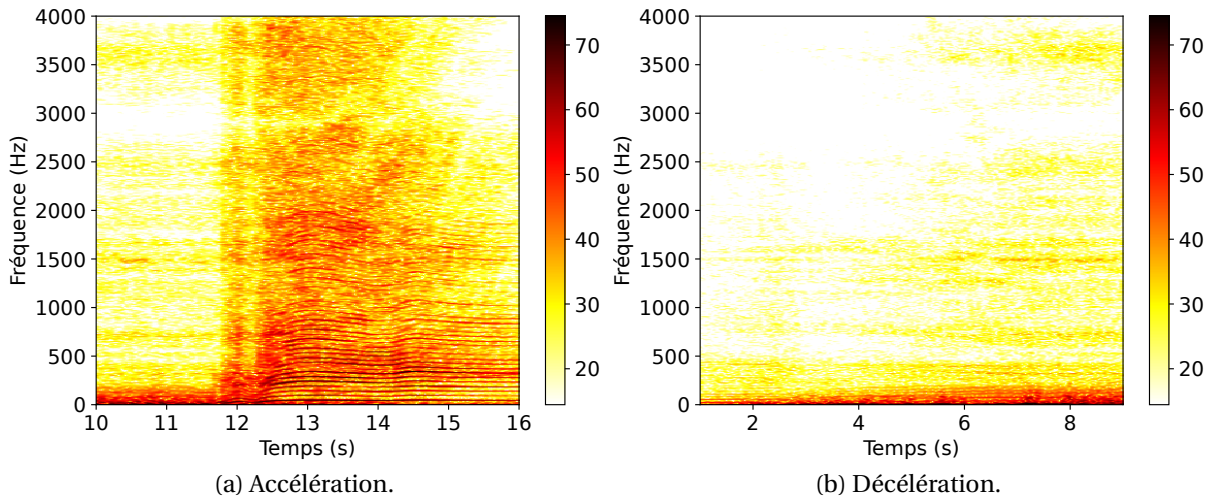


FIGURE 21 – Spectrogramme du microphone de référence lors du passage du DR-Es-2 (400 cm³) en accélération et décélération – Dynamique : 60 dB.

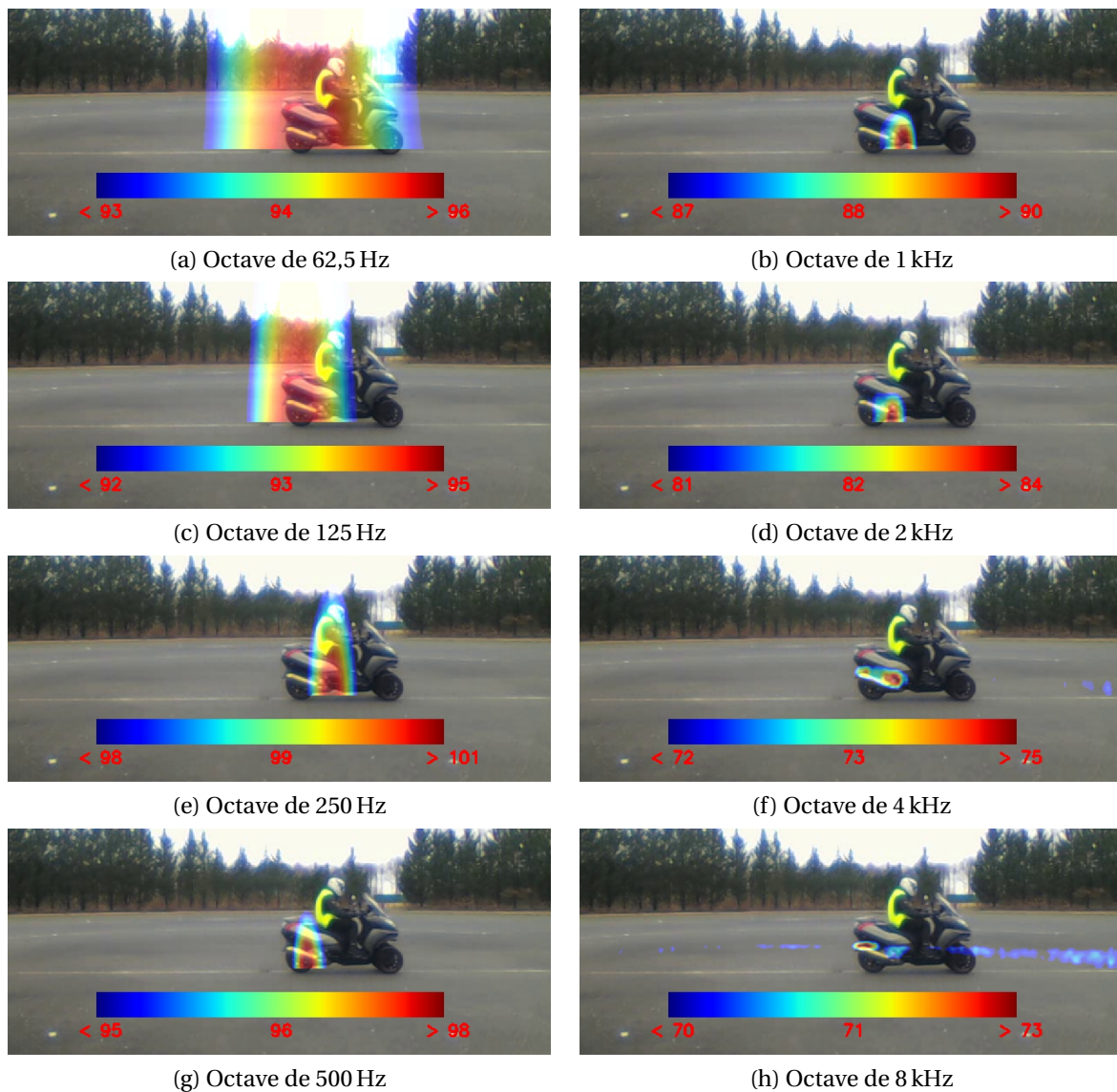


FIGURE 22 – Images acoustiques sur plusieurs bandes d'octaves du DR-Es-2 (400 cm³) – Accélération – Dynamique : 3 dB.

Décélération. Les images présentées figure 23 sont calculées à $t = 5,6$ s (sur la figure 21b) soit environ 5 m avant l'arrêt du véhicule. On peut voir ici que sur tout le spectre, la source principale de bruit est l'échappement. On peut aussi noter en hautes fréquences (figure 23d) la présence des freins comme source de bruit secondaire.

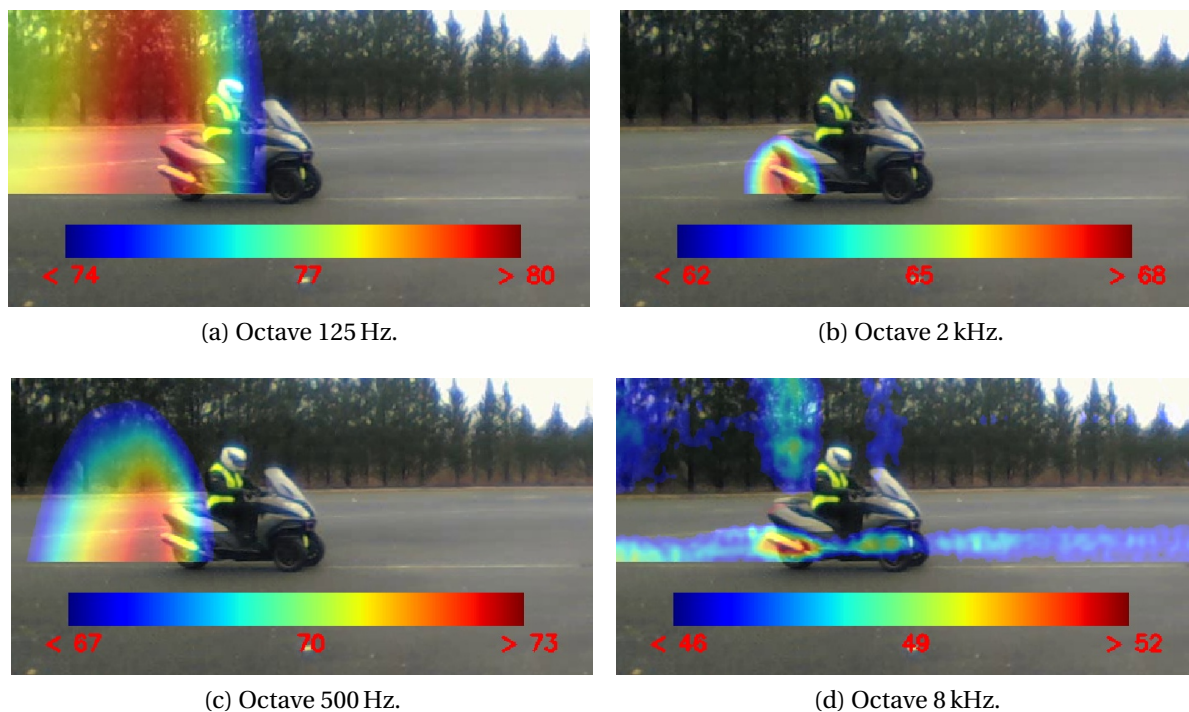


FIGURE 23 – Images acoustiques sur plusieurs bandes d'octaves du DR-Es-2 (400 cm^3) – Dynamique : 6 dB.

Petite cylindrée. Pour le DR-Es-1 (50 cm^3), les mêmes sources de bruit ont été globalement observées mais ayant des niveaux plus faibles d'environ 8 dB. Cette diminution globale est aussi visible sur les spectrogrammes où apparaissent des composantes tonales émergentes du bruit large bande aussi bien à la décélération qu'à l'accélération sur une large bande fréquentielle.

Asymétrie de rayonnement. Nous avons présenté en section 2.3.2 des résultats d'imagerie intégrée sur une large bande fréquentielle (de 0 à 2000 Hz) montrant une asymétrie du rayonnement acoustique sur le DR-Es-1 due à la présence du pot d'échappement d'un côté du véhicule. Ces résultats sont présentés de nouveau figure 24.

La comparaison peut alors être faite avec le deux-roues ayant une cylindrée de 400 cm^3 (figure 25). On peut remarquer que cette asymétrie est aussi présente et équivalente avec une différence, là encore, de 3 dB.

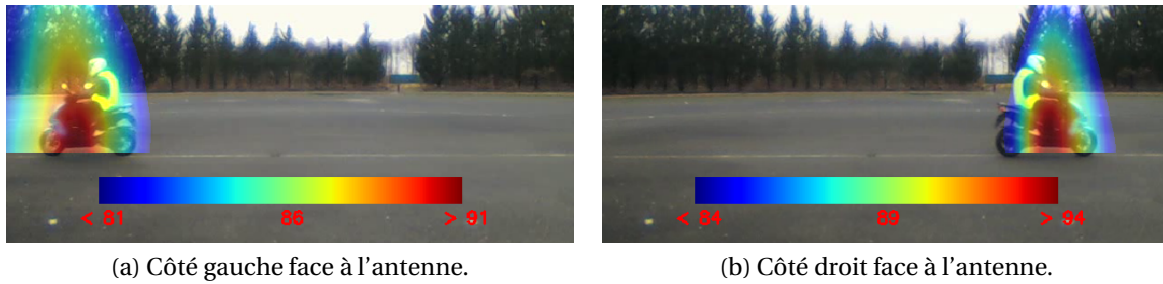


FIGURE 24 – Images acoustiques entre 0 et 2 kHz du DR-Es-1 (50 cm³) – Accélération – Dynamique : 10 dB.

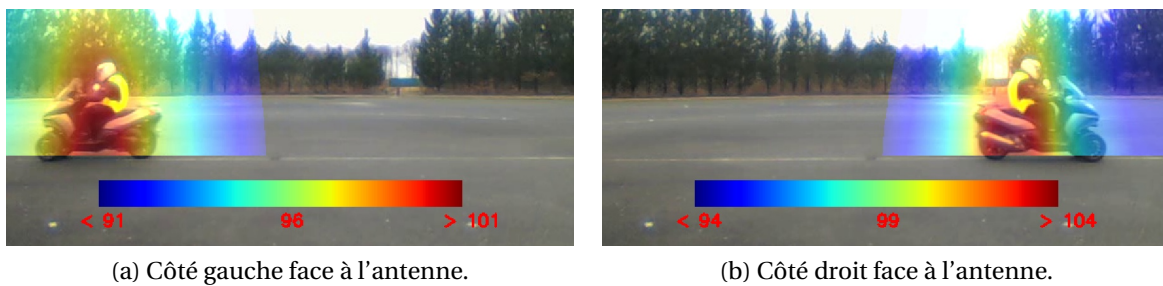


FIGURE 25 – Images acoustiques entre 0 et 2 kHz du DR-Es-2 (400 cm³) – Accélération – Dynamique : 10 dB.

Deux-roues électrique

On s'intéresse ici au DR-E1, le deux-roues électrique utilisé lors de ces essais à La Ferté-Vidame. Dans cette section, l'analyse des sources de bruit est présentée à partir de spectrogrammes disponibles en figure 26. En effet, les images acoustiques n'apportent pas plus d'information.

La figure 26a présente le spectrogramme du son émis par le deux-roues électrique passant à 25 km/h. On peut noter un bruit large bande qui est dû au contact pneu-chaussé. Les composantes tonales sont émises par le bloc moteur. En augmentant la vitesse à 50 km/h (figure 26b), on constate une augmentation du bruit pneumatique qui devient la source dominante au-dessus de 1 kHz. Plus bas en fréquence, c'est le moteur qui est la source principale de bruit. Pendant la phase d'accélération, le spectrogramme du scooter (figure 26c) présente de nombreuses harmoniques dont les fréquences augmentent avec la vitesse. On peut aussi noter des composantes fréquentielles continues à 500, 1000 et 2000 Hz. Lors de la phase de décélération (voir figure 26d) on observe les mêmes sources de bruit que pour le passage à 50 km/h : la source moteur en dessous de 1 kHz et les contacts pneu/chaussée au-delà.

Les figures présentées jusqu'ici se limitent à des fréquences relativement basses. La figure 27 fait apparaître un comportement inattendu en très hautes-fréquences. On observe, sur le spectrogramme présenté en figure 27a centré sur 16 kHz, des composantes tonales énergétiques dont la fréquence augmente pour certaines et diminue pour d'autres avec l'augmentation de la vitesse du deux-roues. Ce comportement est visible dans une moindre mesure autour de 8 kHz. En reprenant la figure 26c dans le détail, on voit aussi trois branches au-

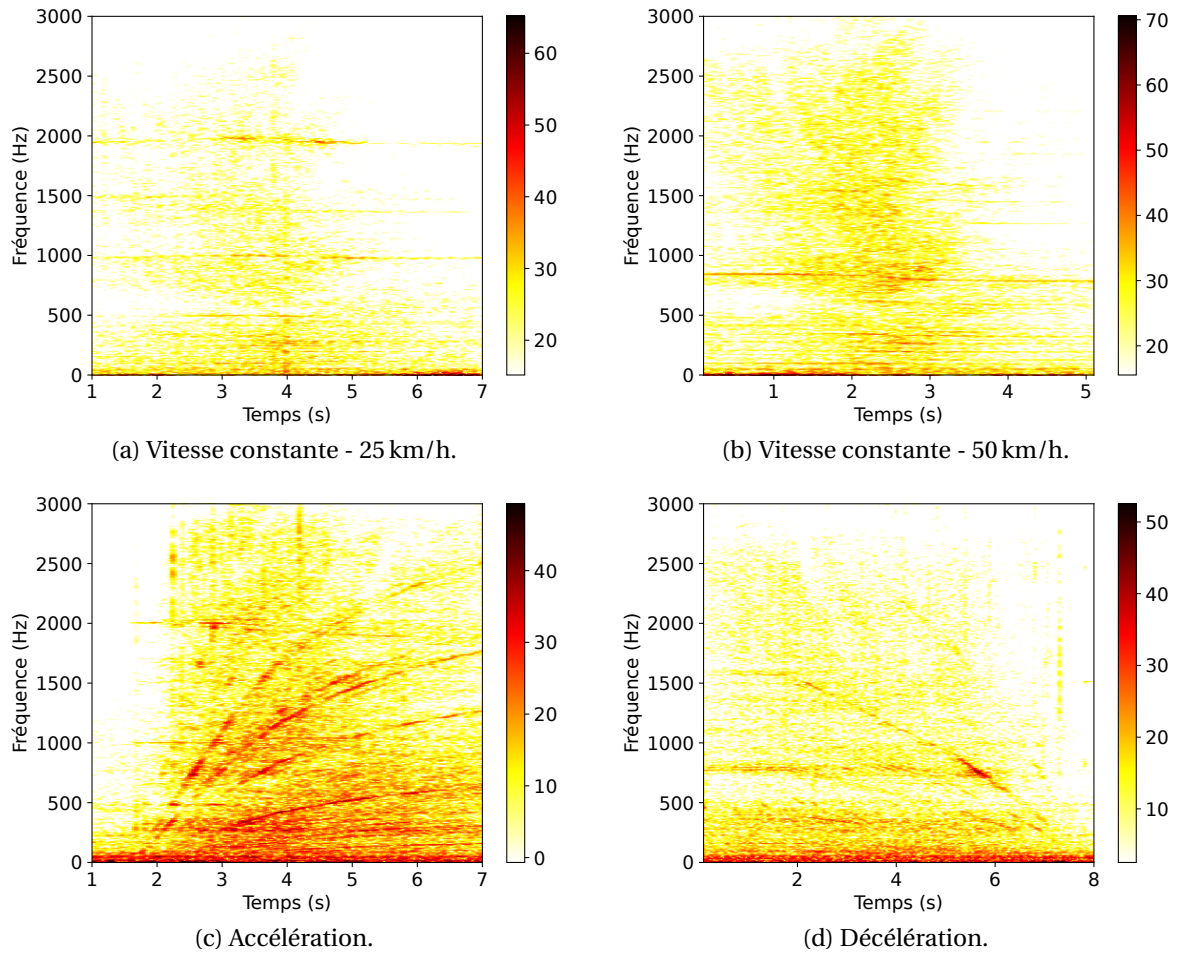


FIGURE 26 – Spectrogrammes du microphone de référence pour différentes allures du DR-E1 - Dynamique : 50 dB.

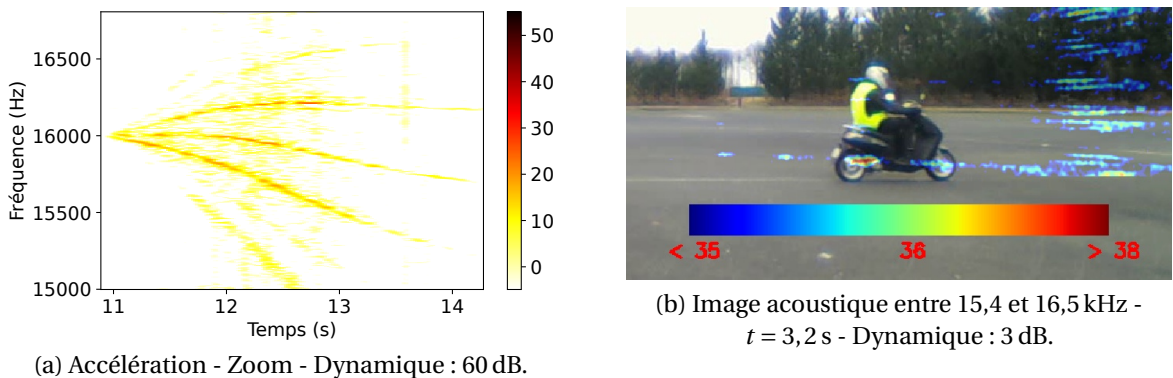


FIGURE 27 – Etude de composantes hautes-fréquences du bruit émit par le DR-E1.

tours de 2 kHz, mais pas autours de 1000 et 500 Hz, peut-être à cause du niveau important du bruit de fond. L'image acoustique intégrée sur la bande [15,4 , 16,5 kHz] est présentée en figure 27b. On peut voir que cette source de bruit vient du moteur. On note que le repliement spatial est important à ces fréquences et que la source est dupliquée de nombreuses fois, notamment horizontalement. Ces harmoniques sont présentes pour les moteurs électriques à modulation de largeur d'impulsion (PWT, *pulse-width modulation*) [23].

Index

- Acuité, 21
- Gêne sonore, 24
 - estimation, 93
 - estimation *in situ*, 110
 - facteurs influençant, 25
 - modèle de Klein *et al.*, 31
 - modèle de Morel *et al.*, 30
 - spécifique, 24
 - totale, 25
- Formation de voies (*beamforming*), 40
 - identification de source, 52
 - limites, 42
 - performances, 40
 - résolution, 41
 - séparation de source, 60
- Bruit
 - carte, 4
 - conséquences sanitaires, 1
 - lois, directives, 2
 - observatoires, 5
- Centre de gravité spectral, 22
- Force de fluctuation, 22
 - modèle, 88
- Megamicos*
 - antennes, 47, 49, 98
 - système, 45, 51
- MFCC, 71
- Apprentissage machine
 - (*machine learning*), 69
 - matrice de confusion, 75
 - noyaux (*kernel*), 73, 105
 - paramètres pertinents, 70
 - SVM, 72
 - vecteur de pondération, 78, 107
- Niveau
 - équivalent, 14
 - jour, soirée, nuit, 17
 - pondéré, 16
 - statistique, 18
- Paysage sonore (*Soundscape*)
 - cartographie, 7
 - concept, 6
- Rugosité, 23
 - modèle, 84
 - modèle théorique, 23
- Sonie, 20
 - courbes d'iso-sonie, 16
 - ISO 532B, 82
 - modèle, 82
 - niveau de, 20
 - partielle, 20