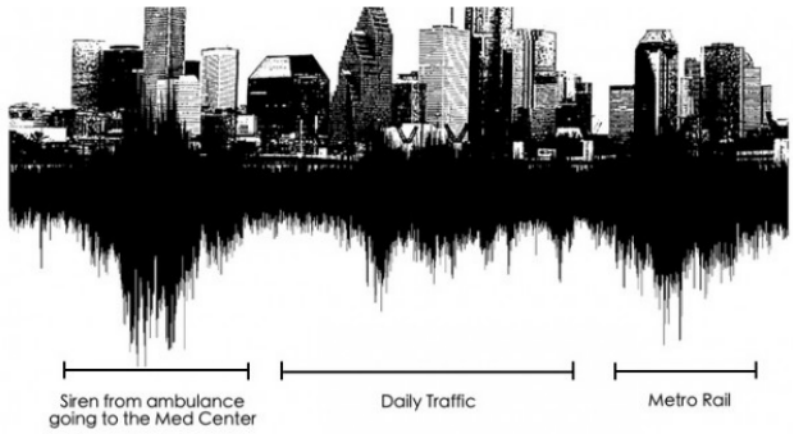
GRÉGOIRE LAFAY

ANALYSE DE SCÈNES SONORES ENVIRONNEMENTALES

[2 septembre 2016 at 4:43 – classicthesis version 4.2]

ANALYSE DE SCÈNES SONORES ENVIRONNEMENTALES

De l'étude perceptive à l'analyse automatique: une approche transdisciplinaire



[1]

GRÉGOIRE LAFAY

Doctorant

Équipe Analyse et Décision en Traitement du Signal et des Images (ADTSI)
Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)
École Centrale de Nantes, France

Décembre 2016

1 issu de http ://www.joesdaily.com/art-design/your-words-turned-into-art/

Grégoire Lafay : *Analyse de scènes sonores environnementales,* De l'étude perceptive à l'analyse automatique: une approche transdisciplinaire, © Décembre 2016

cite.
cite.

— TOTO & TOTO

Dedicated to TOTO.

date – date

ABSTRACT

TODO

RÉSUMÉ

TODO

PUBLICATIONS

Attention : *This requires a separate run of* `bibtex` *for the* `refsections`*, e.g.,* `theseLafay20161-blx`*,* `theseLafay20162-blx` *and* `theseLafay20163-blx` .

Analyse perceptive des scènes sonores environnementales

Lafay, Grégoire, Mathias Rossignol, Nicolas Misdariis, Mathieu Lagrange, and Jean-François Petiot (2014). "A new experimental approach for urban soundscape characterization based on sound manipulation : A pilot study." In: *International Symposium on Musical Acoustics*.

Lafay, Grégoire, Mathieu Lagrange, Jean-François Petiot, Mathias Rossignol, and Nicolas Misdariis (2016). "Investigating soundscapes perception through acoustic scenes simulation. Part I : simulation protocol presentation and case study." In: *The Journal of the Acoustical Society of America, (under revision)*.

— (2016). "Investigating soundscapes perception through acoustic scenes simulation. Part II : complementary experiments." In: *The Journal of the Acoustical Society of America, (under revision)*.

Lafay, Grégoire, Nicolas Misdariis, Mathieu Lagrange, and Mathias Rossignol (2016). "Semantic browsing of sound databases without keywords." In: *Journal of the Audio Engineering Society*.

Lagrange, Mathieu, Grégoire Lafay, Boris Defreville, and Jean-Julien Aucouturier (2015). "The bag-of-frames approach : a not so sufficient model for urban soundscapes." In: *The Journal of the Acoustical Society of America, express letter* 138.5, EL487–EL492.

Analyse automatique de scènes sonores environnementales

Lostanlen, Vincent, Grégoire Lafay, Joakim Anden, and Mathieu Lagrange (2016). "Object-based Auditory Scenes Similarity Retrieval and Classification With Wavelet Scattering."

Simulation automatique de scènes sonores — Utilisation de scènes sonores simulées pour l'évaluation des algorithmes en détection automatique d'événements sonores

Benetos, E., G. Lafay, M. Lagrange, and M. D. Plumbley (2016). "Detection of overlapping acoustic events using a temporally-constrained probabilistic model." In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6450–6454. DOI: 10.1109/ICASSP.2016.7472919.

Benetos, Emmanouil, Grégoire Lafay, and Mathieu Lagrange (2016). "Polyphonic Sound Event Tracking using LDS."

Lafay, G., M. Lagrange, E. Benetos, M. Rossignol, and A. Roebel (2016). "A Morphological Model for Simulating Acoustic Scenes and Its Application to Sound Event Detection." In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* PP.99, pp. 1–1. ISSN: 2329-9290. DOI: 10.1109/TASLP.2016.2587218.

Lagrange, Mathieu and Grégoire Lafay (2016). "Results of the DCASE 2016 challenge."

Rossignol, Mathias, Mathieu Lagrange, Grégoire Lafay, and Emmanouil Benetos (2015). "Alternate level clustering for drum transcription." In: *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, pp. 2023–2027.

Rossignol, Mathias, Grégoire Lafay, Mathieu Lagrange, and Nicolas Misdariis (2015). "SimScene : a web-based acoustic scenes simulator." In: *1st Web Audio Conference (WAC)*.

Cite

— TOTO

ACKNOWLEDGMENTS

TODO

TABLE DES MATIÈRES

I PRÉAMBULE 1

1 INTRODUCTION 3

2 MOTIVATION 5

II ANALYSE SENSORIELLE 7

3 LA PERCEPTION DE L'ENVIRONNEMENT SONORE 9

3.1 Le traitement de l'information auditive 9

3.1.1 Perception et cognition 9

3.1.2 Processus *Bottum-up* et processus *Top-down* 13

3.1.3 L'écoute 14

3.1.4 L'approche Écologique 14

3.1.5 La chaîne traitement de l'information auditive 16 de

3.2 Représentation mentale de l'environnement sonore 17

3.2.1 La notion de catégorie 18

3.2.2 Le processus de catégorisation 19

3.2.3 Organisation de la structure catégorielle 20

3.2.4 Théories de la catégorisation 21

3.2.5 Similarité et catégorisation 23

3.2.6 Catégorisation et contexte sensoriel 23

3.3 Analyse de scènes acoustiques 23

3.3.1 Définition 23

3.3.2 Une approche psychoacoustique 24

3.3.3 Régularités et processus primitifs 25

3.3.4 Perception de la forme 26

3.3.5 Flux auditif et stratégie de groupement 29

3.3.6 Attention et perception 31

3.3.7 L'approche par les neurosciences 31

3.4 L'étude des paysages sonores 31

3.4.1 La notion de paysage sonore 31

3.4.2 Application à la nuisance sonore urbaine 33

3.4.3 Approches catégorielle et dimensionnelles 34

3.4.4 Analyse et Modélisation des propriétés perceptives des paysages sonores 42

3.4.5 Catégoriser les sources et paysages sonores 43

3.4.6 Classifier les sources et environnements sonores 47

3.4.7 Prendre en compte les contributions séparées des différentes sources sonores 50

3.5 Événements et textures sonores 52

3.5.1 Définition 52

3.5.2 Percevoir les textures 53

3.5.3 Période d'attention 55

4 UN MODÈLE MORPHOLOGIQUE DE SCÈNES SONORES ENVIRONNEMENTALES 59
4.1 Motivations 59
4.1.1 Analyse sensorielle 59
4.1.2 Analyse automatique 62
4.2 Proposition d'un modèle de scènes sonores 62
4.2.1 Discrétiser l'environnement sonore 62
4.2.2 Description du modèle morphologique 66
4.3 Du modèle à la simulation : l'analyse sensorielle 70
4.3.1 Simulation et analyse sensorielle 70
4.3.2 Protocole expérimental basé sur la simulation 71
4.3.3 Paramètres de contrôle 73
4.3.4 Interface de sélection aveugle des sons isolés 75
4.3.5 Interface de simulation : l'outil *Simscene* 76
4.4 Du modèle à la simulation : l'analyse automatique 77
5 APPLICATION À L'ÉTUDE PERCEPTIVE DES ENVIRONNEMENTS SONORES URBAINS 79
5.1 Introduction 79
5.2 L'impact de la composition des scènes sur la perception de l'agrément 80
5.2.1 Objectif 80
5.2.2 Banque de données de sons isolés 81
5.2.3 Typologie des sources sonores présentes dans l'environnement urbain 81
5.2.4 Acquisition des sons isolés 82
5.2.5 Planification expérimentale 84
5.2.6 Données et méthodes d'analyses 87
5.2.7 Validité écologique de l'expérience 91
5.2.8 Vérification de l'agrément des scènes simulées 92
5.2.9 Étude comparative entre les descripteurs structurels 92
5.2.10 Influence des descripteurs structurels sur l'agrément perçu 95
5.2.11 Étude comparative entre les descripteurs sémantiques 97
5.2.12 Étude des espaces de représentations induits par les descripteurs sémantiques 101
5.2.13 L'influence spécifique des marqueurs sonores sur l'agrément perçu 102
5.2.14 Discussions 105
5.3 Agir sur l'agrément perçu en modifiant la composition sémantique 108
5.3.1 Objectif de l'expérience 108
5.3.2 Planification expérimentale 108
5.3.3 Données et méthodes d'analyses 108
5.3.4 Discussions 108

5.4 L'impact de la composition sur les processus de catégorisation des scènes 108
5.4.1 Objectif de l'expérience 108
5.4.2 Planification expérimentale 108
5.4.3 Données et méthodes d'analyses 108
5.4.4 Discussions 108

III ANALYSE AUTOMATIQUE 111
6 L'ANALYSE AUTOMATIQUE DES SCÈNES SONORES ENVIRONNEMENTALES, UN ÉTAT DE L'ART 113
7 APPLICATION DU MODÈLE MORPHOLOGIQUE À L'ÉTUDE DES PERFORMANCES DES ALGORITHMES D'ANALYSE AUTOMATIQUE DES SCÈNES SONORES ENVIRONNEMENTALES 115
7.1 De l'utilisation de scènes simulées en analyse automatique 115
8 APPLICATION DU MODÈLE MORPHOLOGIQUE À L'ANALYSE AUTOMATIQUE DES ENVIRONNEMENTS SONORES 117

IV CONCLUSIONS ET PERSPECTIVES 119
9 CONCLUSIONS 121
10 PERSPECTIVES 123

V APPENDICE 125 S
A OUTILS D'ANALYSE STATISTIQUE UNI-VARIÉE 127
A.1 Test paramétriques à deux populations 127
A.2 Test paramétriques à plus de deux populations 127
A.3 Mesures de corrélation paramétrique 127
A.4 Régression linéaire multiple 127
B OUTILS D'ANALYSE DIMENSIONNELLE 129
B.1 Analyse discriminante 129
B.2 Analyse par composante principale 129
B.3 Positionnement multidimensionnel 129
C SÉQUENCE D'ÉVÉNEMENTS OU TEXTURE SONORE : L'INFLUENCE DE LA PÉRIODE D'ATTENTION. 131
C.1 La théorie de la détection du signal 131
C.2 Objectif de l'expérience 131
C.3 Planification expérimentale 131
C.4 Résultats 131

BIBLIOGRAPHY 133

TABLE DES FIGURES

Figure 1 Processus cognitifs et perceptifs 10
Figure 2 Paradigme de la psychologie cognititve 12
Figure 3 Le phénomène de bistabilité : l'illusion du canard-lapin 14
Figure 4 Principaux processus de traitement de l'information auditive et leurs interactions 16
Figure 5 Les trois niveaux d'abstraction de l'axe vertical de la structure catégorielle. 22
Figure 6 Groupement séquentiel : proximité temporelle. 25
Figure 7 Groupement séquentiel : proximité fréquentielle. 26
Figure 8 Groupement simultané : régularité harmonique 27
Figure 9 Groupement ancien-plus-nouveau 28
Figure 10 Compétition entre groupement séquentiel et groupement simultané 30
Figure 11 Tâche de description et tâche de tri ou de catégorisation 38
Figure 12 Catégorisation des paysages sonores urbains, d'après (Raimbault and Dubois, 2005) 48
Figure 13 Taxonomie des sources sonores urbaines, d'après (Brown et al., 2011) 49
Figure 14 Information potentielle contenue dans les séquences d'événements, les textures, et le bruit 53
Figure 15 Plannification expérientale de l'expérience de discrimination de textures sonores et d'exemplaires de textures sonores 55
Figure 16 Événement ou texture sonore : influence de la période d'attention 56
Figure 17 Taxonomie des sources sonores urbaines, d'après (Salamon et al., 2014) 57
Figure 18 TODO 62
Figure 19 Organisation hiérarchique de la banque de sons isolés utilisée pour la simulation 67
Figure 20 TODO 68
Figure 21 Etape de processus de simulation pour l'analyse sensorielle 72
Figure 22 TODO 74
Figure 23 L'interface de sélection aveugle de l'outil de simulation *Simscene* 75
Figure 24 L'outil de simulation *Simscene* 76
Figure 25 Planification expérimentale des éxperiences de simulation et d'évaluation de l'agrément 81

Figure 26 Taxonomies des classes de sons utilisées pour la simulation des environnements sonores urbains 83
Figure 27 TODO 93
Figure 28 TODO 94
Figure 29 TODO 95
Figure 30 TODO 96
Figure 31 TODO 99
Figure 32 TODO 102
Figure 33 TODO 105
Figure 34 TODO 109

LISTE DES TABLEAUX

Table 1 Les catégories sonores les plus citées, d'après (Niessen et al., 2010) 45
Table 2 Résumé des étapes de l'expérience de simulation 85
Table 3 TODO 89
Table 4 Coefficients de corrélation linéaire calculés entre l'agrément perçu moyen A *vs.* TODO 97
Table 5 Classes d'événements identifiées comme étant des marqueurs sonores 100
Table 6 Coefficients de corrélation linéaire calculés entre l'agrément perçu moyen A *vs.* TODO. 103

LISTINGS

ACRONYMS

Première partie

PRÉAMBULE

preamble text here.

1 INTRODUCTION

2 MOTIVATION

Deuxième partie

ANALYSE SENSORIELLE

preamble text here.

3 LA PERCEPTION DE L'ENVIRONNEMENT SONORE

Avant d'aller plus loin dans l'exposé de nos travaux, il nous semble indispensable de dresser ici un état des lieux des connaissances liées à la perception des sons. Nous procédons en cinq étapes.

"partie" -> "dans la 1ere"
"étape" -> "au cours de la 1ere"

Dans la première, nous proposons un tableau général des différents processus intervenant dans le traitement de l'information sonore, processus mis en œuvre dès lors que le signal atteint le tympan.

Dans la seconde, nous interrogeons la manière dont nous nous représentons le monde sonore perçu. Nous démontrons comment cette représentation influe sur notre perception.

Dans la troisième, nous nous intéressons aux processus dits d'Analyse de Scènes Acoustiques (ASA), processus par lesquels le cerveau ségrègue les informations contenues dans l'environnement sonore afin d'en dégager des objets cohérents.

Dans la quatrième, nous introduisons la notion de paysage sonore, et examinons l'impact que cette notion a sur les recherches en matière de perception des environnements. Nos travaux s'inscrivant largement dans cette approche, nous dressons un état de l'art des connaissances en matière de perception des paysages sonores, et tentons de dégager les grands axes méthodologiques suivis par ces études.

Enfin, dans la cinquième, nous nous attachons à définir les concepts d'événement et de texture sonore, notions clefs qui interviennent par la suite dans le modèle de scène sonore proposé.

3.1 LE TRAITEMENT DE L'INFORMATION AUDITIVE

3.1.1 *Perception et cognition*

Perception. Le mot désigne l'ensemble des processus de traitement de l'information sensorielle. La perception du monde sonore qui nous entoure est un phénomène complexe, aujourd'hui encore mal connu.

Cette perception est à l'origine de l'interaction que nous créons avec notre environnement. Elle détermine notre capacité d'adaptation à ce dernier. Cette relation au monde *réel* ne se rompt jamais. Nous percevons des sons en permanence, et ce, même si aucune source sonore n'est présente.

Ainsi tel mélomane dont la tête résonne encore de l'air, bien après que les instruments se soient tus. Ainsi tel usager des transports dont l'oreille anticipe (pour s'en protéger...) les crissements du métro alors que la rame n'est pas encore à quai.

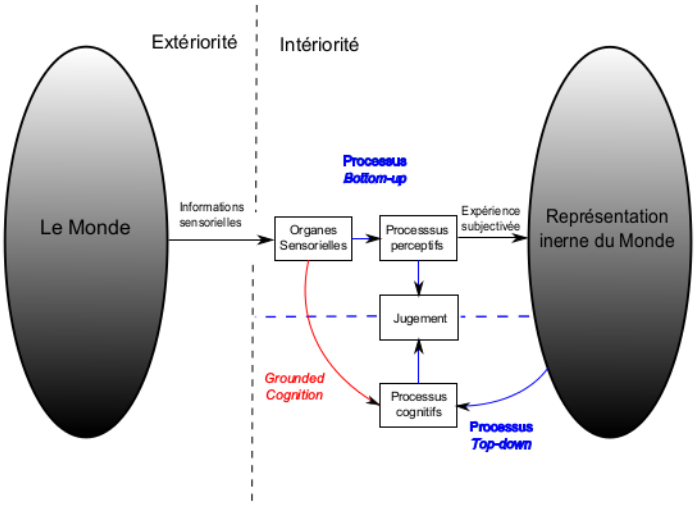


FIGURE 1 : Processus cognitifs et perceptifs

Cognition. Selon U. Neisser[1] dans (Neisser, 1976, p. ??):

> "Cognition is the activity of knowing : the acquisition, organisation and use of knowledge."

Le mot désigne l'ensemble des processus d'acquisition et de développement d'une connaissance du monde.

Selon la théorie classique, perception et cognition dépendent de deux groupes de systèmes fonctionnels du cerveau distincts. La perception mobilise les systèmes de traitement dits modaux, c'est à dire supportés par les organes sensoriels (oreilles, yeux etc ...), tandis que la cognition s'appuie sur des représentations mentales des réalités externes, par essence amodales.

Cette dichotomie entre perception et cognition est aujourd'hui remise en question. Dans une approche "incarnée" de la cognition (*Grounded Cognition*), Barsalou nie le caractère amodal des représentations mentales prônant que ces dernières dépendent également des modalités sensorielles (Barsalou, 2010). Il tente ainsi de réunir les processus perceptifs et cognitifs (Barsalou, 1999; Goldstone and Barsalou, 1998). Les deux approches sont illustrées sur la Figure 1.

3.1.1.1 *Psychologie cognitive et psychoacoustique*

La psychologie cognitive est un domaine de recherche dédié aux phénomènes se rapportant à la connaissance. Elle est née dans les années 50, en réaction au *Béhaviorisme*, théorie fondée sur "l'étude des comportements objectivement observables de l'être humain", et négligeant, de fait, le rôle de la conscience. La psychologie cognitive

1 Ulric Neisser est considéré comme un des pères du cognitivisme notamment grâce à son livre (Neisser, 1967). Il a par la suite critiqué la direction prise par le mouvement, dénonçant une "approche laboratoire" trop éloignée de la réalité terrain.

s'interroge sur des modèles théoriques complexes rendant compte de tous les faits et de toutes les lois connus. Les chercheurs y explorent tout à la fois, la mémoire, le langage, l'intelligence, la perception.

L'approche cognitiviste, dans le domaine de la perception auditive, se distingue de celle plus traditionnelle de la psychoacoustique [2]. Tandis que la psychoacoustique émet l'hypothèse d'une relation directe entre le stimulus et la réponse du sujet, la psychologie cognitive soutient que la réponse, elle, est entièrement corrélée au contexte, à l'expérience, et aux interactions multi-sensorielles (Maffiolo, 1997).

La réponse tient compte non seulement des traitements perceptifs mais aussi des représentations issues et de la mémoire individuelle (i. e. construites en particulier à partir de la relation sensible au monde) et de la mémoire collective, à travers le développement des connaissances partagées (Maffiolo, 1999, p. ? ?).

Latin ("id est") -> italique (comme "in situ")

La psychologie cognitive s'intéresse prioritairement à l'aspect cognitif de la perception en considérant l'individu comme un tout. Elle prend en compte la culture, l'expérience, l'activité de l'individu et ne focalise pas seulement sur la réaction des organes sensoriels comme l'oreille. Elle questionne les aspects qualitatifs plus que quantitatifs de notre compréhension du monde sonore (Maffiolo, 1999, p. ? ?).

Elle envisage l'ensemble des étapes du traitement auditif de manière globale et permet ainsi de faire le lien entre une information sensorielle et une information abstraite (McAdams and Bigand, 1994).

En psychologie cognitive, on distingue les approches cognitivistes, qui s'intéressent plus particulièrement aux processus montant (*botttom-up*, cf. section 3.1.2), relatifs au traitement de l'information perçue, et les approches cognitives, qui interrogent, avant tout, les processus descendants (*top-down*, cf. section 3.1.2) liés à la mémoire du sujet ainsi qu'au contexte (Guastavino, 2003, p. ? ?).

3.1.1.2 *Paradigme de la psychologie cognitive*

Nous l'avons vu, la psychologie cognitive ne conçoit pas le sujet comme une " boîte noire ", mais reconnaît en lui un système de traitement de l'information. Elle admet que l'individu adopte une stratégie afin d'optimiser le traitement des stimuli. Cette stratégie est déterminée par la nature du stimulus, mais aussi par son contexte, et par les connaissances a priori du sujet.

Maffiolo (Maffiolo, 1999, p. ? ?) propose une présentation des présupposés sur lesquels repose la psychologie cognitive (Cf. Figure **??**). Ces présupposés sont résumés ci-après :

- le monde est discrétisé en dimensions ou propriétés issues de la physique, et considérées comme vraies

Tu ne ponctues jamais tes énumérations, ça fait un peu négligé.

Normalement "," ou ";" et "." sur la dernière.

2 La psychoacoustique est une branche de la psychophysique qui applique au domaine de l'acoustique les concepts et les méthodes de la psychophysique

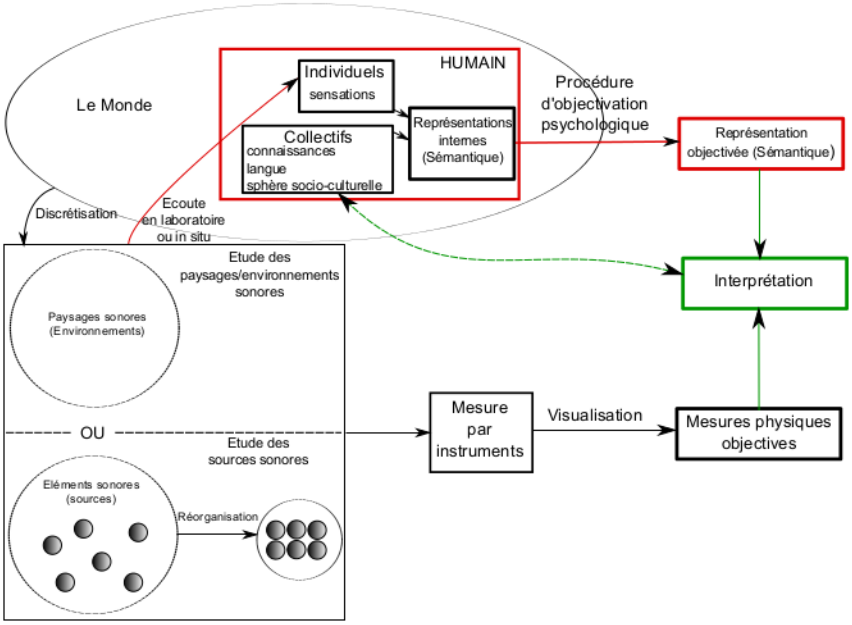


FIGURE 2 : Paradigme de la psychologie cognititve, d'après (Maffiolo, 1999)

- ces dimensions ou propriétés peuvent être mesurées objectivement par des instruments. Elles rendent compte ainsi de la réalité
- le sujet intègre de manière séquentielle ces dimensions ou propriétés en fonction du contexte
- l'évaluation subjective du sujet est interprétée comme un décalage par rapport à la mesure objective considérée comme vraie

Au regard du paradigme classique de la psychologie cognitive, Maffiolo met en évidence quatre points discutables :

- la pertinence des dimensions et propriétés physiques utilisées pour le découpage du monde
- un traitement par les sujets tenant spécifiquement compte de ces dimensions
- une séparation nette entre stimulus et contexte
- le caractère subjectif du jugement humain en comparaison à l'objectivité d'un appareil de mesure.

GL : TODO : commenter

3.1.2 *Processus* Bottum-up *et processus* Top-down

L'interaction entre l'homme et son environnement est fonction d'une part de l'information sensorielle captée par le sujet, d'autre part de la rétroaction exercée par lui sur ces données. Cette rétroaction est déterminée par son expérience sensible du monde. Par "expérience sensible", nous entendons la mémoire interne des interactions passées, mémoire grâce à laquelle nous optimisons l'analyse des stimuli, et intégrons les effets de contexte dus à l'environnement.

Cette mémoire est à la fois :

- individuelle : dépendant de notre expérience propre
- collective : dépendant des connaissances que nous avons acquises sur le monde

La rétroaction est l'expression de l'individualité du sujet, individualité qui explique que deux personnes, ayant des capacités sensorielles semblables, peuvent percevoir différemment un même environnement.

Ainsi la perception mobilise deux formes de traitements (Cf. Figure 1) :

- les traitements dits ascendants (bottom-up) dirigés par les données
- les traitements dits descendants (top-down) dirigés par les concepts ou les représentations

Étudier la perception demande de prendre en compte aussi bien l'information externe (processus ascendant) que l'information interne (processus descendant). Réduire la perception à une simple association de sensations ne permet pas de rendre compte de l'éventail des processus cognitifs entrant dans le décodage de l'environnement. Un exemple concret, emprunté au domaine de la vision, est celui du phénomène dit de bi-stabilité, i. e. La faculté, chez un sujet, de tirer d'un même stimulus deux analyses différentes, mais jamais simultanément (Cf. Figure 3).

GL : TODO : ajouter (Schwartz et al., 2012)

Un autre exemple, emprunté cette fois au domaine de l'audition, nous semble illustrer encore le caractère dual de la perception. Il est donné par McAdams et Bigand (McAdams and Bigand, 1994, p. 2):

> " ...Imaginez vous un instant en pleine forêt amazonienne : vous entendriez exactement les mêmes bruits que le guide qui vous accompagne, mais, étant donné votre manque

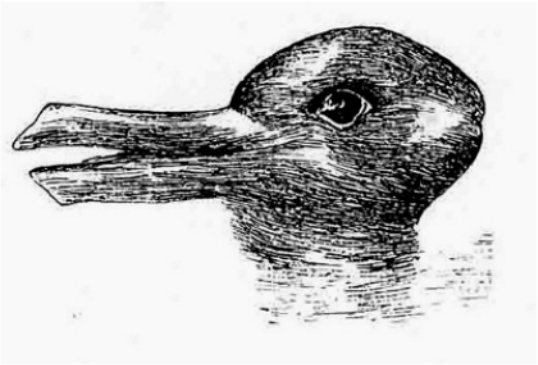


FIGURE 3 : Le phénomène de bistabilité : l'illusion du canard-lapin. Première publication dans *Fliegende Blätter*, 23 octobre 1892, p. 147

> de connaissance du milieu, vous seriez incapable d'extraire du fond sonore les sons correspondant aux cris de l'iguane, aux singes macaques, aux chants des ouistitis ou aux bruissements des arbres tropicaux. De ce fait vous seriez dans l'incapacité d'attribuer une signification à l'ensemble de la structure sonore, ce qui pourrait être important pour votre survie dans l'environnement. "

3.1.3 *L'écoute*

Go : (Gaver, 1993b)

3.1.4 *L'approche Écologique*

L'approche écologique a d'abord été introduite dans le domaine de la vision par Gibson (Gibson, 1966), qui se demande entre autre si les " lois structurant les objets sont porteuses d'informations, ou si cette information est tirée de comparaisons " (Gibson, 1978).

Cette approche reconnaît que la réponse au stimulus dépend à la fois de l'information perçue, et de la connaissance du monde, autrement dit, l'environnement quotidien, et le contexte habituel d'écoute du stimulus.

3.1.4.1 Soundwalk

Appliquée à la perception sonore, l'approche écologique requiert de prendre en compte l'environnement et du sujet, et du stimulus auquel il est exposé. La démarche s'oppose aux méthodes expérimentales traditionnelles, celle de la psychophysique en particulier, sur l'aspect décontextualisant de l'écoute en laboratoire, qui affecte la perception du sujet contraint à un effort d'abstraction supplémentaire pour obtenir l'illusion de la réalité.

Nombre d'études désormais sont réalisées dans un cadre *in situ*. On parle d'ailleurs de *soundwalk* [3] pour désigner les expériences où le sujet est immergé dans l'environnement qu'il doit évaluer (Adams et al., 2008; Jeon et al., 2013).

La méthode des *soundwalk* permet entre autre :

- de contextualiser le sujet, à savoir, l'évaluer dans un environnement qu'il connaît (lieu de vie, de travail)
- d'évaluer l'environnement sonore, tout en maintenant actif les autres sens (vue, olfaction)
- d'éluder le problème de la reproduction des environnements sonores en laboratoire

Malgré tout, les études *in situ*, bien que valides écologiquement, présentent elles aussi des inconvénients. Dans l'hypothèse où tous les sujets ne passent pas l'expérience en même temps, il est impossible de garantir à chacun les mêmes stimuli et/ou le même environnement. Se pose le problème de la reproductibilité des expériences. Inversement, dans l'hypothèse où tous les sujets passent l'expérience en même temps, il peut se poser des problèmes d'organisation, de nature à compromettre une égale réceptivité, disponibilité chez tous les sujets.

3.1.4.2 *Reproduction écologique des environnements sonores en laboratoire*

Le problème de la reproduction écologique des environnements sonores en laboratoire a été particulièrement étudié par Guastavino. (Guastavino and Cheminée, 2003; Guastavino and Katz, 2004; Guastavino et al., 2005). En comparant les descriptions verbales produites à la suite d'écoutes *in situ* , et d'écoutes en laboratoire, via des systèmes de reproduction stéréophoniques et multi-phoniques, (Guastavino et al., 2005) montre que les événements sonores peuplant les scènes sont décrits de la même manière, quel que soit le contexte d'écoute.

Cependant des différences apparaissent au niveau de la description des fonds sonores (*backgrounds*), entre les écoutes stéréophoniques, et les écoutes multi-phoniques et *in situ*, suggérant de fait que le système de reproduction influe sur les processus cognitifs mis en œuvre. Des conclusions similaires sont faites dans (Guastavino and Katz, 2004) s'agissant cette fois de systèmes de reproduction mono, stéréo, et multi-phoniques.

GL : TODO : sélection des stimuli

3 *Soundwalk* est un terme anglais introduit par R. Murray Schafer (Schafer, 1969) signifiant littéralement " marche sonore ". Ce terme étant couramment utilisé en français, il ne sera pas traduit dans ce document

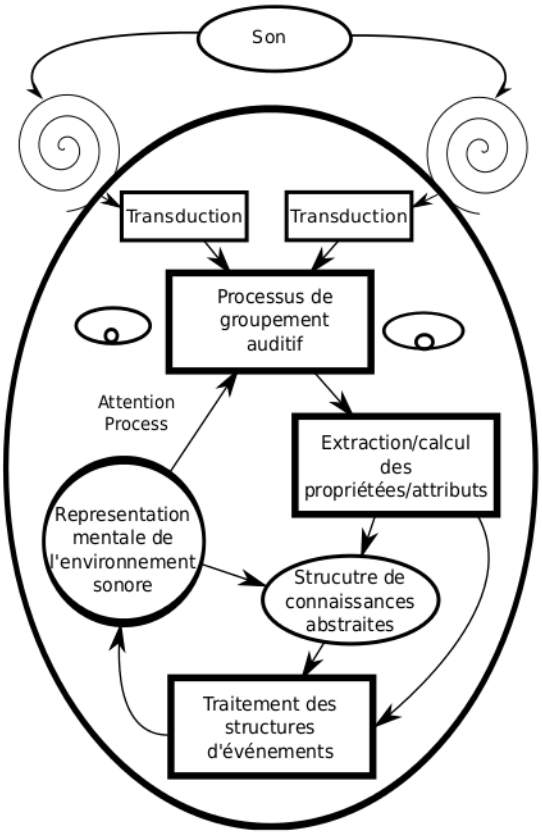


FIGURE 4 : Principaux processus de traitement de l'information auditive et leurs interactions, d'après (McAdams and Bigand, 1994)

3.1.5 *La chaîne traitement de l'information auditive*

Le son est une vibration émise par une source d'excitation, et transmise à l'air. Cette vibration se propage ensuite jusqu'à atteindre un récepteur, le tympan, qui va capter le différentiel de pression résultant de cette vibration. C'est le point de départ du processus de traitement de l'information auditive.

"interconnecté" (sans trait d'union) est dans le dictionnaire

Si on adopte une approche *traitement de l'information*, on peut décomposer ce processus en plusieurs systèmes inter-connectés. Ces systèmes forment une chaîne qui, au fur et à mesure des traitements, interprète le signal acoustique afin d'en extraire l'information sémantique. Plus on se place loin dans la chaîne de traitement, plus on a accès à une information abstraite, potentiellement utilisable par d'autres processus de haut niveau. La figure 4 extraite de (McAdams and Bigand, 1994) nous donne un aperçu des principales fonctionnalités du système de traitement auditif.

Lors de l'étape de *transduction*, les vibrations sonores parvenant au tympa, sont analysées puis traduites en impulsions nerveuses transmises au cerveau. Ces impulsions rendent compte des attributs spectraux et temporels de l'onde. L'extraction des composantes fréquentielles intervient dans la cochlée. C'est à l'intérieur de cette dernière

n

que les différentes parties de la membrane basilaire vont être excitées, en fonction des fréquences composant le signal, suivant un axe tonotopique. Les vibrations, captées à chaque point d'excitation de la membrane basilaire, sont transmises au cerveau via les nerfs auditifs, chaque point codant une information correspondant à une bande fréquentielle limitée.

Vient ensuite le *processus de groupement auditif*. C'est une étape d'intégration temporelle au cours de laquelle l'information est analysée en images auditives cohérentes. Contrairement à ce que pensaient les Grecs, nous ne possédons pas de "canaux" séparés pour chaque objet sonore présent dans l'environnement (Yost, 1994). C'est notre cerveau qui se charge de fusionner et de discrétiser les éléments sonores simultanés afin de créer un flux auditif structuré. En d'autres termes, il détermine le nombre d'objets présents, identifie leur provenance, et en définit le sens.

anciens :-P

Afin d'illustrer notre propos, mettons nous dans la peau du mélomane écoutant cette fois un Choral de Bach. C'est le *processus de groupement auditif* qui, sur la base des paramètres spectro-temporels du signal, nous permet de distinguer les voix de basse, ténor, alto et soprano. Par contre c'est à partir d'une analyse des attributs perceptifs que nous sommes capables de percevoir les mélodies comme des objets unitaires, même si ces dernières sont développées entre les différentes voix du choral. Cette extraction des propriétés perceptives intervient pendant la phase dite *d'extraction/calcul des propriétés/attributs*.

Les travaux menés au titre de l'Analyse de Scènes Auditives (ASA), dont le sujet est abordé plus bas (Cf. section 3.3), ont largement porté sur ces processus de groupement.

Italique, pas de majuscule (mais normalement, maj. à "Section")

GL : TODO : finir

3.2 REPRÉSENTATION MENTALE DE L'ENVIRONNEMENT SONORE

Le cerveau entretien un dialogue constant avec l'environnement sonore. Ce dialogue s'effectue par le biais des représentations mentales. Une définition des représentations mentales est donnée par (Houdé et al., 1998):

> " La représentation mentale peut être vue comme une entité interne, le correspondant cognitif individuel des réalités externes expérimentées par un sujet. "

Ces représentations font office de sauvegardes de l'information. Conservées en mémoire sous une forme hautement abstraite (McAdams and Bigand, 1994, p. ? ?), elles rendent compte à la fois de notre compréhension du monde, et de la manière dont nous l'abordons. Ces connaissances subjectives, non directement observables, restent

néanmoins accessibles au chercheur par le biais d'expériences d'objectivation (voir section 3.4.3.1)

Ces représentations forment une image mentale discrète d'un monde réel continu (Houdé et al., 1998). C'est par le passage du continu au discret que nous sommes à même d'organiser nos connaissances, afin de les réutiliser de manière efficace. Les objets discrets issus de cette organisation sont appelés des catégories. L'action consistant à juger si un événement perçu appartient à une catégorie, est appelée la catégorisation.

3.2.1 *La notion de catégorie*

3.2.1.1 *Définition*

Une des vérités de tout être humain est de segmenter son environnement, i. e. de se bâtir un système de classification permettant de regrouper des objets n'étant pas identiques (Rosch and Lloyd, 1978, p. 1). On appelle catégorisation, l'action consistant à regrouper des objets du monde physique considérés comme équivalents, et catégorie, l'entité mentale contenant le groupe d'objets ainsi rassemblés.

D'un point de vue écologique, la catégorisation est un processus essentiel. Nous sommes constamment en train de catégoriser l'environnement, et devons être à même, à tout moment, de prendre une décision sur l'appartenance catégorielle d'un objet. Ce processus est adaptatif. La prise de décision est toujours fonction du sujet, d'une situation et d'un contexte. Ainsi, un même objet, perçu à deux moments distincts, pourra être affecté à des catégories différentes.

3.2.1.2 *La nature des catégories*

3.2.1.3 *Utilité des catégories*

Avant de parler de l'utilité des catégories, voyons où et comment s'exercent les catégorisations dans notre quotidien. (Anderson, 1991) identifie trois occurrences :

- le langage : Le langage est le lieu, par excellence, des catégories. Catégoriser, c'est considérer un objet comme un élément distinct du monde. Cette distinction s'accompagne généralement (pas toujours) d'une désignation. C'est l'essence même des processus d'identification que de chercher à nommer les objets, une fois qu'ils ont été isolés. Il est raisonnable de penser que, de la même manière, une catégorie possède un label associé. On parlera alors de catégorie sémantique. Cette relation entre langage et catégorie nourrit le débat sur l'universalité présupposée de la catégorisation. L'opération permanente consistant à

isoler un objet, et à lui attribuer un nom, est l'antinomie du "geste adamique", i. e. d'attribution libre du nom, hors toute influence et/ou contexte (d'Adam, le premier homme révélé et nommé dans la bible). La langue est un code partagé par une communauté. Dans une certaine mesure, sa définition, i. e. le sens donné aux mots, peut varier suivant les groupes de cette communauté. Ainsi, catégoriser ne dépend pas seulement d'une réalité physique du monde, mais également d'un contexte socio-culturel. La catégorisation peut être vue comme une action intermédiaire entre, d'une part, l'organisation d'une connaissance individuelle résultant d'une expérience sensorielle personnelle, et, d'autre part, la constitution d'une représentation collective pouvant être partagée par le biais d'un langage commun (Dubois et al., 2006).

Faudra faire une passe à la fin pour vérifier tous les guillemets...

- le regroupement par similarité des caractéristiques : Nous sommes capables de regrouper des objets possédant des caractéristiques similaires, et ce, même si ces objets nous sont inconnus (Fried and Holyoak, 1984).

- le regroupement par similarité fonctionnelle : Nous sommes capables d'interpréter des objets possédant des fonctions similaires comme faisant partie d'un même groupe, et ce même si ces objets ont des caractéristiques distinctes. Pour exemple : deux hommes descendant d'un camion de pompier pour éteindre un feu de forêt seront catégorisés pompiers, ce indépendamment des tenues qu'ils portent. Quand nous parlons de la catégorisation comme du processus de discrétisation du monde réel, ce "monde réel" englobe et la réalité des faits physiques, et la réalité des faits sociologiques.

3.2.1.4 *L'effet de la catégorisation*

3.2.2 *Le processus de catégorisation*

La catégorisation est nécessairement un outil supervisé, i. e. nous ne pouvons catégoriser les objets qu'à partir des catégories que nous connaissons. Le processus catégoriel procède suivant deux mécanismes :

- mécanisme inductif : associer un objet à une catégorie sur la base des caractéristiques perçues de ce dernier

- mécanisme déductif : associer à un objet les caractéristiques de la catégorie à laquelle il appartient

3.2.2.1 *Capacité d'inférence*

3.2.3 *Organisation de la structure catégorielle*

Le cerveau doit en permanence faire sens d'une information riche et variée, et ce de manière productive. Afin de satisfaire à cette exigence d'efficacité, l'organisation catégorielle doit répondre à deux grands principes (Rosch and Lloyd, 1978, p. 29):

- L'économie cognitive
- La redondance structurelle

3.2.3.1 *L'économie cognitive*

La catégorisation doit fournir un maximum d'informations pour un minimum d'efforts. C'est pourquoi la logique catégorielle prend en compte le contexte sensoriel. En résumé, le traitement de l'objet s'opère et par rapport à lui, et par rapport au traitement des objets perçus et catégorisés simultanément. Comme énoncé par D. Dubois (Dubois, 1991, p. 33):

> "Catégoriser un stimulus signifie le considérer dans la finalité de cette catégorisation, non seulement comme équivalent des autres stimuli de la même catégorie, mais également différent des stimuli qui n'appartiennent pas à cette catégorie".

Du principe d'économie cognitive, il découle que la catégorisation de l'objet n'est pas une catégorisation dans l'absolu. Elle ne dépend pas uniquement de l'observation des propriétés particulières de l'objet, mais également du contexte dans lequel il est appréhendé.

3.2.3.2 *La redondance structurelle*

L'ensemble des objets physiques ne vit pas dans un espace fini, identifié, et dont les valeurs seraient équiprobables. Le monde ne se résout pas à des paramètres dimensionnés, indépendants et manipulables, comme dans le cadre d'études en laboratoire. Au contraire. Il peut exister des discontinuités saillantes entres objets, de même que ces objets peuvent être liés entre eux par des patterns de co-occurrence de propriétés (exemple : un chien possède "quatre pattes et un museau" plus souvent que "deux pattes et un museau"). Ces discontinuités et corrélations, présentes dans les propriétés perçues, étayent la structure catégorielle de notre représentation mentale, et gouvernent ainsi le processus de catégorisation.

3.2.3.3 *Catégorie et abstraction*

Rosch propose de voir la structure catégorielle suivant deux axes (Rosch and Lloyd, 1978, p. 30-41):

- *axe vertical* : L'axe vertical fixe l'organisation hiérarchique des catégories, et permet d'appréhender l'imbrication de ces catégories les unes par rapport aux autres. Ce faisant, il en dresse la taxonomie, les catégories de haut-niveaux représentant des objets abstraits ou concepts, et incluant un grand nombre de sous catégories, et les catégories de bas-niveaux représentant des objets concrets, incluant peu de sous catégories. Ainsi, plus le niveau d'abstraction est grand, plus les similitudes entre objets d'une même catégorie (intra-catégorielle), ainsi que les similitudes entre objets de catégories distinctes (inter-catégorielle) sont faibles. Inversement, plus le niveau d'abstraction est faible, plus les similitudes intra- et inter-catégorielles sont élevées. Rosch décompose cette dimension verticale en trois niveaux d'abstraction (Cf. Figure 5) : Superordonné, Base et subordonné. Le niveau superordonné regroupe les catégories à haut niveau d'abstraction. Les périmètres de ces catégories sont larges (Mobilier, Véhicule...) i. e. les objets qu'elles contiennent peuvent être très distincts. Le niveau subordonné regroupe les catégories à bas niveau d'abstraction, ou catégories concrètes. Les périmètres de ces catégories sont plus précis (Chaise longue, Cabriolet...), et les objets qu'elles contiennent sont nécessairement très similaires. On notera ici que les objets de la classe Cabriolet présentent plus de propriétés communes avec les objets de la classe Berline (c'est un exemple...), que n'en sauraient partager les objets des classes Mobilier et Véhicule.

- *axe horizontal* : L'axe horizontal fixe, lui, l'organisation " géographique " des catégories, et permet d'appréhender, d'une part, les périmètres de ces catégories au sein d'un même niveau d'abstraction, d'autre part, la typicité des objets contenus dans une même classe. Bien sûr, les catégories ne sont pas des objets strictement discrets, et les propriétés des objets qu'elles regroupent peuvent se trouver attribuées à d'autres objets, dans d'autres catégories. Ainsi, les frontières entre les différentes catégories ne sont pas figées, et peuvent même se recouvrir.

3.2.4 *Théories de la catégorisation*

3.2.4.1 *Théorie classique*

Go : (Goldstone and Kersten, 2003)

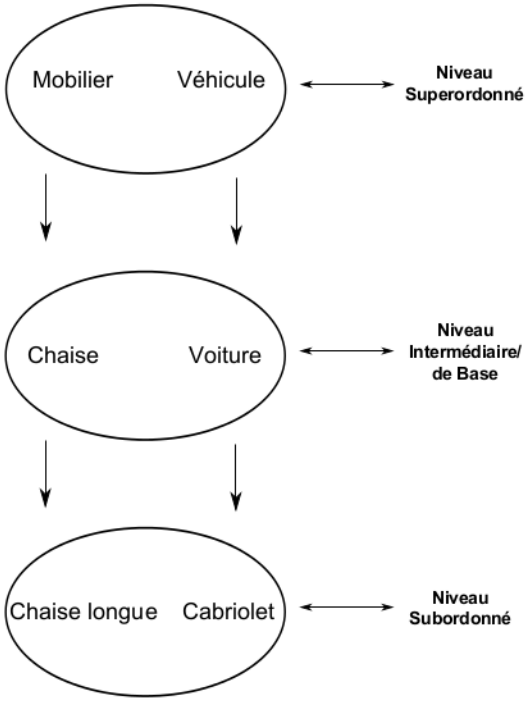


FIGURE 5 : Les trois niveaux d'abstraction de l'axe vertical de la structure catégorielle.

3.2.4.2 *Théorie prototypique*

Nous traitons ici du processus de catégorisation formalisé par E. Rosch et B. B. Lloyd (Rosch and Lloyd, 1978), et nommé théorie prototypique de la catégorisation.

Pour discriminer les catégories, Rosch propose de ne pas raisonner en terme de frontières, mais plutôt de décrire chaque catégorie par un nombre de cas non ambigus (clear case) (Rosch and Lloyd, 1978, p. 36). Tous les objets d'une catégorie ne sont pas également représentatifs de cette dernière. Les cas non ambigus peuvent être vus comme les objets les plus typiques de la catégorie. Ce fait est empirique. Il a été montré que des sujets peuvent très bien s'accorder sur la typicité d'un objet par rapport à une catégorie, tout en n'étant pas d'accord sur les frontières de cette dernière (Rosch, 1975; Rosch and Lloyd, 1974). Le terme prototype, qui donne son nom à la théorie, vient de l'assertion que, parmi ces cas non ambigus, il en existe un, le prototype, plus représentatif que les autres, et qui forme le noyau de la catégorie.

italique

Les catégories sont structurées en interne, en référence à un prototype, i. e. l'objet possédant les attributs typiques de celle-ci. L'appartenance d'un objet à une catégorie dépend alors de la ressemblance qu'entretient ce dernier avec le prototype. Plusieurs propositions ont été faites afin de définir le prototype d'une catégorie : Pour Tversky (Tversky, 1977), l'élément prototype est celui dont la somme des si-

milarités avec les autres éléments de la catégorie est la plus élevée. Pour (Rosch and Mervis, 1975), il s'agit de l'objet possédant le plus de propriétés en commun avec les objets de la catégorie, et le moins de propriétés avec les objets des catégories externes. La typicité d'un élément d'une catégorie s'évalue à la fois en fonction de son degré d'appartenance à celle-ci, et de son degré de différenciation vis à vis des autres catégories.

En se limitant à l'observation d'attributs vivant dans un espace métrique, (Reed, 1972; Rosch et al., 1976) ont montré que le prototype est un centroid, un objet défini comme étant la moyenne des attributs des objets de la catégories.

Mais il faut bien comprendre que cette théorie prototypique de la catégorisation, bien que se basant sur des faits expérimentaux, est avant tout une vision pratique, un concept qui n'a pas été clairement défini et dont l'implication dans les processus de catégorisation reste floue (Rosch and Lloyd, 1978, p. 36-40) (Dubois, 1991, p. 49-54). familier

3.2.4.3 *Théorie de l'exemplaire*

La théories des exemplaires ou *context theory* (Medin and Schaffer, 1978), et son extension multidimensionnelle (Medin and Schaffer, 1978)

3.2.4.4 *Autres théories*

3.2.5 *Similarité et catégorisation*

Go : (Gygi et al., 2007)

3.2.6 *Catégorisation et contexte sensoriel*

G1 : (Ballas and Howard, 1987; Gygi and Shafiro, 2011; Niessen et al., 2008)

3.3 ANALYSE DE SCÈNES ACOUSTIQUES

3.3.1 *Définition*

L'analyse de scènes est d'abord apparue dans le domaine des recherches informatiques en vision. Le terme fait référence aux différentes stratégies par lesquelles un ordinateur parvient à isoler un objet d'une image (McAdams and Bigand, 1994, p. 12). Considérant les processus perceptifs, le pendant auditif de l'analyse de scènes est l'Analyse de Scène(s) Acoustique(s). L'ASA a été introduite par A. S. Bregman dans son ouvrage de référence(Bregman, 1994). ou des objets, ou une structure

L'ASA désigne l'ensemble des traitements perceptifs permettant d'isoler, dans une mixture sonore, les informations émanant de sources

distinctes, et de les "organiser" en un tout cohérent. Ces regroupements sont nécessaires au cerveau, et donc au sujet, pour comprendre, pour donner sens à l'environnement. On parle de processus de ségrégation ou de processus de groupement (Winkler et al., 2009). Comme vu précédemment section 3.1.5, ces processus mobilisent à la fois les traitements ascendants ou bottom-up, intervenant au niveau de l'information auditive transduite, et les traitements descendants ou top-down, intervenant au niveau du bagage mémoriel. Les processus *bottum-up* sont appelés *processus primitifs*, et les processus *top-down*, *processus basés sur des schémas*.

plutôt mettre les noms français entre guillemets

Les *processus primitifs* sont innés, et opèrent à partir des régularités du signal, afin d'en regrouper les composantes fréquentielles produites par une même source. Le mot régularité désigne ici les propriétés constantes de l'environnement, perçues par tous les hommes, et en tous lieux. Par exemple, si la fréquence fondamentale d'un son harmonique change au cours du temps, toutes ses harmoniques changeront également afin de maintenir la structure harmonique du son [p. 38](Bregman, 1994).

sujets, individus

Les *processus basés sur des schémas* sont eux conditionnés, et opèrent sur la base de connaissances, ou schémas, partie intégrante de notre représentation mentale du monde, représentation construite sur la base des écoutes antérieures.

3.3.2 *Une approche psychoacoustique*

Si la plupart des recherches sur l'ASA adoptent une approche cognitiviste, se concentrant sur l'étude des *processus primitifs*, elles suivent généralement une méthodologie expérimentale très inspirée de la psychoacoustique. De fait, les sujets sont soumis à des stimuli décrits analytiquement dans un espace multidimensionnel de dimensions physiques (fréquence, intensité etc ...) (Dubois et al., 2006). Dans la majorité des cas, ces stimuli ou sons, qu'ils soient purs ou qu'ils soient complexes,[4] sont synthétisés en laboratoire.

Ces sons sont proposés à l'écoute de manière séquentielle. Au cours de l'expérience, les paramètres en sont modifiés (intensité, hauteur fréquentielle, espace entre les séquences...) afin d'évaluer le seuil à partir duquel la capacité du sujet de distinguer les sources sonores est altérée.

à

L'étude de l'ASA se borne donc à l'analyse de l'effet de descripteurs "bas niveau" dans le processus d'intégration, sans tenir compte d'attributs perceptifs "haut niveau", comme la valeur sémantique attribuée aux sons, sans tenir compte non plus de considérations écologiques **??**. C'est pourquoi il est difficile de faire le lien entre la notion de source sonore ayant cours dans le domaine de la psychoacoustique,

4 Un son pur est un son composé d'une seule sinusoïde, i. e. possédant une seule fréquence. Un son complexe est, lui, composé de plusieurs composantes fréquentielles

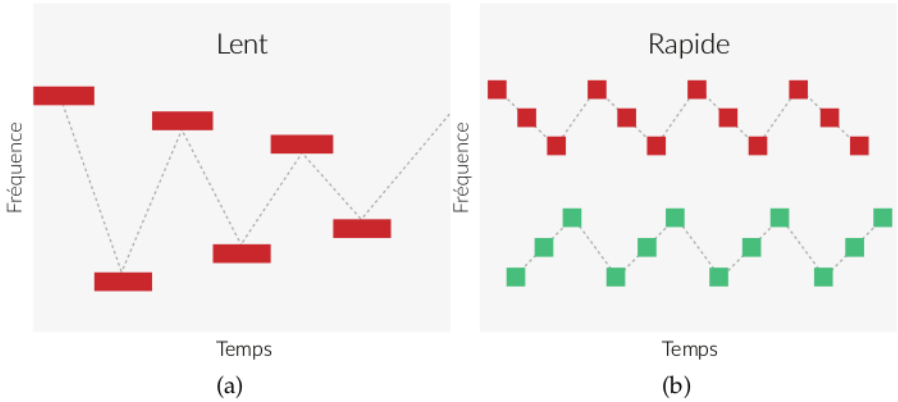


FIGURE 6 : Groupement séquentiel : proximité temporelle. Dans l'exemple (a), la durée entre les événements est telle qu'aucun groupement n'est effectué, les sons sont perçus comme des événements distincts. Dans l'exmple (b) la durée entre les événements est réduite, et un groupement s'opère suivant la proximité fréquentielle. Deux flux sont ainsi créés, le premier (rouge) regroupe les sons haute fréquence, et le deuxième (vert) les sons basse fréquence.

et celle admise dans le domaine de la psychologie cognitive. En théorie, dans l'une et l'autre approche, le terme source sonore se réfère à l'objet (e. g. voiture) supposé en être la source. Mais, en psychologie cognitive, cette relation est naturelle, le stimulus étant le plus souvent un enregistrement de ladite source, tandis qu'en psychoacoustique cette relation est dénaturée, le stimulus, synthétisé, étant un objet abstrait (agglomérat de sinusoïdes) dont l'existence n'est avérée que dans la mesure où il est interprété par le sujet comme un tout, une entité. Les résultats, obtenus à partir de tels stimuli, éloignés de la réalité des phénomènes acoustiques perçus, sont difficiles à généraliser à des applications plus incarnées. Nous touchons là aux limites de l'approche psychoacoustique appliquée aux études des *processus primitifs* de l'ASA. Les résultats, obtenus à partir de stimuli très éloignés de la réalité des phénomènes acoustiques perçus, étant difficiles à généraliser à des applications plus incarnées.

phrase sans verbe

3.3.3 *Régularités et processus primitifs*

L'existence des *processus primitifs* est une conséquence de l'efficience, dans le monde sonore, de régularités universelles affectant l'ensemble des stimuli auditifs. Bregman distingue 4 types de régularités [p. 19,21,31,33](McAdams and Bigand, 1994):

1. *Synchronicité* : Il est rare que des sons n'ayant aucun rapport entre eux démarrent et s'arrêtent au même moment.

2. *Continuité* :

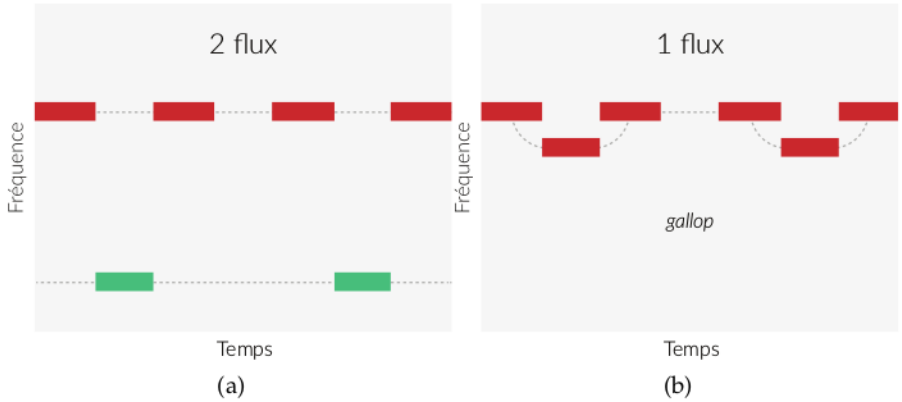


FIGURE 7 : Groupement séquentielle : la proximité fréquentielle. Deux groupes de sons sont joués à deux fréquences. Dans l'exemple (a), la distance entre les fréquences des sons est telle que deux flux sont créés, le premier (rouge) regroupe les sons haute fréquence, et le deuxième (vert) les sons basse fréquence. Dans l'exemple (b) la distance entre les fréquences est réduite et un groupement s'opère suivant la proximité fréquentielle. Un seul flux est créé, et l'on perçoit un pattern temporel rappelant le galop d'un cheval.

motif

- Les propriétés d'un son isolé tendent à se modifier lentement et de façon continue.
- Les propriétés d'une séquence de sons émis par la même source tendent à se modifier lentement.

3. *Harmonicité* : Lorsqu'un corps sonore vibre à une période répétée, ses vibrations donnent naissance à un pattern acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale

4. *Uniformité* : La plupart des modifications qui surviennent dans un signal acoustique affectent tous les composants du son résultant, de manière identique et simultanée.

Notre perception du monde est assujettie à ces lois. Les *processus primitifs*, sensibles aux stimuli exclusivement, nous permettent d'isoler du monde sonore des objets cohérents perçus à travers le prisme desdites régularités(Ballas and Howard, 1987). Le fait est qu'un principe similaire de perception des formes s'applique également au domaine de la vision.

3.3.4 *Perception de la forme*

Que ce soit en vision ou en audition, notre cerveau est en permanence stimulé par une multitude de sources distinctes. Percevoir un objet dans cet agglomérat, c'est être capable d'isoler tous les signaux

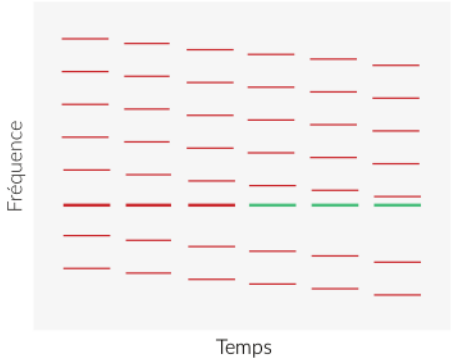


FIGURE 8 : Groupement simultané : régularité harmonique. Un son complexe est joué plusieurs fois. A chaque occurence, on abaisse les hauteurs de la fréquence fondamentale et des harmoniques de manière uniforme. Une harmonique est conservée à fréquence constante (trait gras). Au début, un flux est créé, i. e. les harmoniques et la fréquence fondamentale sont perçus comme étant un seul objet. Au fur et à mesure que la régularité harmonique est brisée, le cerveau tend à percevoir l'harmonique à fréquence constante dans un flux séparé, i. e. comme étant un second objet.

émis par une même source, et de les réunir en une unité perceptive cohérente.

Parmi les premiers travaux qui se sont intéressés à ces processus de groupement, on trouve la psychologie de la forme, en allemand *Gestalt Theory*. Cette théorie, introduite par Ernst Mach et Christian von Ehrenfels à la fin du XIXème siècle, se propose d'expliciter les principes selon lesquels des stimuli sensoriels sont combinés afin de former un pattern mental rendant compte de la présence d'un objet présent dans l'environnement.

Theory ça fait pas très allemand... "Gestalttheorie"

Mis en évidence en perception visuelle, ces principes restent vrais en perception auditive(Bregman, 1994, ch. 1). Parmi ces principes, nous en détaillons ici cinq :

1. *Proximité* : Des éléments proches les uns des autres ont tendance à être groupés ensemble. En audition, ce principe de proximité opère suivant les différentes caractéristiques du son, à savoir, la fréquence, l'*onset* [5] et l'intensité.

2. *Similarité* : Des éléments qui se ressemblent ont tendance à être groupés ensemble.

   Dans le domaine de la vision, la proximité est une notion de spatialité. La similarité, elle, s'applique aux caractéristiques physiques de l'objet (forme, couleur, etc ...), qui ne peuvent se décrire dans une dimension unique.

5 En traitement du signal audio, on désigne par le mot anglais *onset* le début du signal. Ce terme étant couramment utilisé en français, nous ne le traduirons pas dans ce document.

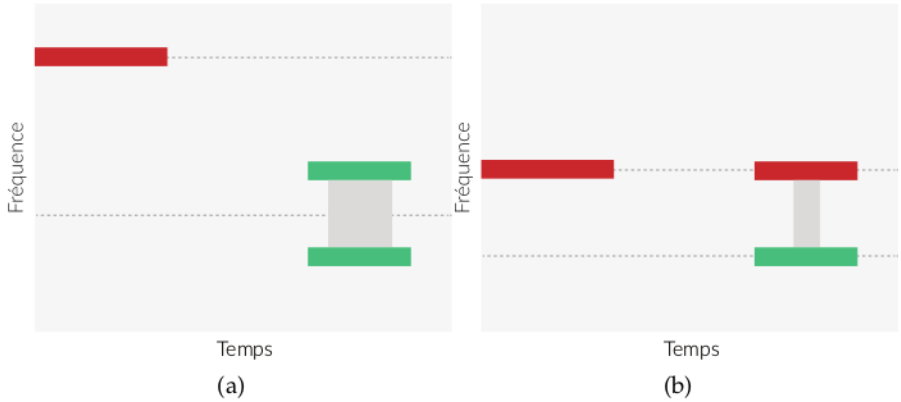


FIGURE 9 : Groupement ancien-plus-nouveau. Dans l'exemple (a), les sons A et B ont des fréquences éloignées. Le cerveau génère deux flux, le premier relatif au son A (rouge), et le deuxième comprenant les sons B et C (vert). Dans l'exemple (b), les sons A et B ont la même fréquence. Le cerveau interprète le son B comme étant une continuité du son A. L'attraction entre B et C en est réduite. Le cerveau génère toujours deux flux, le premier regoupant cette fois ci les sons A et B (rouge), et le deuxième le son C (vert).

Dans le domaine de l'audition, la proximité est généralement admise comme notion de temporalité (*onsets* proches). Pour le reste des descripteurs, il est cependant difficile de distinguer les principes de proximité et de similarité. Bregman propose de parler proximité lorsqu'on traite d'une dimension physique particulière, et de parler similarité dès lors qu'on considère un ensemble de descripteurs, ou lorsque l'on traite d'attributs qui ne peuvent être clairement décomposés suivant des dimensions distinctes, comme le timbre.

de

3. *Continuité* : Des éléments qui varient de manière non abrupte ont tendance à être groupés ensemble. Par ce principe, des objets distincts mais proches temporellement (respectivement spatialement pour la vision) ont tendance à être perçus comme le prolongement des uns par rapport aux autres. C'est ce principe qui nous permet de percevoir comme une seule entité, un objet dont les caractéristiques varient dans le temps,e.g. le son d'une sirène. A contrario, un changement abrupte indique généralement l'apparition d'une nouvelle source.

par les autres

pas de ","

De récentes études en neurosciences ont montré l'importance de ce principe dans les processus de groupement. Notamment (Winkler et al., 2009) qui propose de voir l'ASA comme un processus prédictif, le cerveau cherchant à anticiper la nature des stimuli qui lui parviennent, sur la base de régularités extraites des objets détectés dans l'instant précédant.

phrase sans verbe

4. *Clôture* : Des éléments discontinus qui suggèrent la forme d'un objet continu ont tendance à être groupés ensemble. De manière automatique, le cerveau tend à percevoir un ensemble d'objets distincts comme un tout. En audition, ce principe est très lié à la notion de masquage. En effet, les sons que nous percevons sont régulièrement masqués par d'autres sons concurrents, éventuellement plus forts. Le principe de clôture nous permet de compenser ce phénomène de masquage et de percevoir le signal sans discontinuité. Ainsi lorsqu'un son pur est régulièrement entrecoupé de silences, nous percevons une série de sons purs, mais, si ces silences sont comblés par un bruit blanc, nous percevons un son pur continu. Ce phénomène est parfois appelé " l'illusion de continuité " (Dannenbring, 1976), et s'applique particulièrement dans le contexte de la perception de la parole (Carlyon et al., 2002).

5. *Destin commun* : Des éléments qui varient de manière synchrone et uniforme ont tendance à être groupés ensemble. Comme évoqué à la section 3.3.1 c'est ce principe qui permet de percevoir comme un tout les différentes harmoniques qui composent un son complexe. C'est également ce principe qui nous incite à percevoir de manière unie des stimuli ayant le même *onset* temporel.

Ces principes agissent de concert afin de grouper les composantes du son en flux auditifs.

3.3.5 *Flux auditif et stratégie de groupement*

L'une des notions les plus importantes en ASA est le concept de flux auditif (*auditory stream*), que Bregman définit comme " le groupement perceptif que nous faisons des parties du spectre qui vont ensemble ".

Contrairement au terme " son ", qui peut faire référence aux réalités physiques des phénomènes acoustiques, aussi bien qu'aux représentations mentales que nous nous en faisons, le flux auditif désigne spécifiquement une entité perceptive.

Le terme flux se veut volontairement généraliste. Ce dernier peut désigner aussi bien un son isolé (un coup de marteau), que plusieurs sons à condition que ces derniers soient perçus comme formant une seule entité (une série rapprochée de coups de marteau).

On désigne par " formation de flux auditive " (*auditory streaming*) , le processus à l'origine de la création de flux. (Winkler et al., 2009) propose la définition suivante en ce qui concerne la formation de flux auditifs :

> " Un phénomène perceptif dans lequel une séquence de sons est perçue comme étant composée de deux ou plusieurs flux auditifs "

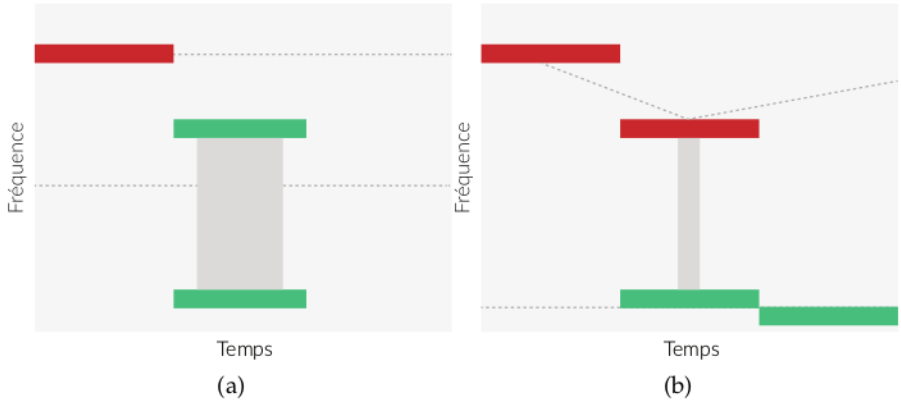


FIGURE 10 : Compétition entre groupement séquentiel et groupement simultané. Dans l'exemple (a), le cerveau perçoit deux flux, le premier regroupant le son A (rouge) et le deuxième regroupant les sons B et C (vert). Dans l'exemple (b), un quatrième son (D) est ajouté, ce dernier possédant une fréquence très proche de celle de (D). Le groupement sequentiel par proximité fréquentielle entre les couples A-B et C-D est favorisé au détriment du groupement simultané entre les sons B-C.

Comme nous venons de le voir, le flux désigne une représentation perceptive d'un son, et, à ce titre, il est l'équivalent auditif de l'objet pour la vision (Bregman, 1994, p. 11). Cependant, nous notons que la notion de flux, et plus particulièrement celle de formation de flux, est souvent utilisée dans un contexte où une dimension temporelle est sollicitée. On parle d'ailleurs de construction de flux (*build-up of streaming*) pour désigner la période durant laquelle le cerveau accumule des indices afin de générer et stabiliser les flux auditifs (Cusack et al., 2004; Snyder and Alain, 2007). Ainsi dans ce document, nous réservons les termes " flux " pour désigner les représentations mentales des stimuli en train d'être traités (intégrés temporellement), et " formation de flux auditifs " pour désigner les processus perceptifs de groupement des sons. Nous conservons le mot " objet " pour désigner de manière générale les représentations mentales, stockées en mémoire, des phénomènes acoustiques.

Considérant les processus primitifs, on distingue trois stratégies de groupement.

- *groupement séquentiel* : désigne le groupement de sons ne partageant pas un même *onset*. Dans le cas de sons purs, le groupement séquentiel s'appuie beaucoup sur le principe de *proximité*, et notamment de *proximité* fréquentielle, principe qui veut que, plus deux sons sont proches en fréquence, plus ils ont tendance à être regroupés dans le même flux (Cf. Figure 7). Pour des sons complexes, c'est plutôt le principe de similarité qui entre en jeux. La *proximité* temporelle, i. e. la durée séparant chaque son, rentre cependant en ligne de compte. Plus cette dernière est

faible, plus deux sons ont des chances d'êtres regroupés (Cf. Figure 6).

- *groupement simultané* : aussi appelé groupement spectral, désigne le groupement des composantes fréquentielles qui partagent un même *onset*. Le groupement simultané s'appuie principalement sur la régularité harmonique. Une composante fréquentielle venant briser la suite harmonique tend à être isolée dans un autre flux auditif (Cf. Figure 8).

- *groupement ancien-plus-nouveau* : Ce groupement est l'application directe du principe de continuité. Lorsque le spectre de l'environnement sonore s'enrichit subitement, tout en conservant ses composantes fréquentielles de départ, le cerveau tend à interpréter l'ajout comme une continuation de l'ancien, et à l'intégrer dans le même flux sonore (Cf. Figure 9).

Dans le cas d'une compétition entre un groupement séquentiel et un groupement simultané, c'est l'organisation issue du groupement séquentiel qui prime (Cf. Figure 10). Ce phénomène fait sens d'un point de vue écologique. En effet la plupart des sons, et en particulier ceux utilisés pour la communication, n'existent que dans une certaine durée, et sont intermittents. Il est alors nécessaire de faire des associations entre des sons parfois séparés par un intervalle de temps long afin de percevoir le sens du message (Winkler et al., 2009).

Les exemples cités plus haut se placent tous dans un contexte *bottom-up*, en évacuant d'éventuels processus attentionnels. Bien que cela ait été encore très peu étudié, il paraît cependant évident que ces stratégies de groupement s'opèrent également dans un contexte *top-down*, en s'appuyant sur une mémoire à plus long terme.

3.3.6 *Attention et perception*

3.3.7 *L'approche par les neurosciences*

review : (Snyder and Alain, 2007)

3.4 L'ÉTUDE DES PAYSAGES SONORES

3.4.1 *La notion de paysage sonore*

La notion de paysage sonore (*soundscape*) a été introduite par Schafer dans les années soixante-dix dans son livre (Schafer, 1969), et détaillée dans l'ouvrage de référence (Schafer, 1977). la question que pose Schafer est :

> " Quelle est la relation entre l'homme et les sons de l'environnement qui est le sien, et que se produit-il lorsque ces sons viennent à changer ? "

Une première définition du paysage sonore a été donnée par (Truax, 1978)

> " [a]n environment of sound (or sonic environment) with emphasis on the way it is perceived and understood by the individual, or by a society "

Aujourd'hui, cependant, on s'accorde sur la définition suivante (Aletta et al., 2016):

> " Un environnement sonore tel qu'il est perçu, expérimenté et/ou compris par un individu ou une communauté, dans son contexte. " [6]

L'une et l'autre définition sont larges. Tout environnement peut être considéré comme un paysage sonore dès lors qu'on lui associe un ensemble de sons entendus par un sujet donné. Le problème est d'envisager l'environnement sonore par rapport à l'évaluation subjective de l'auditeur, et non, uniquement, par rapport à ses paramètres acoustiques. Schafer, déjà, explique la nécessité de ne plus considérer le bruit seul, mais aussi la perception de ce bruit par les individus, et le contexte dans lequel il est perçu, ceci afin d'améliorer la qualité de leur environnement. On parle d'environnement sonore lorsqu'on se réfère au phénomène acoustique physique, et de paysage sonore lorsqu'on se réfère à la représentation que l'on se fait de l'environnement.

Ainsi, les études sur les paysages sonores suivent le paradigme de la psychologie cognitive (Dubois et al., 2006; Maffiolo, 1999) (cf. section 3.1.1.2). L'environnement sonore est décrit en utilisant à la fois des descripteurs acoustiques (mesures), et des descripteurs perceptifs, l'analyse de l'interaction entre ces descripteurs permettant de comprendre les processus cognitifs mis en œuvre dans l'évaluation perceptive des paysages sonores.

L'approche étant ainsi centrée sur le sujet, les recherches sur les paysages sonores sont par essence interdisciplinaires (Aletta et al., 2016; Davies et al., 2013), faisant appel à des outils et des méthodes provenant de champs de recherches variés comme l'acoustique, la psychologie cognitive, la psycho-linguistique, la sociologie, et plus récemment, l'intelligence artificielle.

6 Cette définition a été publiée dans le cadre de la norme *ISO-12913* (Standardization, 2013)

3.4.2 *Application à la nuisance sonore urbaine*

GL : TODO : citer (Tae Hong Park et al., 2014), senseur de l'espace urbain

La ville est un environnement bruyant. Elle l'a été de tous temps. GL : TODO : citation romain Ce qui a changé, par contre, c'est la perception du bruit. Dans les années 80, le bruit " devient " pollution, facteur de dégradation de la qualité de vie. Les chercheurs se concentrent alors sur l'identification des sources du bruit, et sur les moyens d'abaisser les niveaux sonores. Les premières législations anti-bruit apparaissent, qui proposent/imposent une réduction du niveau des bruits produits, essentiellement, par les Transports et l'Industrie. familier

Mais le problème persiste, et pour cause, le bruit demeure un phénomène subjectif, autrement dit dépendant de l'appréciation de l'auditeur. Le bruit est affaire de contexte et beaucoup de lieux urbains sont appréciés aussi pour leur atmosphère vivante, c'est à dire, quelque part, "bruyante". Ville agréable ne rime pas nécessairement avec ville silencieuse. fam. fam.

Corriger l'environnement sonore uniquement suivant des paramètres acoustiques, par définition objectifs, ne suffit donc pas. Il est aujourd'hui communément admis que des mesures comme le L_{Aeq} ne peuvent, seules, rendre compte du confort sonore, et que vouloir " travailler " un environnement sonore uniquement sur la base des paramètres acoustiques, par définition objectifs, ne suffit pas(Aletta et al., 2016; Kang and Zhang, 2010; Schulte-Fortkamp and Fiebig, 2006; Yang and Kang, 2005).Il faut désormais envisager le bruit non plus seulement comme un objet physique, mais encore comme un objet cognitif(Guastavino, 2003). Le problème n'est plus de savoir à partir de quand un son est gênant, mais pourquoi il est perçu comme tel, et par tel individu. Les recherches se sont ainsi portées sur la notion de paysage sonore, envisageant la nuisance sonore et à travers les aspects qualitatifs, et à travers les aspects sémantiques des phénomènes acoustiques. pas encore introduit, si ?

De plus, si beaucoup d'efforts sont faits afin de réguler les niveaux bruits des sons non-désirés, l'approche inverse, i. e. ajouter des sons positivement connotés reste très peu considérée. Cette approche, consistant à identifier et agir sur les sons acceptés, ou plaisants, afin d'améliorer la qualité d'un environnement, est nommée l'approche positive par Schafer. De récentes études ont donné des résultats prometteurs, notamment (Hong and Jeon, 2013) qui montre que l'ajout de sons d'oiseaux ou d'eau à des sons de trafics urbain permet de significativement améliorer l'appréciation de ces derniers. ","

GL : TODO : citer (Galbrun and Ali, 2012)

Depuis vingt ans,l'approche par les paysages sonores a permis de développer une base de descripteurs qualitatifs et acoustiques grâce auxquels nous jugeons mieux, et sommes mieux à même d'améliorer l'environnement sonore urbain. (Kang, 2006; Schulte-Fortkamp et al., 2007)

Un des enjeux actuels de l'analyse des paysages sonores est de relier ces données perceptives, établies à partir d'enquêtes, à des mesures acoustiques, afin de pouvoir établir une politique de réduction du bruit efficace, adaptée à chaque situation (Schulte-Fortkamp, 2013). Cependant, le caractère pluridisciplinaire de ces recherches, et l'utilisation de protocoles expérimentaux variés pour évaluer l'environnement sonore, rendent l'intégration des résultats difficile (Davies et al., 2013). De plus, il n'y a toujours pas de consensus sur les descripteurs (acoustiques ou perceptifs) à utiliser pour caractériser un paysage sonore (Aletta et al., 2016; Brocolini et al., 2012), ce qui empêche la communauté, d'une part, de présenter aux décideurs des indicateurs génériques d'évaluation des paysages sonores, d'autre part, d'élaborer/proposer des modèles crédibles sur la base de ces expertises.

Récemment, plusieurs projets internationaux ont été lancés afin de standardiser les pratiques expérimentales des recherches portant sur les paysages sonores, notamment *the European Cooperation in Science and Technology Action*[7] (Schulte-Fortkamp and Kang, 2010) et *the Positive Soundscape project* (Davies et al., 2009; Davies et al., 2013), mais les difficultés persistent (Ribeiro et al., 2013; Schulte-Fortkamp, 2013).

3.4.3 *Approches catégorielle et dimensionnelles*

Deux grandes problématiques intéressent la recherche sur les paysages sonores :

- la première concerne la *représentation mentale des paysages sonores*. Comment nous représentons nous, en mémoire, un paysage sonore perçu ? La question en amène deux autres :
  1. Quelles sont les différentes catégories de paysages sonores ?
  2. Comment caractériser ces catégories ?

Je mettrais plutôt en footnote

Ouvrons là une parenthèse. Répondre à cette deuxième question revient à comprendre quels sont les éléments qui constituent un paysage sonore, et comment la nature de ses éléments influe sur le processus de catégorisation de l'environnement. Intuitivement, les éléments constitutifs d'un paysage sonore sont les sources sonores. Il s'agit alors également d'étudier la manière dont nous nous représentons ces sources. Fermons la parenthèse.

7 TD0804 : *soundscape of European Cities and Landscapes* : http://www.cost.eu/COST_Actions/tud/TD0804

- La deuxième concerne les *dimensions perceptives*. Quelles *dimensions perceptives* entrent en jeu dans l'évaluation subjective des paysages sonores ? Là encore la question en amène deux autres :
  1. Quels descripteurs perceptifs permettent de caractériser les dimensions perceptives à partir desquelles nous appréhendons l'environnement sonore ?
  2. Quels indicateurs influent sur ces descripteurs perceptifs ?

fam.

Développons. Les descripteurs perceptifs caractérisent les dimensions selon lesquelles nous interprétons l'environnement. Pour exemple, un des descripteurs perceptifs communément utilisé est l'agrément (Cf. Section 3.4.4 pour plus de détails sur les descripteurs couramment utilisés). Un des enjeux de l'approche dimensionnelle est de trouver les indicateurs qui influent sur ces descripteurs perceptifs. On distingue quatre types d'indicateurs.

- *Les indicateurs acoustiques/physiques* : Il s'agit d'indicateurs objectifs, obtenus via des mesures. Parmi ces indicateurs, certains caractérisent le niveau sonore par une approche holistique (L_{Aeq}), d'autres par une approche statistique (L_{A10-90}), d'autres encore en considérant séparément les différents canaux fréquentiels. On inclue, d'autre part, dans les indicateurs acoustiques/physiques, des indicateurs permettant de décrire les caractéristiques spectrales du son.
- *Les indicateurs perceptifs* : Les dimensions affectives suivant lesquelles nous percevons l'environnement sonore ne sont pas indépendantes. Ainsi certains descripteurs, comme l'agrément ou le confort, peuvent eux-mêmes servir d'indicateurs pour d'autres descripteurs plus généraux comme la qualité sonore. On inclue, d'autre part, dans les indicateurs perceptifs, des indicateurs procédant d'une évaluation subjective d'un attribut physique (e. g. niveau sonore perçue).
- *Les indicateurs psychoacoustiques* : Ces indicateurs sont à mi-chemin entre les indicateurs acoustiques et les indicateurs perceptifs. Comme les premiers, ils sont objectifs, calculés sur le signal sonore. Comme les seconds, ils sont perceptivement inspirés, i. e. construits afin de rendre compte d'une réalité perceptive. Pour exemple on cite la *loudness de Zwicker* qui rend compte du niveau sonore perçu.
- *Les indicateurs extra-sonores* : on regroupe ici tous les indicateurs qui ne sont pas liés au son. Certains sont liés au sujet (âge, genre, humeur), d'autres aux stimuli visuels, d'autres encore au temps. Contrairement aux indicateurs acoustiques, psychoacoustique et perceptifs, qui sont tous évalués/mesurés suivant des échelles ordonnées, discrètes

ou continues, certains indicateurs extra-sonores sont eux évalués sur des échelles de catégories[8] (e. g. Genre : homme/femme). On parlera alors plutôt de contexte extra-sonore.

Ces problématiques, pour rappel, la représentation mentale des paysages sonores, et les dimensions perceptives, sont à la base des deux grandes approches méthodologiques adoptées par la communauté scientifique, l'approche catégorielle, et l'approche dimensionnelle. On note cependant qu'avec le temps, la communauté privilégie l'approche dimensionnelle.

GL : TODO : Tableau résumant la totalité des indicateurs.

3.4.3.1 *Méthodologie de l'approche catégorielle*

Les objectifs de l'approche catégorielle sont triples. Il s'agit

- d'appréhender les principes psychologiques qui sous-tendent la formation des représentations mentales
- d'objectiver la nature de ces représentations
- de comprendre l'influence de ces représentations sur le traitement de l'information sonore

A ce titre, l'approche catégorielle peut être vue comme une approche cognitiviste.

Afin d'objectiver la nature des catégories mentales, représentant des paysages sonores ou des sources sonores, L'approche catégorielle a recourt à deux types d'expériences (Cf. Figure 11) :

- *Tâche de description* : On demande au sujet de décrire l'environnement sonore auquel il a été exposé (Axelsson et al., 2005; Guastavino, 2006; Raimbault, 2006; Raimbault and Dubois, 2005), soit de la manière la plus libre possible, soit en contraignant la description par le biais d'un questionnaire. Là encore, les réponses peuvent être libres (questionnaire semi-dirigé) ou à choix forcés (questionnaire dirigé). Plus la description est libre, plus on accède à des représentations mentales spécifiques au sujet. A contrario, plus le questionnaire est contraint, plus on accède à des représentations stéréotypées.

  L'analyse linguistique et lexicale des données ainsi collectées, permet d'en faire émerger les catégories. La richesse des descriptions, résultant de la liberté de réponse laissée au sujet, rend

8 Le terme catégorie employé pour décrire un type d'échelle n'a rien à voir avec le terme catégorie relatif au représentations mentales

cependant ce travail d'analyse difficile. Ces expériences de descriptions peuvent être réalisées en laboratoire, ou dans un cadre *in situ*.

- *Tâche de tri ou catégorisation* : On demande au sujet d'organiser les stimuli auxquels il vient d'être soumis, via une interface graphique le plus souvent, (Guastavino, 2007; Maffiolo, 1999), en groupes ou paquets, suivant une consigne fixée en fonction des objectifs mêmes de l'expérience. L'analyse de ces groupes permet d'en faire émerger les catégories, et de comprendre quels sont les attributs perceptifs à l'origine de l'organisation catégorielle proposée par le sujet. Il est par ailleurs possible de demander au sujet de nommer, voire de décrire ces groupes, afin d'acquérir encore plus de connaissances sur la nature des groupements effectués. On parle de catégorisation forcée lorsque que le nombre de groupes est contraint, et de catégorisation libre lorsque le sujet reste libre d'organiser les stimuli comme il l'entend. Ces expériences de tri sont pratiquées en laboratoire, en utilisant habituellement des enregistrements sonores comme stimuli.

L'avantage de ces pratiques expérimentales est double :

1. elles laissent une grande liberté au sujet dans ses réponses. En particulier, la tâche de catégorisation peut permettre de caractériser des stimuli sans imposer au sujet des dimensions ou attributs particuliers à partir desquels évaluer les sons, comme c'est notamment le cas pour l'analyse sémantique différentielle (Cf.Section 3.4.3.2). Ces dimensions imposées ne font pas forcément sens pour le sujet.

2. GL : TODO : citer (Dubois, 1991), justifier l'utilisation du langage comme senseur médian permettant d'objectiver les représentations mentales

GL : TODO : ajouter : discussion sur les outils d'analyse mds et analyse discriminante, ajouter les comparaisons par pairs

3.4.3.2 *Méthodologie de l'approche dimensionnelle*

L'approche dimensionnelle tente, elle, de caractériser les environnements sur la base de dimensions perceptives pré-établies. Comme nous l'avons vu, ces dimensions sont décrites par des descripteurs perceptifs.

Pour élaborer ses descripteurs, l'approche dimensionnelle a communément recours à l'analyse sémantique différentielle. Au cours de ces expériences, le sujet, à partir de ses ressentis, doit évaluer des descripteurs imposés en s'aidant d'échelles sémantiques bipolaires, ou

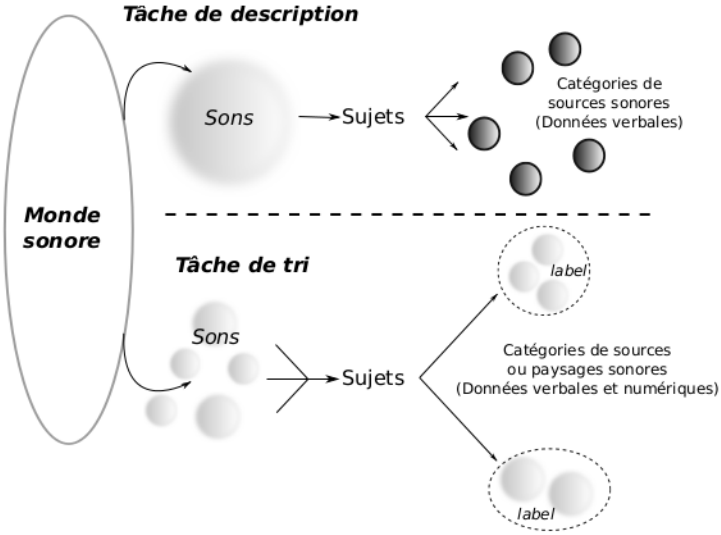


FIGURE 11 : Tâche de description et tâche de tri ou de catégorisation

échelles de Likert. Ces échelles répertorient l'ensemble des valeurs pouvant être prises par les différents descripteurs d'une scène sonore en cours d'évaluation. Elles formes un questionnaire à réponses fermées. En fonction des besoins de l'étude, elles sont discrètes ou continues, paires ou impaires. Cependant dans le cadre de l'évaluation des environnements sonores, on utilise généralement des échelles impaires et graduées en 7 (Raimbault, 2006), 9 (Hall et al., 2013) ou 11 points (Ricciardi et al., 2015).

La valeur sémantique des échelles tient du fait que les extrémités en sont bornées par des mots. Pour exemple, (Ricciardi et al., 2015) évalue la qualité d'un environnement ainsi que la présence de voitures à partir de deux échelles de 11 points chacune, et délimitées, l'une, par les termes désagréable (1) / plaisant (11) (*unpleasant/pleasant*), l'autre, par les termes rare (1) / fréquent (11) (*rarely/frequently*). Ces termes cadrent les réponses des sujets afin de s'assurer que tous interprètent l'échelle de la même manière, i. e. en attribuant peu ou prou la même valeur à chacune des graduations. Ce fait est néanmoins difficilement vérifiable en pratique, les termes extrêmes pouvant revêtir un sens différent en fonction des sujets.

L'évaluation à partir d'échelles sémantiques peut être réalisée en laboratoire, via une interface machine, ou dans un cadre *in situ*, par le biais de questionnaires papiers (Jeon et al., 2013; Torija et al., 2013), ou, comme c'est de plus en plus le cas, au moyen d'une application sur téléphone portable (Kardous and Shaw, 2014; Ricciardi et al., 2015). L'outil présente plusieurs avantages. Il permet une collecte des données sur serveur directement, et il offre la possibilité d'enregistrer les environnements entrain d'être évalués, ce qui répond en partie aux problèmes inhérents aux études *in situ*, notamment en ce qui concerne la reproductibilité des stimuli (Cf. section 3.1.4).

L'intérêt que suscite l'approche dimensionnelle réside dans le fait que les résultats obtenus sont facilement analysables et interprétables. Évaluer un environnement sonore au moyen d'échelles sémantiques et d'indicateurs objectifs permet d'obtenir une description sous la forme de descripteurs quantitatifs. Or Il existe nombre de tests statistiques (Cf. Annexe A) ou d'outils d'analyse dimensionnelle (Cf. Annexe B) directement applicables à ces données.

En fonction des objectifs de l'étude, on peut distinguer trois approches méthodologiques :

- *L'identification des descripteurs perceptifs*. Dans cette approche, on distingue déjà deux méthodes :
  - Dans la première, l'expérimentateur identifie des descripteurs perceptifs pertinents, sans avoir d'idée pré-établie sur leur nature. Ces descripteurs sont habituellement détectés à partir de l'analyse lexicale des descriptions verbales fournies par les sujets.
  - Dans la seconde, l'expérimentateur, au sein d'un groupe de descripteurs perceptifs et/ou objectifs donné, sélectionne ceux qui rendent compte au mieux de l'évaluation des paysages sonores. Les descripteurs objectifs sont calculés à partir du signal sonore, les descripteurs perceptifs sont évalués sur des échelles sémantiques. Différentes techniques d'analyse statistique multidimensionnelle, comme l'analyse par composantes principales ou le positionnement multidimensionnel (*Multidimensional scaling*, cf. Annexe refapp :statmulti), permettent de faire émerger des dimensions linéairement non-corrélées qui rendent compte au mieux de la variabilité des données (Cain et al., 2013; Torija et al., 2013). Ces nouvelles dimensions n'ayant pas de valeurs physiques ou perceptives a priori, une inspection qualitative des scènes sonores est alors nécessaire afin de les caractériser. Il est par ailleurs possible de tester d'éventuelles corrélations entre les descripteurs perceptifs et/ou entre les descripteurs objectifs. Par exemple, (Torija et al., 2013) évalue les corrélations entre 15 descripteurs perceptifs et 49 indicateurs acoustiques via l'utilisation du coefficient de Pearson.

- *L'étude de l'influence des indicateurs sur le descripteur* : L'expérimentateur, à partir d'un descripteur perceptif, et d'une série d'indicateurs objectifs ou perceptifs donnés, évalue dans quelle mesure l'évolution du descripteur est contrainte par les indicateurs, l'objectif a terme étant de d'obtenir un modèle prédictif de la variation du descripteur. Dans ce but, (Lavandier and Defréville, 2006; Ricciardi et al., 2015) se servent tous les deux de la régression linéaire multiple (Cf. Annexe B) afin de modéliser

la variation de la qualité sonore en fonction d'indicateurs perceptifs (niveau sonore perçu, familiarité avec l'environnement, présence des sources sonores de voix) et objectifs (L_{Aeq}, L_{A10}-L_{A90}).

- *La classification non-supervisée des scènes sonores*. Là encore, dans cette approche, on distingue deux méthodes :
  - Dans la première, l'expérimentateur, considérant des classes de paysages sonores données (e. g. parc, rue, marché), vérifie si les descripteurs varient d'un type d'environnement à l'autre. Il dispose d'outils statistiques (Cf. Annexe A) permettant de tester l'existence de différences significatives entre tel ou tel type d'environnements sonores (Hong and Jeon, 2013).
  - Dans la seconde, l'expérimentateur, considérant un ensemble de paysages sonores, analyse directement l'espace décrit par l'ensemble des descripteurs afin de faire émerger des groupes de scènes sonores similaires, au sens des descripteurs. Il peut avoir recours à des techniques de clustering, comme par exemple le clustering hiérarchique ascendant (Torija et al., 2013) ou encore des méthodes non-supervisées, inspirées des réseaux de neurones, comme les cartes auto-organisées (*Self Organized Map*, *SOM*) (Ricciardi et al., 2015). Les différents descripteurs pouvant être corrélés entre eux, il lui est également possible d'utiliser des outils d'analyse dimensionnelle, comme l'analyse par composante principale, permettant de générer de nouvelles dimensions décorrélées, et de sélectionner celles qui expliquent le mieux la variance des données.

Contrairement à l'approche catégorielle, l'approche dimensionnelle laisse peu de liberté au sujet, ce dernier étant contraint d'utiliser les échelles qui lui sont présentées pour décrire l'environnement. L'utilisation de ces échelles suppose que les attributs qu'elles décrivent puissent être évalués de manière linéaire et uni-dimensionnelle, ce que le sujet n'est pas toujours en mesure de réaliser. (Raimbault, 2006) montre notamment que des échelles sensées évaluer la structure temporelle d'une scène sonore (stable/instable ou figé/évolutif) ne conviennent pas, cette notion n'étant pas comprise par les sujets comme étant bipolaire.

L'utilisation d'échelles comprend par ailleurs plusieurs risques :

- Les échelles peuvent être mal interprétées par le sujet, ou même ne pas faire sens. Une description détaillée des échelles, ainsi que l'utilisation de plusieurs mots pour en définir les extrémités permettent de pallier ces difficultés. (Hall et al., 2013) évalue ainsi l'agrément en utilisant une échelle de 9 points dont les

extrémités sont décrites par les triplets désagréable-mécontent-insatisfait / agréable-content-satisfait. Ce biais peut encore être réduit en apportant un soin particulier à la sélection des termes extrêmes afin de s'assurer que ces derniers soient bien appropriés, par exemple, en menant une expérience intermédiaire sur la base d'un questionnaire libre (Guastavino and Katz, 2004), réalisable en condition *in situ* (Hong and Jeon, 2013; Kang and Zhang, 2010), ou en demandant au sujet d'expliquer verbalement sa notation (Raimbault, 2006).

- Tous les sujets peuvent ne pas utiliser les échelles de la même manière. Certains sont portés à en utiliser toutes les valeurs. D'autres peuvent n'en privilégier que certaines, et notamment écarter les extrêmes (sans que, d'ailleurs, il soit possible de déterminer si ces variances entre les sujets sont involontaires ou, au contraire, décidées). Une normalisation des données, avant analyse, est possible, pour réduire l'impact de ce biais (Defréville et al., 2004; Hong and Jeon, 2013; Lavandier and Defréville, 2006; Nielbo et al., 2013). Cette normalisation est obligatoire, s'agissant d'échelles de notation (e. g. attribution d'une note entre 0-10, 0-100 etc ldots) a priori non bornées de termes aux extrémités. Rien ne garantit, en effet, que la valeur subjective donnée à une note (e. g. 5/20) soit la même pour tous les sujets. Elle est moins pertinente s'agissant d'échelles sémantiques. S'ajoute à cela le fait que les données provenant d'analyses sensorielles comprennent souvent des réponses extrêmes (*outliers*). La normalisation, dans ce cas, peut fausser sensiblement les données.

- En général, pour un environnement donné, la valeur finale d'une échelle est calculée en moyennant les réponses de plusieurs sujets. Pour être valide, cette approche suppose que la distribution des réponses sur l'échelle soit unimodale. Or il a déjà été montré que ces distributions peuvent être multi-modales, du fait, entre autre, des variations d'interprétations de l'échelle entre les sujets, ou des différences d'appréciation relatives à d'autres facteurs (Raimbault, 2006). Il peut être utile d'inspecter les distributions des réponses avant de considérer des résultats moyennés.

Ainsi, dans le cadre de l'approche dimensionnelle, il est important de s'assurer que :

1. les échelles soient aptes à décrire les attributs qu'elles décrivent
2. les échelles soient correctement interprétées par les sujets

3.4.4 *Analyse et Modélisation des propriétés perceptives des paysages sonores*

Nous détaillons dans la suite de cette section les descripteurs perceptifs ayant fait l'objet d'une attention particulière dans les approches dimensionnelles. Il est a noter qu'il n'existe pas de consensus dans la communauté sur :

1. la définition de ces descripteurs
2. les pratiques expérimentales permettant d'étudier ces descripteurs

Par pratique expérimentale, nous comprenons entre autre, la nature des échelles à utiliser (nombre de points, termes aux extrémités), leur analyse, ainsi que l'application d'éventuelles étapes de normalisation (Aletta et al., 2016).

3.4.4.1 *Gêne et bruit*

La gêne provoquée par un paysage sonore, et en particulier par un paysage sonore urbain est un des descripteurs perceptifs les plus étudiés. Ce fait est notamment lié au besoin pressant de trouver une solution à la pollution sonore en ville. Une récente étude indique d'ailleurs que 86% des français se disent gênés par le bruit extérieur lorsqu'ils se trouvent chez eux (Bendavid and Chasles-Parot, 2014).

G1 : (Marquis-Favre et al., 2005a,b) États de l'art
G2 : (Gille and Marquis-Favre, 2016; Gille et al., 2016a,b; Klein et al., 2015; Marquis-Favre and Morel, 2015; Morel et al., 2016; Trollé et al., 2015) trafic et bruit

3.4.4.2 *Qualité sonore*

G1 (Brocolini et al., 2012; Ricciardi et al., 2015)
G2 (Hong and Jeon, 2013) utilise la notion de préférence, lien avec le confort acoustique.
G3 (Ozcevik and Can, 2012)
G4 (Yu and Kang, 2010) preference
G5 (Nilsson et al., 2007; Nilsson and Berglund, 2006) niveau sonore mieux que contenu spectral et fluctuation temporelle

3.4.4.3 *Agrément*

G1 : (García Pérez et al., 2012)
G2 : (Lavandier and Defréville, 2006)
G3 : (Guillén and López Barrio, 2007)

3.4.4.4 *Confort acoustique*

G1 : (Meng et al., 2013; Yang and Kang, 2005)
G2 : (Jeon et al., 2011, 2013)
G3 : (Tse et al., 2012)
G4 : (Yu and Kang, 2009) modèle du confort

3.4.4.5 *Calme et tranquillité*

G1 : (Delaitre et al., 2012) définition du concept
G2 : (Pheasant et al., 2009; Pheasant et al., 2008)
G3 : (Memoli et al., 2008)
G4 : (De Coensel and Botteldooren, 2006; De Coensel et al., 2013) influence musicale
G5 : (Payne, 2013) notion de restauration

3.4.4.6 *Propriétés combinées*

G1 : (Davies et al., 2013)
G2 : (Cain et al., 2013)
G3 : (Axelsson et al., 2010)
G4 : (Hall et al., 2013) émet des réserves sur la justesse de cette approche.

3.4.4.7 *Autre descripteurs*

G5 : (Kang and Zhang, 2010) autres

3.4.4.8 *Influence d'attributs extra-sonores*

G1 : Contexte environnementale (température, chaleur, humidité), (Jeon et al., 2011; Meng et al., 2013)
G2 : Contexte socio-culturel (Hall et al., 2013; Yu and Kang, 2010)
G2 : Contexte socio-culturel (Guillén and López Barrio, 2007) pas d'effet
G3 : Contexte spatial (Hall et al., 2013)
G4 : affordances/appropriateness (Hall et al., 2013) G5 : visuel (Guillén and López Barrio, 2007)

3.4.4.9 *Réponse physiologique*

G1 : (Hume and Ahtamad, 2013)

3.4.5 *Catégoriser les sources et paysages sonores*

3.4.5.1 *Catégories de sources sonores*

Parmi les travaux les plus influents sur la catégorisation des sources sonores, on trouve ceux de W. W. Gaver (Gaver, 1993a,b). Se plaçant

dans le cadre de l'"écoute de tout les jours" (cf. section 3.1.3), il propose d'envisager le problème sous un angle phénoménologique, en considérant l'action physique à l'origine, plutôt que l'objet. L'action étant très dépendante de la nature physique de l'objet, il trie les causes suivant trois types d'objets :

- Solide : heurter, gratter, rouler, déformer
- Liquide : goutter, éclabousser, clapoter
- Gaz : exploser, venter bof... (souffler ?)

D'autres études ont montré l'importance des phénomènes physiques originels dans les processus de catégorisation et d'interprétation des sons (Lemaitre et al., 2010; Marcell et al., 2000). (Houix et al., 2012) notamment demande à des sujets de catégoriser librement 60 sons environnementaux en se concentrant sur l'action, et de commenter leurs groupes. Une analyse des groupements révèle que la catégorisation s'opère suivant deux étages hiérarchisés, le premier, le plus général, comprenant des catégories d'objets très proches à celles proposées par Gaver (Solide, Liquide, Gaz et Machine), et le deuxième, plus spécifique, comprenant des catégories d'actions. Une seconde expérience similaire, réalisée uniquement sur des objets solides, montre que la nature du pattern temporel (continu ou discret) résultant de l'action à l'origine du son influe de manière significative sur la catégorisation. Ce dernier point est également observé par (Gygi et al., 2007). Sur la base d'une matrice de similarité obtenue à partir de comparaisons par paires, et via un positionnement multidimensionnel en trois dimensions, Gygi et *al.*montre que les sons s'organisent en trois clusters incluant les sons harmoniques, les sons d'impacts (discret) et les sons continus. Une épreuve de catégorisation semi-libre (les sujets devant réaliser au minimum 5 clusters), avec verbalisation pratiquée dans la même étude, montre par ailleurs que les sujets catégorisent les sons principalement en fonction du type de sources (animaux, homme, véhicule, mécanique, musique, eau), moins fréquemment en fonction du contexte et du lieu (extérieur, sport, bar) et rarement sur la base de caractéristiques physiques isolées (hauteur, fréquence) ou d'émotions ressenties (ennuyant, alarmant) eux

L'influence de la nature de la source, et de sa sémantique associée, sur les processus de catégorisation a particulièrement été étudiée. A partir d'une étude de 35 papiers traitant de catégories sonores, (Niessen et al., 2010) établie une liste de 20 catégories de sons les plus citées. La liste est présentée dans le tableau 1.

La grande majorité de ces catégories sont des catégories de sources sonores. Seules deux font référence à des catégories d'objets sonores abstraits (Événement, Background). Ces catégories de sources ne s'expriment pas toutes au même niveau d'abstraction. Certaines sont précises (*Aboiement chien*), d'autres sont très larges (*Sons naturels*). Par

Nom des catégories les plus citées		
Sons naturels	Voix	Voiture
Sons humains	Enfant	Machine
Sons technologiques/artificielles	Cloche	Vent
Trafic	Background	Aboiement chien
Oiseaux	Événement	Bruits de pas
Musique	Avion	
Travaux	Eaux	

TABLE 1 : Les catégories sonores les plus citées, d'après (Niessen et al., 2010)

ailleurs, certaines sont incluses dans d'autres (*Bruits de pas<Sons humain*), les trois catégories ayant le périmètre le plus large, et englobant toutes les autres sont *Sons naturels*, *Sons humains* et *Sons technologiques/artificielles*. Comme nous allons le voir (cf. section 3.4.5.2 et 3.4.7) c'est en partie suivant ce découpage catégorielle que s'opère la perception de l'environnement.

Plusieurs études se basent sur une analyse linguistique de descriptions spontanées et libre d'environnements sonores, afin d'établir des catégories de sources sonores. Dans ces études, il est d'usage de demander explicitement au sujet de distinguer les aspects plaisants et désagréables du paysage étudié.

En réalisant une étude *in situ* d'environnements de parcs, (Szeremeta and Zannin, 2009) mettent en évidence 9 catégories de sources sonores. Certaines sont systématiquement positivement connotées (*oiseaux*, *nature*), d'autres négativement connotées (*machine*, *alarme/signaux*, *train*), d'autres encore peuvent être jugées soit positives soit négatives comme *personne* (majoritairement positive), *trafic véhicule* (majoritairement négative), *musique*, *trafic aérien*.

L'étude de (Guastavino, 2006) utilise une méthode d'analyse similaire, mais en demandant aux sujets de décrire un environnement urbain idéal (plaisant), sur la base de leur mémoire uniquement. Des résultats similaires sont observés, i. e.les catégories *oiseaux* et *nature* sont systématiquement positivement perçues, les catégories *klaxon* et *travaux* négativement perçues, les catégories *personne* (majoritairement positivement connotée) et *musique* ayant une connotation variable.

L'auteur fait remarquer que les sujets décrivent les sons en s'appuyant sur la source émettrice de ces derniers. Il y a donc une assimilation entre l'objet et le phénomène acoustique. En conséquence la sémantique (le sens) liée à l'objet intervient dans le processus perceptif (dans ce cas le jugement hédonique) au même titre que les propriétés acoustiques. L'observation des appréciations inhérentes aux catégories de véhicules vont dans ce sens : les catégories *trafic* (*voiture*, *moto/scooter*, *camion*) sont systématiquement négativement perçues, à

la différence des catégories *transports publics* (*bus* et *train*), toujours bien perçues. La représentation positive que nous avons des *bus* fait que ces sons, bien que proches de ceux de véhicules individuels, sont largement bien acceptés.

Nous noterons cependant que l'étude de (Guastavino, 2006) est réalisée sans support sonore, les sujets n'ayant que leur mémoire pour se représenter l'environnement urbain idéal. On peut penser que dans ce cas, les attributs sémantiques sont particulièrement sollicités. Nous

approfondissons

questionnons ce point à la section 5.2.11.1.

3.4.5.2 *Catégories de paysages sonores*

Outre les catégories de sources sonores, plusieurs études s'intéressent à la formation de catégories d'objets plus complexes, les paysages sonores.

V. Maffiolo (Maffiolo, 1999) montre l'existence de deux processus distincts engagés, en fonction de la capacité de l'auditeur à identifier des événements sonores. Dans cette étude, les sujets doivent 1) catégoriser des enregistrements d'environnements sonores urbains, et 2) décrire les groupements effectués. A partir d'une analyse linguistique des descriptions verbales, Maffiolo montre l'existence de deux catégories cognitives abstraites d'environnements sonores respectivement appelées : "les séquences événementielles" et "les séquences amorphes". Les séquences événementielles sont des environnements composés d'événements saillants et identifiables (*démarrage de voiture*, *voix d'homme*). Les séquences amorphes sont des environnements dont il est difficile d'isoler des éléments distincts.

Chacune de ces catégories a été sous catégorisée suivant différentes stratégies :

"-"

- les scènes événementielles ont été sous catégorisées en fonction
  1. du type source présent
  2. de l'agrément perçu (agréable/désagréable)
- les scènes amorphes ont été sous catégorisées en fonction
  1. de l'évaluation des propriétés acoustiques (intensité sonore, contenu spectral)
  2. de l'agrément perçu (agréable/désagréable)

On remarque ainsi que les scènes événementielles profitent d'une analyse descriptive basée sur l'identification des sources sonores, alors que les scènes amorphes bénéficient d'une analyse holistique, à partir d'indicateurs acoustiques (subjectifs) globaux. On note que les deux catégories suscitent un jugement hédonique.

Cette distinction (événementiel/amorphe) s'opère aussi au niveau de la source sonore. Analysant des descriptions libres des sources

sonores peuplant l'environnement urbains, Guastavino montre que les descriptions des sons à basse fréquence peuvent se diviser en deux catégories appelées " événements sonores " et " bruit de fonds ". Dans les derniers, aucune source ne peut être identifiée.

Raimbault et Dubois (Raimbault and Dubois, 2005), combinant les résultats obtenus par trois thèses ((Maffiolo, 1999),(Raimbault, 2002),(Guastavino, 2003)), montrent que la catégorisation des paysages sonores s'opère, suivant leurs compositions, en terme de sources sonores (cf. figure 12). Une première distinction s'opère entre d'un coté les paysages sonores comportant des sons de *transports motorisés* et/ou de *travaux*,et de l'autre, des paysages sonores comportant des sons suggérant une présence humaine. Ces derniers se subdivisent encore entre d'une part les paysages " vivant " et d'autre part les paysages " relaxant ", composés également de sons de *nature*. Le rôle prédominant joué par l'activité humaine dans la catégorisation des environnements était déjà pressentie par Schaefer (Schafer, 1977).

Des résultats très similaires sont obtenus par (Guastavino, 2007). Passant par une tâche de catégorisation libre avec verbalisation, Guastavino montre que la catégorisation s'opère suivant la présence/absence de sons d'origine humaine, ainsi que sur un jugement hédonique (plaisant/non-plaisant) des sources. La présence de sons humains agit sur deux niveaux, 1) les environnements sont divisés entre ceux dominés par les sons humains, et ceux dominés par les sons mécaniques. 2) Les premiers se subdivisent en fonction de l'activité et du lieu (parc calme, marché actif). Les seconds se subdivisent encore à partir de la présence ou non de sons humains.

mettre cette explication à ton premier usage de "hédonique", page préc.

3.4.6 *Classifier les sources et environnements sonores*

Contrairement à la section précédente, où il est question de catégories, i. e. représentation mentale, nous traitons ici de classes. Par classe nous entendons un groupe d'objets qui ne fait pas référence à une entité mentale particulière, mais dont le regroupement vient d'une volonté de classer/d'organiser des environnements, ou des sources, suivant leurs caractéristiques physiques, morphologiques, ou encore suivant leurs fonctions. Le but est alors, sur la base de descripteurs objectifs, d'étudier les similarités existant entre ces groupes,(cf. section refsec :appDimensionelle).

3.4.6.1 *Classes de sources sonores*

Un des buts premiers des études sur les classes sonores est d'établir la typologie complète de tous les types de sources peuplant un environnement donnée.

Sur la base de l'étude de (Raimbault and Dubois, 2005), et dans l'idée de proposer une nomenclature générique pour décrire les sources sonores présentes en milieu urbain, (Brown et al., 2011) propose une

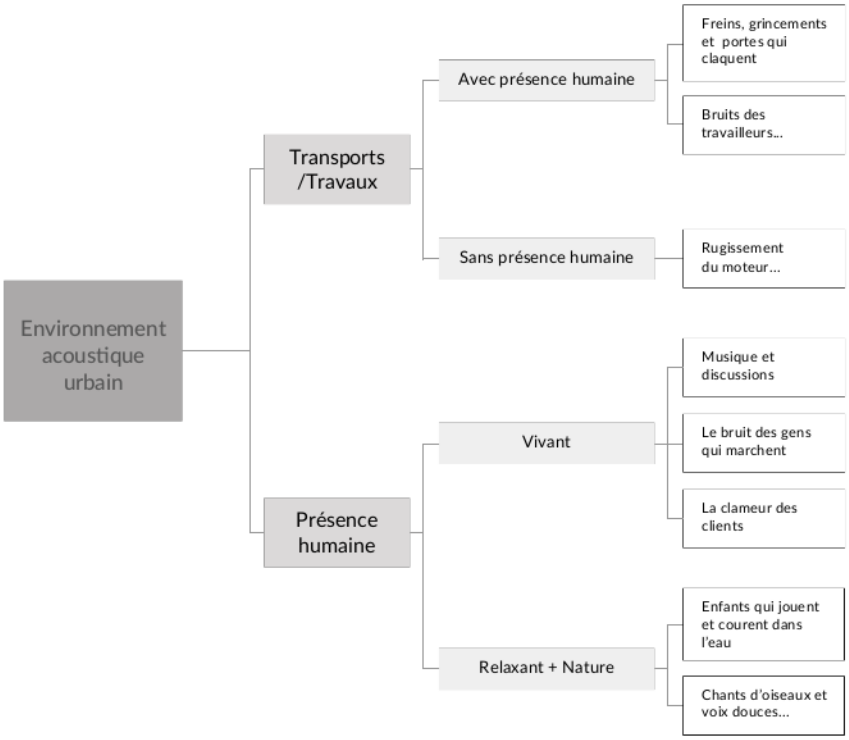


FIGURE 12 : Catégorisation des paysages sonores urbains, d'après (Raimbault and Dubois, 2005)

taxonomie reprise à la figure 13. Cette classification est centrée sur l'objet. En partant de la taxonomie proposée par (Brown et al., 2011), (Salamon et al., 2014) propose une nouvelle taxonomie, plus détaillée, centrée, elle, à la fois sur l'objet et sur l'action (cf. figure 17). Les auteurs partent de l'idée que la réalité sonore d'un objet diffère en fonction de son utilisation (*passage de voiture vs. freinage de voiture vs. klaxon de voiture*). Pour rendre compte de ce fait, certaines classes d'objets du plus bas niveau sont subdivisées en classes d'actions, labellisées par des verbes.

Outre organiser les sources, il est aussi utile de comprendre quelles sont les différences acoustiques qui peuvent se manifester entre plusieurs classes de sons. (Yang and Kang, 2013), sur la base d'indicateurs acoustiques et psychoacoustiques, compare des classes sons provenant d'environnements urbains (*musique*, *mécaniques* et *trafic*) et d'environnements naturels (*eau*, *vent* et *oiseaux*). Chaque indicateur est calculé sur le signal, à l'aide d'une fenêtre glissante, et moyenné. En réalisant une analyse par composante principale sur ces indicateurs, les auteurs montrent que l'intensité (*loudness de Zwicker*), le contenu spectral *sharpness* et la structure temporelle *fluctuation* sont les trois principaux indicateurs permettant d'expliquer la variance entre ces différents types de sons. Ce fait avait déjà été observé dans

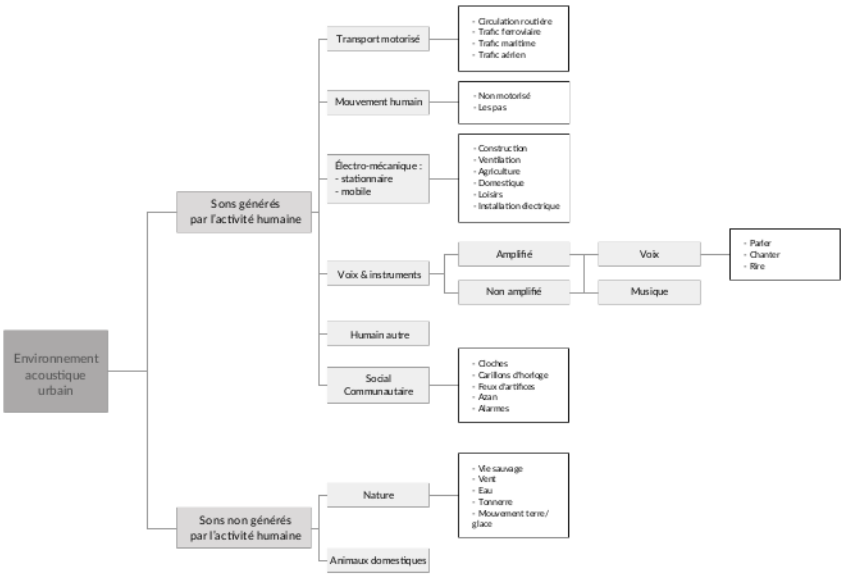


FIGURE 13 : Taxonomie des sources sonores urbaines, d'après (Brown et al., 2011)

d'autres études mêlant différents stimuli (Botteldooren et al., 2006; De Coensel and Botteldooren, 2006).

3.4.6.2 *Classes de paysages sonores*

Beaucoup d'études analysent l'existence de similarités entre environnements sonores à partir d'indicateurs quantitatifs, qu'ils soient objectifs (Rychtáriková and Vermeir, 2013), subjectifs (Jeon et al., 2013), ou les deux à la fois (Ricciardi et al., 2015; Torija et al., 2013). La méthodologie est presque toujours la même :

1. Pour chaque environnement, calculer des indicateurs acoustiques/psychoacoustiques, ou évaluer des indicateurs perceptifs à l'aide d'échelles sémantiques

2. Utiliser des outils de clustering afin d'établir des classes d'environnements similaires.

Sur la base de descripteurs subjectifs uniquement, (Jeon et al., 2013) identifient quatre classes comprenant respectivement, des environnements dominés par le bruit urbain, des environnements comprenant majoritairement des composantes naturelles, des environnements urbains ouverts (place), ou encore des environnements équilibrés (sons urbains et naturels). La distinction se fait en grande partie à partir d'indicateurs de préférence liés au confort acoustique, mais également à l'impression visuelle et à la configuration spatiale du lieu.

Sur la base de descripteurs subjectifs et objectifs, (Torija et al., 2013) établit 15 classes de paysages sonores. Il apparaît que La distinction correspond à des différences au niveau des sources sonores présentes/absentes (trafic, oiseaux, fontaine, moto, sirène, parc, humain). Parmi les indicateurs acoustiques, ceux tenant compte de la dynamique du niveau sonore (*crest factor*) ainsi que du niveau des basses fréquences (L à 125Hz) permettent à eux seuls d'expliquer 84% de la variance des données. Les auteurs concluent que l'utilisation de descripteurs acoustiques peut permettre, seule, d'isoler les paysages sonores similaires, conclusion reprise par *rychtarikova2013soundscape*.

GL : TODO : contredire cette vérité

3.4.7 *Prendre en compte les contributions séparées des différentes sources sonores*

Comme nous l'avons vu, plusieurs études adoptant l'approche catégorielle, ont permis de montrer que l'identification de certaines sources sonores, ainsi que leur sémantique associée, joue un rôle important dans l'évaluation perceptive des paysages, particulièrement au niveau de l'agrément perçu (Guastavino, 2006; Szeremeta and Zannin, 2009).

Continuant dans ce sens, les études adoptant l'approche dimensionnelle cherchent de plus en plus à compléter les indicateurs globaux avec des indicateurs caractérisant les contributions spécifiques de différentes sources sonores. Pour ce faire, elles partent toutes d'une liste de catégories de sources pré-établie. A partir de cette liste, elles calculent des indicateurs acoustiques spécifiques à ces sources, et/ou demandent à des sujets d'en évaluer les caractéristiques perceptives.

En menant différentes études *in situ* sur la qualité de différents environnements, (Nilsson et al., 2007; Nilsson, 2007) montrent que l'identification des sources sonores permet de mieux prévoir la qualité globale de l'environnement que le niveau sonore. En particulier les sons *technologique/mécanique* ont un impact négatif sur l'environnement alors que les sons *naturels* ont un impact positif. Les sons *humain* restent cependant neutre. De plus, l'étude montre que dans le cas d'une exposition modérée au bruit de trafic, l'ajout de sons positivement perçus (*naturels* dans leur cas) peut potentiellement améliorer la qualité de l'environnement. Une observation déjà effectuée par d'autres études. Cependant, pour une exposition élevée au bruit, une politique de réduction des niveaux est obligatoire.

verbe

(Defréville et al., 2004; Lavandier and Defréville, 2006) évaluent l'impact séparé de différentes sources de trafic (*voiture*, *Moto*, *scooter*, *bus*), de sons humains (*voix adultes*, *voix enfants*) et de sons naturels (*oiseaux*) sur l'agrément perçu. Pour chacune de ces sources ils calculent des indicateurs objectifs de niveaux (L_{Aeq}, L_{A10}) et de présence (nombre d'occurrences, pourcentage de temps présent), ainsi

que des indicateurs perceptifs (présence, proéminence, proximité). Des indicateurs globaux relatifs au niveau (objectif : *loudness de Zwicker* ; subjectif : niveau perçu) sont également pris en comptes. La régression linéaire multiple est utilisée afin de mesurer l'influence des indicateurs sur l'agrément.

Que l'on considère les indicateurs subjectifs ou objectifs, l'utilisation combinée de l'indicateur de niveau global avec les indicateurs spécifiques aux différentes sources permet d'augmenter la capacité de prédiction de la qualité sonore, comparé à l'utilisation de l'indicateur de niveau global seul. Là encore les auteurs montrent que dans le cas où les environnements sont peu exposés au trafic, les sons d'*oiseaux* et d'*humain* ont un effet positif, la qualité augmentant en fonction de leur présence. Ils notent également que l'appréciation des *voitures* diffère en fonction du type d'environnement : dans un parc, elles ont un effet négatif alors que dans une rue, elles sont comprises comme faisant partie de l'environnement et n'influence pas (de manière individuelle) la qualité perçue.

Dans une étude d'envergure, comprenant 3400 réponses collectées sur deux cites (Paris et milan), et utilisant une méthodologie proche de celle de (Lavandier and Defréville, 2006), Ricciardi et *al.*(Ricciardi et al., 2015) testent plusieurs modèles permettant de prédire la qualité sonore. Ces modèles sont tous bâtis à partir d'indicateurs perceptifs globaux, sonores et visuels, ainsi que d'indicateurs perceptifs sonores spécifiques à différentes sources. Les modèles tenant compte des indicateurs visuels produisent des sorties corrélées à 72% avec la qualité mesurée. Cette corrélation décroît à 58% si l'on supprime les indicateurs visuels, et tombe à 19% si l'on ne considère plus que le niveau sonore global (sans les indicateurs spécifiques aux sources). Les auteurs clusterisent les différents environnements sur la base de ces indicateurs. 6 classes sont mises à jours, les regroupements étant encore une fois relatifs à la présence/absence de diverses sources sonores. Plus spécifiquement, certains groupements sont liés :

1. à la possibilité de distinguer, ou non, des sources sonores dans les scènes (scènes événementielles *vs.* amorphes)
2. à la présence majoritaire d'une classe de sons en particulier (*trafic*, *humain*, *nature*)
3. à la présence simultanée de plusieurs sources

En recalculant des modèles pour chacune des classes, les auteurs montrent que les indicateurs relatifs à des sources sont plus représentés dans le cas des modèles par classes, mais varient significativement d'une classe à l'autre. Par exemple, l'indicateur relatif aux sources d'*oiseaux* n'apparaît, dans le modèle, que pour la classe dominée par des sons *naturels*. Ces résultats questionnent l'utilité et l'efficacité d'une modélisation de la qualité sonore qui se voudrait générale, i. e. applicable pour tout type de situations et d'environnements.

3.5 ÉVÉNEMENTS ET TEXTURES SONORES

3.5.1 *Définition*

S'éloignant de l'approche des paysages sonores, plusieurs études se sont concentrées sur l'analyse perceptive d'un certain type de sons, appelés texture sonore.

Pour définir la texture sonore, nous nous appuyons sur la définition donnée par (Saint-Arnaud, 1995, p. 25):

- " Les textures sonores sont des objet composites, formés d'éléments de base appelés atomes "
- " Les atomes apparaissent suivant un pattern haut-niveau pouvant être soit périodique (galop), soit aléatoire (pluie), voire les deux "
- " Les caractéristiques haut niveaux des textures restent constantes sur de longue périodes de temps, ce qui implique qu'elles ne peuvent comporter aucun message complexe "
- " Le pattern haut-niveau doit être présenté au moins une fois dans sa totalité pendant une période de temps n'excédant pas les quelques secondes. Cette période est nommée période d'attention (*attention span*) "

Cette définition est avant tout morphologique, la texture étant définie en fonction de ses caractéristiques physiques. Cela vient, entre autre, du fait que la texture a d'abord été étudiée dans le cadre du traitement du signal, beaucoup d'applications multimédia ayant besoin de modèles permettant de synthétiser de tels sons (Schwarz, 2011). La notion de texture s'oppose intuitivement à celle d'événement sonore et de séquence d'événements. Par opposition, l'événement est vu comme un élément discret, un son court et non homogène.

C'est par la notion d'information transmise que semble se faire la distinction entre texture et séquence d'événements. Les caractéristiques des textures restant stables au cours du temps, l'information transmise finit éventuellement par atteindre une asymptote. A contrario, une succession d'événements distincts, comme c'est le cas pour une séquence musicale ou de parole, transmet de plus en plus d'informations dans le temps (Cf. Figure 14). En poussant le raisonnement à l'extrême, le bruit blanc peut être vu comme la représentation la plus " pure " d'une texture, ce dernier étant porteur d'une information très limitée.

Cette dimension événement/texture est orthogonale à celle de " bruit de fond " / " événements de premier plan " (*background/foreground*), utilisée dans le langage courant pour discriminer l'environnement urbain. Concernant les notions de background et de foreground, nous

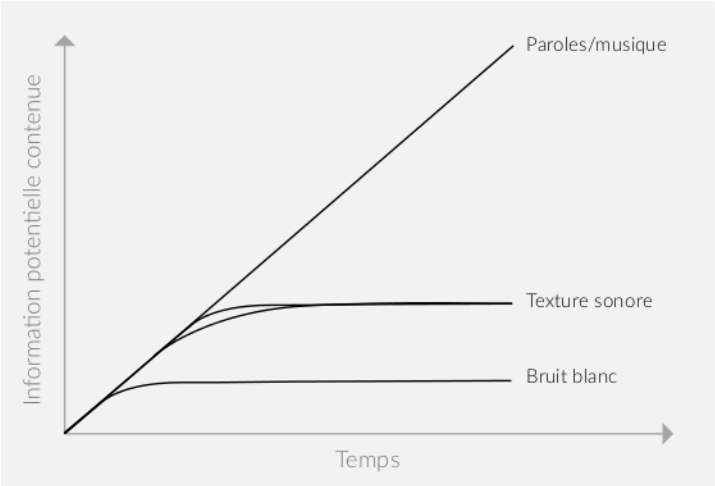


FIGURE 14 : Information potentielle contenue dans les séquences d'événements, les textures, et le bruit. D'après (Saint-Arnaud, 1995)

considérons que l'une et l'autre peuvent être vues comme des flux auditifs, ces derniers pouvant être composés à la fois de textures et d'événements regroupés dans le but de faciliter le traitement auditif de la scène.

3.5.2 *Percevoir les textures*

Contrairement aux événements sonores, la texture est un objet simple, dont le traitement cognitif ne requiert pas une analyse poussée.

Cela a été mis en évidence par Josh H. McDermott et ses-co auteurs (McDermott and Simoncelli, 2011; McDermott et al., 2013). S'inspirant du fonctionnement de l'oreille humaine, et notamment des processus auditifs intervenant depuis la cochlée, jusqu'au thalamus, ils ont pu établir un modèle permettant de re-synthétiser des textures sonores en ne se servant que de statistiques simples, calculées à partir de représentations temps-fréquence de signaux de textures enregistrés.

Dans une première expérience (McDermott and Simoncelli, 2011), la capacité des sujets à identifier les textures synthétisées a été testée . Les résultats ont montré que les sons de synthèse étaient aussi bien identifiés que les sons enregistrés. McDermott démontre ainsi qu'une information résumée sous la forme de statistiques, est utile, d'un point de vu cognitif, à la reconnaissance. Dans le cas des textures, ces statistiques constituent même l'unique information disponible, le cerveau ayant fait fi de toute autre représentation plus détaillée (Nelken and Cheveigné, 2013).

Dans une seconde expérience (McDermott et al., 2013), les sujets ont du reconnaître, parmi une triade de sons synthétisés, celui produit par une source différente (i. e.un type de texture différent, Cf. Figure 15). Les résultats ont montré que la capacité de discrimination

est fonction de la durée des textures. Plus cette dernière est élevée, plus la capacité à discriminer est importante. Ce constat valide les hypothèses formulées par (Saint-Arnaud, 1995) sur l'existence d'une période d'attention, nécessaire au cerveau afin de percevoir le stimuli comme une texture. Ces résultats ont aussi montré que le processus de traitement de l'information sonore comprend une prise de décision quant à la nature des stimuli, décision qui va ensuite influer sur la manière d'analyser l'information montante. L'expérience prouve que cette prise de décision n'a rien d'anodine, car, dans le cas où le cerveau perçoit une texture, il décide sciemment de dégrader l'information, en la résumant de manière statique.

Le fait qu'un jugement perceptif s'améliore avec la durée des stimuli est un principe bien connu en perception des sons(Moore, 1973). Une troisième expérience de (McDermott et al., 2013) a montré cependant que cette vérité n'était pas toujours vérifiée. Au cours de cette expérience, les sujets, soumis à trois exemplaires d'un même type de texture (e. g. trois sons synthétisés de pluie), dont deux étaient produits à partir des mêmes statistiques extraites, ont du identifier le troisième, issu de statistiques différentes (Figure 15). Les résultats ont montré que la capacité des sujets à discriminer le bon stimulus décroît avec la durée des stimuli. Ce fait, qui peut sembler paradoxal, est une conséquence directe du choix du cerveau de ne traiter les textures que sur la base de statistiques. Partant du principe que le signal sonore est analysé suivant des fenêtres d'intégrations successives (Poeppel, 2003; Yabe et al., 1998), plus les stimuli sont longs, plus le cerveau est confiant dans le fait qu'il a à faire à des textures, et plus il tend à conserver une information réduite. La réduction de cette information finit éventuellement par gommer les différences fines qui existent entre les stimuli, ce qui ne permet plus de faire la distinction entre eux.

"Le cerveau fait ci ou ça" : c'est une hypothèse. Et puis évitons d'anthropomorphiser le cerveau

Une des avancés majeures de ces études est qu'elles apportent de nouvelles réponses sur la nature des représentations sonores stockées en mémoire. Dans le cas des textures, il s'agirait ainsi de descripteurs bas-niveaux, résumés sous la forme de statistiques simples. Cette découverte fait sens d'un point de vue écologique, car elle respecte le principe d'économie de moyens. Le cerveau, reconnaissant que les caractéristiques des textures n'évoluent pas au cours du temps, ne conserve en mémoire qu'une information condensée, qui lui permet pourtant de traiter des sons potentiellement longs.

Il a été montré que le cerveau peut stocker bien plus que des statistiques. (Agus et al., 2010) a mis en évidence qu'un bruit blanc, écouté de manière répétée, pouvait être reconnu encore plusieurs semaines après l'écoute, et ce parmi d'autres bruits blancs. Dans ce cas le cerveau emmagasine bien la totalité du signal acoustique.

plus d'information, c'est tout ce qu'on peut dire -- pas forcément la totalité

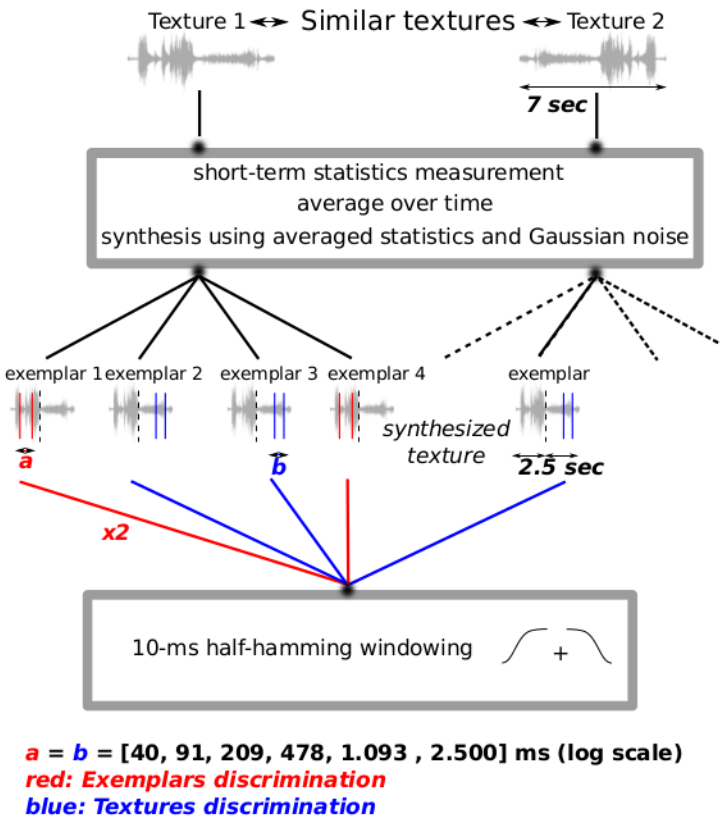


FIGURE 15 : Plannification expérientale de l'expérience de discrimination de textures sonores et d'exemplaires de textures sonores, menée par (McDermott et al., 2013)

3.5.3 *Période d'attention*

Nos travaux ont également porté sur cette notion de période d'attention. Comme nous l'avons vu, la texture est un objet composite. En poussant cette vision à l'extrême, une texture peut être vue comme un empilement d'événements sonores, ayant cessés d'être perçus de manière distincte, dès lors qu'ils forment un tout homogène et stable.

Nous avons suivi cette idée, afin de bâtir un protocole permettant d'analyser la période d'attention. En considérant comme stimuli une mixture d'événements du même type, nous faisons l'hypothèse qu'à partir du moment où le cerveau parvient à isoler un événement de cette mixture, il ne perçoit plus la mixture comme une texture, mais comme une succession d'événements. Inversement, s'il ne parvient pas à distinguer un événement isolé, alors la mixture est perçue comme une texture.

Euh, on est dans la biblio, là, non ? Pourquoi tu commences à parler de tes travaux ?

Sur la base de cette hypothèse, une expérience de reconnaissance de type oui/non (Cf. Annexe C pour une description exhaustive de l'expérience) a été montée. Dans cette expérience, chaque stimuli est composé d'un son cible, suivi d'une séquence d'événements enchevêtrés. Tous les événements sont des sons isolés ayant une durée de

(si c'est un travail que tu as fait mais qui est d'une certaine manière déconnecté de ce qui est présenté dans la thèse, il faut introduire clairement comme tel -- là on a l'impression de rentrer de but en blanc dans la description de ton travail...)

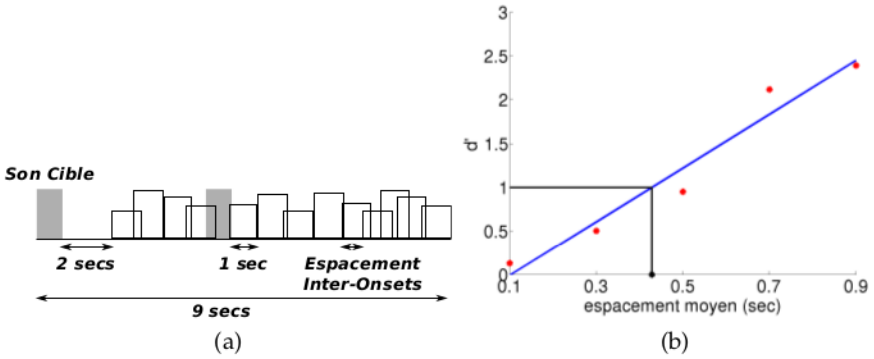


FIGURE 16 : Événement ou texture sonore : influence de la période d'attention. (a) la nature des stimuli utilisés. (b) le seuil d'espacement moyen permettant de faire la distinction entre une séquence d'événements et une texture

1 seconde. La séquence dure 6 secondes (Cf. 16a). L'objectif pour le sujet est d'indiquer si oui ou non il a entendu le son cible dans la séquence d'événements.

Les séquences donnent à entendre des scènes de trafic. Ces scènes sont simulées en agglomérant des sons de voiture isolés. La simulation est contrôlée par un paramètre réglant l'espacement temporel inter-onset moyen entre les événements. Cinq valeurs d'espacement sont considérées : 0.1, 0.3, 0.5, 0.7 et 0.9 secondes. Pour chaque espacement, nous simulons 20 séquences de trafics, chaque sujet devant alors écouter 100 stimuli. La moitié de ces stimuli sont des pièges, le son cible y étant absent. Nous mesurons les performances des sujets en utilisant la mesure de sensitivité d' (Cf. Annexe C.1). Les résultats sont très encourageants. Ils montrent qu'il existe bien un espacement limite à partir duquel la mixture cesse d'être perçue comme une texture. Pour des sons isolés d'une seconde, cet espacement limite est de 0.42 secondes, soit la moitié de la durée des événements utilisés (Cf. Figure 16b).

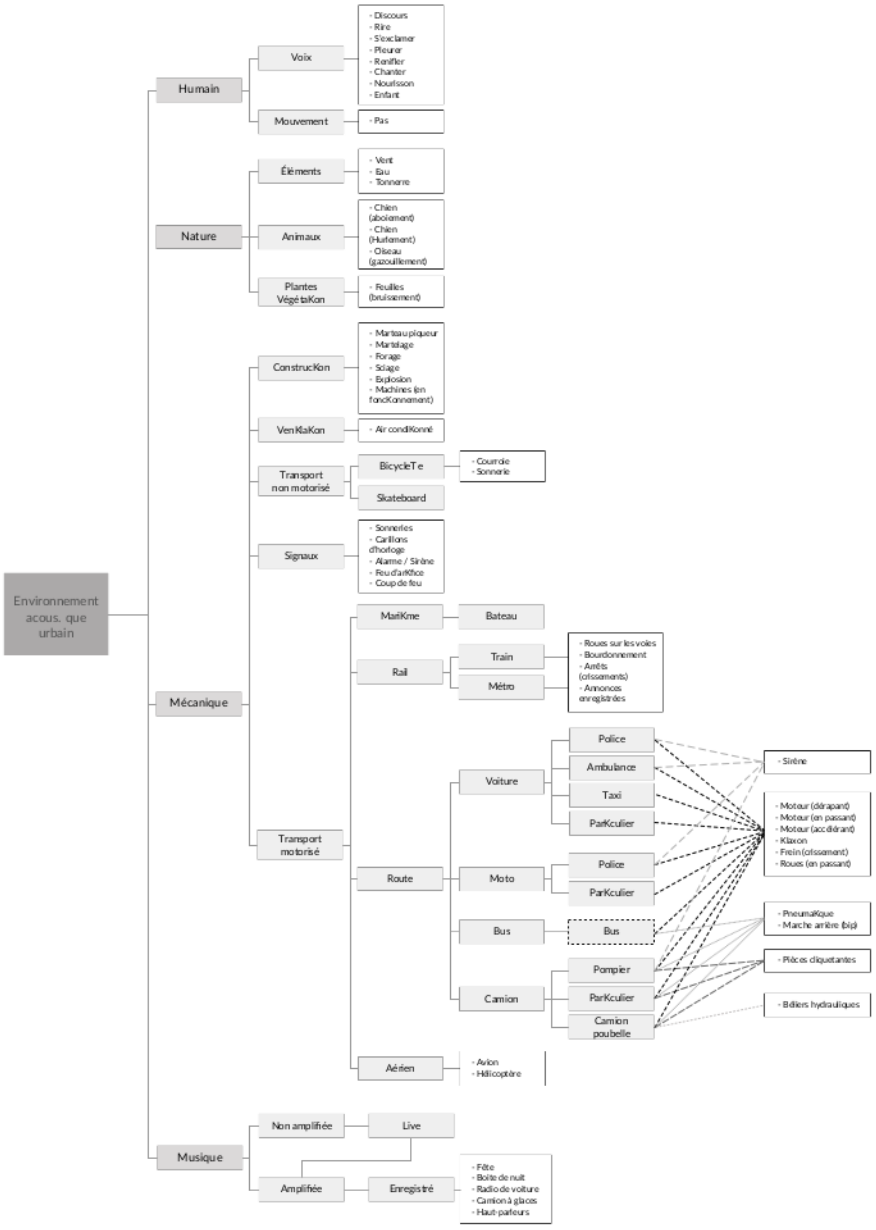


Figure 17 : Taxonomie des sources sonores urbaines, d'après (Salamon et al., 2014)

4

UN MODÈLE MORPHOLOGIQUE DE SCÈNES SONORES ENVIRONNEMENTALES

4.1 MOTIVATIONS

4.1.1 *Analyse sensorielle*

Comme nous l'avons vu à la section **??**, les approches expérimentales étudiant les paysages sonores sont de deux natures :

- Les approches catégorielles : où il s'agit de mettre en évidence des catégories de paysages sonores ou de sources sonores
- Les approches dimensionnelles : où il s'agit d'identifier les dimensions perceptives engagées dans les processus cognitifs, ainsi que les indicateurs dont elles dépendent.

Concrètement, ces approches suivent des objectifs différents :

- l'approche catégorielle cherche à identifier les objets d'intérêts de l'environnement sonore.
- l'approche dimensionnelle cherche à identifier les dimensions caractérisant les qualités affectives perçues d'une scène.

Nous pensons que la simulation offre un cadre expérimental élégant permettant de faire le lien entre les deux approches.

Si l'on interroge les influences qu'ont les différents éléments constituant une scène sur les qualités affectives perçues, alors les deux approches se rejoignent naturellement. L'approche catégorielle permet d'établir la liste des éléments d'intérêt, laquelle liste peut servir de base à une annotation des stimuli utilisés par l'approche dimensionnelle afin d'étudier les contributions spécifiques de leurs éléments respectifs.

D'une part, comme illustré à la figure 18, la simulation peut être vue comme l'approche inverse des épreuves catégorielles (cf. section **??**). Quand les épreuves catégorielles discrétisent l'environnement sonore sur la base d'un tri et/ou d'une description verbale réalisés par un sujet, la simulation, elle, recompose cet environnement, à partir d'une banque imposée de sons isolés. Ainsi, les catégories sonores, point de sortie des épreuves catégorielles, constituent-elles la banque de sons, point d'entrée de la simulation.

D'autre part, la finalité de la simulation est de produire un environnement complet, dont la partition, i.e. ses caractéristiques structurelles et compositionnelles, est parfaitement connue. Les stimuli

(au sens musical)
-- ça vaut le coup de préciser parce que tu parles aussi par endroits de clustering...

ainsi produits peuvent être utilisés par l'approche dimensionnelle, afin d'étudier de manière fine les contributions des différents éléments.

La simulation se pose donc comme un outil intermédiaire, faisant le lien entre les connaissances issues des études adoptant l'approche catégorielle, et les stimuli requis par l'approche dimensionnelle.

Passer par la simulation présente par ailleurs plusieurs intérêts :

- *Intérêt pratique* : Afin d'étudier l'importance relative des différentes sources, il est indispensable de disposer de stimuli dont la partition est connue. Une première solution, notamment adoptée par (Lavandier and Defréville, 2006), est alors d'annoter les stimuli. L'annotation est une solution cependant limitée :

  1. Cette opération est fastidieuse, très coûteuse en temps, et difficile à réaliser pour de grandes banques de données.
  2. Connaître la position des différentes sources dans une mixture sonore ne permet pas d'isoler leurs caractéristiques physiques respectives, et donc de calculer des indicateurs acoustiques dédiés. En traitement du signal, la séparation de sources reste un problème encore très étudié.

  GL : Passer par la simulation nous permet d'obtenir directement le stimuli et son annotation, cette dernière étant produite directement par le sujet, et non par un individu tiers.

- *Intérêt écologique* : La validité écologique des stimuli est un problème fondamental en analyse sensorielle. Dans le cas de l'analyse des qualités affectives perçues, où l'on demande directement au sujet " que pensez vous de la qualité Q de cet environnement ", il s'agit de garantir que les stimuli proposés fassent sens par rapport à la représentation mentale que le sujet se fait :

  – du monde sonore
  – de Q

  Il est possible, dans les approches classiques, de résoudre ces problèmes en étudiant de manière préalable les stimuli à enregistrer (cf. section 3.1.4). La simulation, en renversant la question posée (" générer un environnement qui correspond à une certaine valeur de Q "), garantit, quant à elle, la validité écologique des stimuli. Le sujet est directement responsable de la nature du stimuli. Ce dernier est ainsi directement connecté à la représentation sonore du sujet.

- *Représentativité des stimuli* : Toute étude sensorielle, quelle soit *in situ*, ou en laboratoire, doit sélectionner un nombre restreint d'environnements sonores à évaluer. Il s'agit alors, tant que faire

se peut

ce peu, de garantir que le substrat de stimuli proposé soit représentatif de l'ensemble des environnements étudiés, un déséquilibre dans l'élaboration dudit substrat pouvant affecter, in fine, l'évaluation des stimuli.

Dans le cas des études sur la perception des environnements urbains, il est d'usage d'isoler des zones d'intérêts (parc, rue, place, cf. section 3.1.4), et de répartir équitablement les stimuli parmi ces zones. Cependant, l'environnement d'une même zone est changeant, il s'agit alors également de contrôler la diversité structurelle des stimuli provenant d'une même zone, et ce particulièrement si l'on cherche à étudier l'influence spécifique des différentes sources. Indiscutablement, cette étape est ardue.

inutile (parlé)

Si la structure interne des paysages sonores est variable, la diversité des sources sonores qui les composent est plus contrôlable. Des environnements sonores de parcs et de rues peuvent comprendre des voix humaines, des bruits de pas, des sons de voiture etc. Seuls les caractéristiques physiques, ainsi que les patterns d'occurrences de ces sources, vont varier. Évaluer des scènes simulées, à partir d'une banque de sons isolés (sources sonores), peut se révéler une solution pratique au problème de diversité des stimuli. Prenons l'exemple de l'étude de l'agrément sonore dans l'environnement urbain. Dans un premier temps, les stimuli sont obtenus via une épreuve de simulation. Dans cette dernière, seule la qualité affective des stimuli est fixée (agréable/désagréable). Les sujets construisent alors les scènes directement en fonction de l'image qu'ils se font d'un environnement urbain agréable/désagréable, adaptant ainsi la structure de la scène à la qualité de l'environnement. Dans un deuxième temps, ces scènes peuvent alors servir de stimuli pour une analyse sémantique différentielle de l'agrément. Cette approche est notamment celle utilisée dans nos travaux (cf. chapitre 5).

Enfin, la plupart des environnements que nous percevons sont relativement neutres, et ne provoquent pas en nous de réactions particulières. Et il peut n'être pas évident d'évaluer des dimensions perceptives comme l'agrément, la gène ou le confort, de ces environnements. Des scènes simulées, sur la base d'une qualité affective imposée (e. g. agréable), proposent, quant à elles, des versions stéréotypées des environnements ainsi qualifiés. On peut voir dans ces scènes un "résumé cognitif", riche et condensé des environnements étudiés. Isoler les éléments d'intérêt de ces scènes peut s'avérer plus facile.

par ex.
(anglicisme inutile :-P)

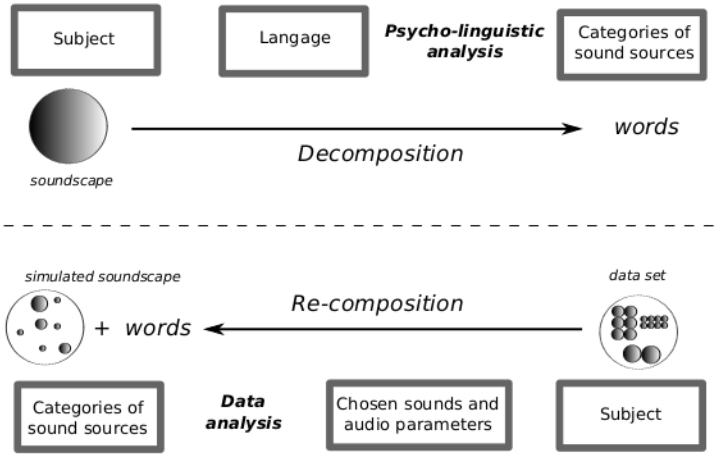


FIGURE 18 : TODO

4.1.2 *Analyse automatique*

4.2 PROPOSITION D'UN MODÈLE DE SCÈNES SONORES

4.2.1 *Discrétiser l'environnement sonore*

4.2.1.1 *L'unité de bases : la source sonore*

Les études portant su l'ASA, et plus spécifiquement sur les processus de ségrégation, montrent, d'une part, que l'humain fait sens de son environnement en isolant les informations relatives aux différentes sources qui composent la scène, d'autre part, que ce groupement intervient très tôt dans la chaîne de traitement, et se base sur des règles génériques innées.

GL : TODO : Neurosciene

De même, la recherche en matière de paysages sonores, adoptant l'approche catégorielle, met en évidence que les processus de catégorisation s'appuient également sur la composition sémantique des scènes, i. e. les sources sonores identifiées.

Il nous apparaît alors normal, pour notre modèle, de considérer comme élément de base la source sonore. Or, comme nous l'avons vu précédemment (cf. section refXX), la notion de source sonore est variable, un même objet pouvant être reconnu suivant plusieurs degrés d'abstraction.

Tu n'as pas encore explicitement parlé de "on va se monter une collection de sons, etc", du coup parler d'enregistrement là arrive un peu comme un cheveu sur la soupe.

4.2.1.2 *Typologie : niveaux d'abstractions et nomenclature source-action*

Avant d'acquérir ces sources, i. e. de les enregistrer, il est nécessaire de les identifier. Une démarche naïve consisterait alors à établir la liste exhaustive de toutes les classes de sources sonores composant l'environnement. Une telle approche soulèverait deux problèmes :

- Une source sonore peut se décrire en fonction de plusieurs niveaux d'abstraction. Comme nous l'avons vu (cf. section **??**), identifier et nommer sont des actions déterminées par notre représentation mentale du monde. Cette représentation s'organise, entre autre, suivant l'axe vertical des niveaux d'abstraction sur lequel s'appuie notre travail de catégorisation. Ainsi, si deux individus entendent un même son de voiture, il est possible que le premier le nomme "voiture" et le deuxième "moteur". Le dénombrement de l'ensemble des sources pouvant être utilisées par le modèle doit prendre en compte ce fait. Ces sources doivent être regroupées en classes hiérarchisées, afin de bâtir une structure taxonomique.

- Il n'existe pas de taxonomie standardisée des sources sonores. C'est une tradition des sciences modernes de classer et nommer les éléments avant de les étudier. Dans les domaines de la faune ou de la flore, une observation longue et minutieuse des sujets d'étude a permis d'élaborer un système de classification, et ainsi d'organiser et trier les objets en fonction de leurs propriétés partagées. Elle a permis encore d'élaborer une terminologie précise des classes d'objets. Grâce à quoi, la Biologie est devenue une science, laquelle science a donné naissance à la théorie de l'évolution (Lecointre and Le Guyader, 2006). Dans le domaine du son (ainsi que dans celui des odeurs), en revanche, point de système de classification (Dubois, 2000; Niessen et al., 2010). Nous trouvons à cela deux explications :

  on parle de taxonomie

  - *Champ lexical limité* : L'identification et la description d'un son sont des processus subjectifs étroitement liés au langage (cf. section **??**). Deux sujets appartenant à deux groupes sociaux différents n'utiliseront pas les mêmes mots pour décrire un même objet. Pour établir un système de classification, il faut prendre une décision quant à la définition des termes utilisés. Or, contrairement au domaine de la vision, où une terminologie de base pour décrire les objets (couleur, forme etc.) est globalement partagée, le champ lexical applicable aux phénomènes acoustiques est, d'une part, limité (durée, fréquence...) (Dubois, 2000), d'autre part, emprunté, dans une large mesure, à d'autres domaines perceptifs. On parle ainsi de brillance, ou de rugosité de sons. La diversité des termes descriptifs, et l'absence de consensus sur ce qu'ils désignent, rend difficile l'élaboration d'une classification standardisée.
  - *Influence du contexte* : L'identification et la description d'un son, section 3.2.6 sont dépendantes du contexte, i.e. de la nature des sources cooccurrentes dans la scène (Ballas and Howard, 1987; Gygi and Shafiro, 2011; Niessen et al., 2008).

Il apparaît clairement que les classes de sons peuplant notre environnement doivent être organisées autour d'une taxonomie : un système de classes hiérarchisées. Cependant il y a un choix à faire quant à la manière de regrouper les sons à l'intérieur de cette taxonomie.

Comme vu à la section **??**, plusieurs études ont montré que la catégorisation des sources sonores s'opère suivant des attributs sémantiques. Parmi ceux-ci, deux reviennent souvent :

- La source (agent, objet, fonction), i. e. l'objet émettant son.
- L'action, i. e. le mouvement physique à l'origine du son.

Ces deux attributs fonctionnent de pair. S'inspirant de l'organisation catégorielle verticale à trois niveaux de Rosch (cf. section 3.2.3.3), Guyot et al. (Guyot et al., 1997) propose un système de catégorisation où les auditeurs identifient des groupements de sources abstraits, au niveau superordonné (" Bruit généré par une excitation mécanique "), des actions au niveau de base (" gratter "," frotter ") et des sources au niveau subordonné (" vaisselle "," stylo "). Reprenant à son tour ce système, (Houix et al., 2012) montre que les sons sont catégorisés, en premier lieu, à partir du type de sources, et ensuite, seulement, à partir de l'action.

L'association source-action semble être une bonne base sur laquelle bâtir une taxonomie. Les classes haut-niveaux étant des classes abstraites de sources sonores (" véhicule "), les classes intermédiaires des classes de sources sonores (" voiture "), et les classes basses des actions sonores (" passage "). Pour les classes de bas niveau, la perméabilité intra-classe est minimale.

verbe

L'association source-action n'est cependant pas suffisante. Le choix des labels utilisés doit faire l'objet d'une sélection particulière. Ces label doivent être génériques, compréhensibles, et décrire de manière non ambiguë les objets de la classe considérée. Afin de les choisir, on peut se référer aux travaux de Gaver (Gaver, 1993b), qui propose une taxonomie phénoménologique des sons, on peut se référer aussi aux travaux de Niessen et *al.*(Niessen et al., 2010) qui, sur la base d'une étude bibliographique de près de 35 publications, établit la liste des catégories sonores les plus utilisées, on peut se référer enfin aux travaux de Salamon et *al.*(Salamon et al., 2014), qui, partant des travaux de (Brown et al., 2011) et reprenant l'association source-action, élabore une taxonomie de sons urbains.

oral

Afin de contourner le problème, on peut là encore s'appuyer sur des considérations perceptives afin d'établir, dans un contexte expérimental donné, quels sons requièrent d'être enregistrés séparément, et quels groupes de sons peuvent être enregistrés ensemble.

4.2.1.3 *événements, textures et scènes amorphes*

Nous avons montré que l'utilisation de la nomenclature basée sur l'association source-action nous permet de dénombrer et trier l'ensemble des sons présents dans l'environnement.

encore... il faudrait vraiment plus tôt dans le chapitre (fin d'intro, typiquement), expliquer ton plan d'action concret, annonçant SimScene, etc.

Il s'agit alors, sur la base de la taxonomie établie, d'enregistrer pour chacune des classes un nombre de sons suffisant. Considérant des environnements denses comme la ville ou la forêt, cette approche pose des problèmes pratiques de faisabilité.

Tout les sons n'ont pas le même intérêt. Une voix humaine peut facilement être isolée du reste des sons concurrents (Carlyon, 2004). De même, un fond sonore de trafic urbain est moins informatif que d'autres sons ponctuels et proches (Southworth, 1969).

"inversement", plutôt ?

Comme vu à la section 3.4.5.2, Maffiolo montre l'existence de deux processus cognitifs distincts dont l'activation dépend de la nature des environnements : l'analyse holistique, s'agissant de scènes amorphes,i. e. sans événement apparent, l'analyse descriptive, sur la base d'une information sémantique extraite à partir des événements connus, s'agissant de scènes événementielles, i. e. comportant des événements identifiables.

et

Ces résultats nous amènent à penser que les processus de ségrégation dépendent également de la nature structurelle de l'environnement. Lorsque des événements émergent d'environnement sonore, le cerveau traite l'information des différentes sources de manière séparée. Plusieurs flux auditifs sont ainsi générés i. e. un pour chaque séquence d'événements émis par la même source. A l'inverse, quand le cerveau ne parvient pas à isoler d'événement, la scène est traitée globalement, i. e. tous ces éléments étant agglomérés dans un même flux.

"un" ou pluriel

"sont", ou pas de "i.e."

De même, les différents travaux sur les textures sonores (cf. section 3.5), montrent que le cerveau a tendance à résumer l'information extraite, lorsqu'il détecte qu'une séquence n'est composée que d'un mélange de sons similaires, et n'apporte pas d'information au cours du temps.

Outre les événements sonores, deux autres types de sons semblent pouvoir être isolés :

Une scène amorphe = un son ? Je dis pas que c'est faux, mais c'est surprenant, ça mérite d'etre un peu plus "emballé"

- *Scènes amorphes* : un son contenant une faible information sémantique,
- Textures sonores : un son dont les caractéristiques physiques restent stables au cours du temps, analysé à partir de statistiques extraites d'une représentation temps-fréquence" puisque explicité plus bas.

définition: tout singulier, ou tout pluriel :-P

Nous pensons qu'il existe des connexions entre les notions de textures/événements, et celles de scènes amorphes/événementielles. Les séquences événementielles peuvent être vues comme des séquences

composées soit uniquement d'événements, soit d'événements et de textures, la présence d'événements, porteurs d'une information plus riche, primant quant au choix du processus à mettre en œuvre.

Les textures et les scènes amorphes sont traitées, elles, de manière holistique, à partir de propriétés acoustiques globales, pour les scènes amorphes (Dubois et al., 2006; Maffiolo, 1999), et sur la base d'une information résumée statistiquement pour les textures(McDermott et al., 2013). Toutes deux portent une information limitée (Nelken and Cheveigné, 2013; Saint-Arnaud, 1995). Cependant, les séquences amorphes sont spontanément décrites par des sujets comme des " fonds sonores " (Guastavino, 2006; Maffiolo, 1999), induisant qu'elles n'existent que grâce à un processus de construction de flux auditifs, alors que les textures sont des objets seulement définis sur la base de leur nature physique. Un exemple de texture souvent cité est le sons du " galop ", qui selon le contexte peut se retrouver au premier plan de la scène.

impliquant ?

Considérant cela, nous pensons qu'il est possible d'assimiler une scène amorphe à une texture, ses caractéristiques physiques demeurant stables au cours du temps, nombre de scènes amorphes (" brouhaha de rue ", " brouhaha de trafic ") étant, par ailleurs, citées comme textures". Cependant l'inverse, considérer une texture comme une scène amorphe, n'est pas forcément vrai.

Afin de limiter le nombre d'enregistrements nécessaires, il est donc possible d'enregistrer directement des mixtures de sons, à condition que ces dernières puissent être considérées comme des textures, la définition de cette dernière notion englobant les scènes amorphes.

4.2.2 *Description du modèle morphologique*

4.2.2.1 *Classe et collection de samples*

Dans le modèle proposé, la scène sonore est vue comme une somme de sources sonores, ou autrement dit, " un squelette d'événements sur un lit de textures " (Nelken and Cheveigné, 2013).

D'un point de vue pratique, ces éléments sonores sont enregistrés. Ces éléments sont nommés des samples.

Définition 1 *Un sample est un enregistrement d'un son isolé, qu'il s'agisse d'un événement, ou d'une texture.*

Ces samples, regroupés en classes de sons hiérarchisées, forment une taxonomie. Un exemple est donné figure 19. Les niveaux hiérarchiques de la taxonomie sont appelés niveaux d'abstraction. Les classes ayant un niveau d'abstraction élevé constituent un regroupement conceptuel de samples ayant potentiellement des caractéristiques variées (i. e. Humain). Plus le niveau de la classe est bas, plus le groupement induit est précis, regroupant des samples similaires (i. e. *voix-adulte-cri*).

plutôt "par ex."

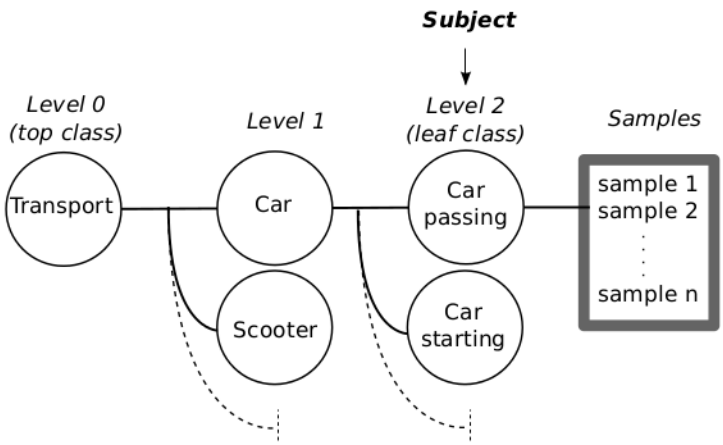


FIGURE 19 : Organisation hiérarchique de la banque de sons isolés utilisée pour la simulation

Pourquoi "samples" plutôt qu' "échantillons" ?

Définition 2 *Une classe est une collection de samples, jugés équivalents. Si le niveau d'abstraction d'une classe est tel que cette dernière possède des sous-classes, alors sa collection de sample est la somme des collections respectives de chacune des sous classes.*

Les classes de niveau d'abstraction élevé sont nommées uniquement à l'aide de termes abstraits désignant, de manière globale, les samples qu'elles regroupent (e. g. *transport*). Les classes de bas niveau utilisent la nomenclature source-action (e. g. *voiture passe*). Quant aux classes du dernier niveau, elles correspondent à des collections de samples, par définition, équivalents les uns aux autres.

4.2.2.2 *Séquences de samples*

Chaque classe de sons sélectionnée pour faire partie de la scène simulée est liée à une piste. Cette piste est une séquence temporelle où sont positionnés les différents samples. Elle est la contrepartie simulée du flux auditif.

Définition 3 *Une piste est une séquence temporelle composée de sample appartenant à une même classe de sons.*

La construction de la taxonomie (nombre de classes, nombre de niveaux d'abstraction), dépend, évidemment, de la tâche considérée.

L'ensemble des pistes, ainsi que leurs paramètres, forment ce que nous appelons une partition.

Définition 4 *La partition désigne l'ensemble des propriétés des pistes composant une scène, à savoir, les classes de sons liées aux pistes, et leurs paramètres structurels (niveau, espacement, début et fin, cf. section 4.2.2.3).*

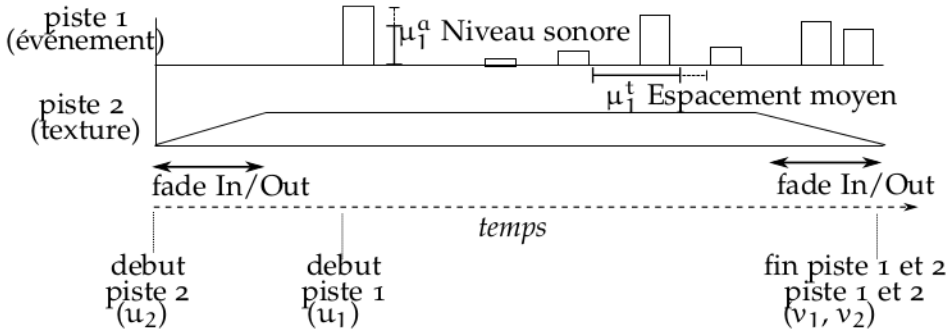


FIGURE 20 : TODO

4.2.2.3 *Paramètres*

En suivant la terminologie précédemment introduite, une scène sonore est vue comme une somme de pistes. Chaque piste étant une séquence temporelle, dont la structure dépend d'une série de paramètres (cf. figure 20). Le modèle ne propose pas d'interagir avec un sample en particulier, mais toujours avec une séquence de samples.

verbe

Nous isolons trois attributs permettant de contrôler un piste :

- *Le niveau* : la moyenne/variance des niveaux des samples
- *L'espacement* : la moyenne/variance des espacements inter-onsets entre les samples
- *La durée* : le début et la fin de la piste

Le modèle fait une distinction explicite entre la gestion des pistes d'événements et la gestion des pistes de textures. En effet, la notion de texture ne peut se comprendre que pour un son continu. Une piste de texture est donc composée de samples concaténés les uns aux autres sans espacement (cf. figure 20).

plus que ça (fondu)

Pour qu'une piste de texture soit "plausible", i. e. qu'on ne détecte pas de discontinuité flagrante, la piste doit être une séquence composée de samples provenant de la même source, et obtenus avec un matériel (et des réglages) identique(s).

4.2.2.4 *Formalisation du modèle*

Tout au long de nos travaux, nous avons utilisé plusieurs modèles, ainsi que différents paramètres afin de simuler des scènes sonores. Nous formalisons, dans cette partie, une version générale du modèle proposé. Les diverses modifications appliquées au modèle, en fonction de son utilisation, sont indiquées dans les sections 4.3 et 4.4.

La formalisation présentée vaut uniquement pour les classes d'événements sonores. Nous décrivons par la suite les diverses contraintes qui s'appliquent pour une classe de textures.

En considérant s, une scène composée de C classes de sons, le modèle de s se définit comme suit :

Conventionnellement, les majuscules \mathcal sont plutôt utilisées pour des ensembles / collections, pas pour des valeurs numériques

$$s(n) = \sum_{i=1}^{C} p_i(n) \tag{1}$$

avec n un indice temporel discret, et p_i la piste correspondant à la classe c_i. La classe c_i est composée de $|c_i|$ samples $c_{i,m}$, $1 < m < |c_i|$.

Une piste p_i est définie comme une séquence de n_i samples d'événements $e_i^k(n)$ ($k = (1, 2, \ldots, n_i)$), choisis aléatoirement parmi les $|c_i|$ samples de la classe c_i. Considérant $\mathcal{U}(x, y)$, une distribution uniforme d'entiers allant de x à y ($x < y$), on a alors :

$$e_i^k = c_{i,\mathcal{U}(1,|c_i|)} \tag{2}$$

Pour chaque piste p_i, un facteur d'amplitude est tiré aléatoirement à partir d'une distribution normale de moyenne μ_i^a et de variance σ_i^a. De même, les espacements inter-onsets sont tirés d'une distribution normale de moyenne μ_i^t et de variance σ_i^t. Les indices temporels de début et de fin de chaque piste sont notés u_i et v_i respectivement. Formellement, une piste p_i se définit comme suit :

$$p_i(n) = \sum_{j=1}^{n_i} \mathcal{N}(\mu_i^a, \sigma_i^a) c_{i,\mathcal{U}(1,|c_i|)}(n - n_i^j) \tag{3}$$

$$n_i^j = n_i^{j-1} + \mathcal{N}(\mu_i^t, \sigma_i^t) \tag{4}$$

où $n_i^0 = u_i$ par convention. Le signal d'une piste est défini de telle sorte que $p_i(n) = 0$ si $n > v_i$. Les paramètres du modèle sont, μ_i^a, σ_i^a, μ_i^t, σ_i^t, u_i et v_i, et doivent être réglés pour chaque piste p_i. La figure 20 offre une illustration de l'action des paramètres introduits.

Pour les textures, deux distinctions sont à observer avec le modèle défini précédemment :

1. L'amplitude du signal (μ_i^a, σ_i^a) n'est tirée qu'une seule fois, et la valeur est appliquée à tous les samples.

2. Afin d'éviter toute sentiment de discontinuité, deux samples de texture sont concaténés en considérant un recouvrement fixé. Ce recouvrement est choisi afin de pouvoir appliquer un fondu enchaîné (*cross-fade*) à valeur d'énergie constante entre les samples, et ce afin de donner l'illusion de continuité.

sensation

"avec un recouvrement, sur lequel est appliqué un..."

oral :-P

GL : a revoir avec mathieu et mathias

4.3 DU MODÈLE À LA SIMULATION : L'ANALYSE SENSORIELLE

Dans cette section nous présentons une version du modèle précédemment introduit, afin qu'il puisse servir de base à un outils de simulation, nommé *SimScene*, utilisable dans le cadre de l'analyse sensorielle des scènes sonores.

Nous commençons par présenter différents outils déjà existant permettant de simuler des environnements sonores. Nous proposons, par la suite, un protocole expérimental, décrivant le cadre applicatif des épreuves perceptives basées sur la simulation. Enfin, nous relions ce protocole au modèle de scènes sonores (cf. section 4.2.2.3), et présentons les fonctionnalités de l'outil *SimScene*.

4.3.1 *Simulation et analyse sensorielle*

Plusieurs outils de simulation de scènes sonores ont déjà été proposés(Finney and Janer, 2010; Misra et al., 2006, 2007; Schirosa et al., 2010; Valle et al., 2009). Ils ont souvent pour but de générer automatiquement l'ambiance sonore d'un environnement virtuel. (Finney and Janer, 2010; Valle et al., 2009). Ils peuvent être vus comme des systèmes semi-autonomes : la simulation pouvant être contrôlée par un utilisateur, mais dépendant aussi soit d'un environnement visuel à illustrer, soit d'un environnement sonore à reproduire. D'autre outils, entièrement contrôlés par un utilisateur, servent eux d'aide à la composition(Misra et al., 2006, 2007). Ces systèmes s'éloignent tous sensiblement du cadre expérimental de l'analyse sensorielle.

A notre connaissance, seuls Bruce et *al.*(Bruce and Davies, 2014; Bruce et al., 2009) se sont servis de la simulation afin d'étudier la perception des paysages sonores. Ils proposent ainsi un système permettant à un sujet d'agir sur un environnement en ajoutant ou de supprimant des sources sonores spécifiques. Le sujet peut par ailleurs modifier le niveau sonore des sources, et leurs positions spatiales.

A l'aide de cet outil, les auteurs demandent à des sujets de manipuler des sources afin de recréer un environnement urbain. Les résultats montrent que l'inclusion ou l'exclusion des sources dépend plus de considérations sociales/sémantiques, que des caractéristiques physiques des sources. Ils soulignent néanmoins que le manque d'enregistrements disponibles limite l'analyse. Ils suggèrent par ailleurs de regrouper les enregistrements similaires en "groupes sémantiques" afin de faciliter l'analyse.

GL : TODO G1 : compléter Bruce et *al.*, cf wac
GL : TODO G2 : (Davies et al., 2014) montre que quand on demande à des participants de simuler un paysage sonore, les simulations font références à ce que les participants s'imaginent être un environne-

ment typique, sans tenir compte de leur propre préférence pour des sons particuliers.

4.3.2 *Protocole expérimental basé sur la simulation*

4.3.2.1 *Organisation des sons isolées*

L'objectif de l'expérimentation est de permettre à un sujet de simuler un environnement sonore cible, à partir de sons isolés. La banque de sons suit l'organisation décrite à la section 4.2.2.1. Les éléments sont regroupés en classes hiérarchisées, afin de former une taxonomie. Plus le niveaux d'abstraction de la classe est élevé, plus la variabilité des enregistrements appartenant à la classe est importante (cf. Figure 19).

Nous conservons la distinction observée entre les événements et les textures en créant deux taxonomies (i. e. deux banques de sons) séparées.

4.3.2.2 *Sélection des sons isolées*

L'objectif de la simulation est d'obtenir une image sonore de la représentation mentale que ce fait un sujet d'un environnement donné. Afin que cette image soit la plus "juste" possible, il faut que le protocole limite les biais pouvant influer sur les choix du sujets.

Un de ces biais intervient dans le processus de sélection. La grande majorité des outils permettant de parcourir une banque de données propose une recherche textuelle sur la base de mots clefs. L'efficacité de ce principe repose avant tout sur la structure typologique et la nomenclature de la base de données. Dans le cadre d'une expérience sensorielle visant à objectiver une représentation interne d'un sujet, cette approche pose trois problèmes majeurs :

- Les sons ne peuvent être tagués d'une manière satisfaisante. En effet, sémantiquement, un son peut être décrit de plusieurs façons. Nous pouvons en désigner la source (une portière de voiture), comme nous pouvons désigner l'action de cette source (le claquement d'une portière de voiture) ou encore son environnement (le claquement d'une portière de voiture dans un garage). Concevoir un système de recherche par mots clefs efficace suppose une description précise de chaque son, qui plus est adaptable à la représentation que s'en fait chaque sujet. Ce qui est difficilement envisageable pour nous. verbe

- Lors d'une recherche par mots clefs, le sujet doit objectiver un nom décrivant l'objet recherché. Or cette objectivation dépend des connaissances collectives du sujet, connaissances liées à sa sphère socioculturelle, et en particulier à sa langue. L'expérience

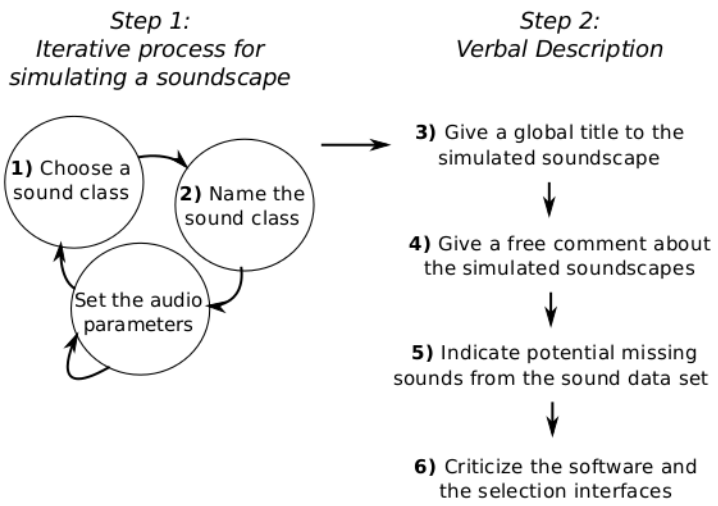


FIGURE 21 : Etape de processus de simulation pour l'analyse sensorielle

visant une diffusion internationale, cette contrainte pose problème.

- La description verbale du son, si elle est accessible par le sujet, peut potentiellement influencer sa sélection. Nous voulons éviter les situations biaisées où, par exemple, pour construire une scène environnementale " calme ", le sujet sélectionne a priori les sons référencés sous le vocable *parc*.

Nous pensons qu'imposer au sujet une terminologie exprimée à travers les labels décrivant les classes est un risque. Notre avis est que la sélection doit s'éloigner le plus possible d'un ancrage sémantique, et s'effectuer à l'aveugle, i. e.sur la base uniquement de l'écoute. Une interface développée spécialement dans ce but est présentée à la section 4.3.4.

Enfin, il est important de noter que le sujet ne peut accéder qu'aux classes du niveau d'abstraction le plus bas, i. e. qui ne possèdent pas de sous-classes et sont directement liées à une collection de samples.

i.e. inutile

L'organisation hiérarchique sert alors deux buts :

- faciliter le parcours, par les sujets, des banques de sons isolées (cf. section 4.3.4).
- faciliter le travail d'analyse de l'expérimentateur, en lui permettant d'observer la composition en terme de sources sonores des scènes suivant différents niveaux d'abstraction.

4.3.2.3 *Processus de simulation*

Trois étapes composent le processus de simulation (cf. Figure 21) :

- *Sélection* d'une classe de sons. Une fois une classe sélectionnée, une piste est générée.

- *Identification* de la classe de sons sélectionnée. Le sujet nomme la classe de sons qu'il a sélectionnée.
- *Paramétrisation* de la piste liée à la classe de sons. Le sujet fixe les paramètres de la piste (pour plus de détails sur les paramètres proposés cf. section 4.3.3).

Ces étapes peuvent être répétées, et ce dans n'importe quel ordre, le sujet pouvant agir rétrospectivement sur les pistes déjà crées. A la fin de la simulation, et afin d'accumuler le maximum de connaissances possible sur la scène simulée, le sujet peut :

- nommer l'environnement simulé
- fournir un commentaire libre décrivant son processus de création, ainsi que le paysage sonore qu'il a voulu illustrer

4.3.3 *Paramètres de contrôle*

Les paramètres permettent au sujet de contrôler la structure de chaque piste. Ils agissent sur tous les samples à la fois, et non sur un en particulier.

Parmi ces paramètres, on retrouve les trois précédemment introduits pour le modèle initial de scène sonore (cf. section 4.2.2.3 et 4.2.2.4), à savoir :

- *niveau sonore* (dB) : pour chaque sample, les niveaux sont tirés aléatoirement à partir d'une distribution normale, paramétrée par le sujet en terme de moyenne variance
- *espacement inter-onset* (seconde) : (piste d'événements seulement) comme pour les niveaux, les espacements sont tirés aléatoirement à partir d'une distribution normale, paramétrée par le sujet en terme de moyenne variance
- *début et fin* (sec) : le sujet fixe le début et la fin de chaque piste.

Afin de faciliter la simulation, deux paramètres supplémentaires sont proposés :

- *fondu par événement* (seconde) : (piste d'événements seulement) le sujet fixe une durée de fondu (entrée et sortie), appliquée à chaque sample d'une piste d'événements.
- *fondu global* (seconde) : le sujet fixe les durées de fondu séparément pour l'entrée et la sortie de la piste. Ces fondus s'appliquent ainsi à l'ensemble des samples de la piste.

Deux de ces paramètres ne s'appliquent que pour les pistes d'événements (*fondu par événement* et *espacement inter-onset*), les samples des textures étant séquencés sans espacement (cf. section 4.2.2.3)

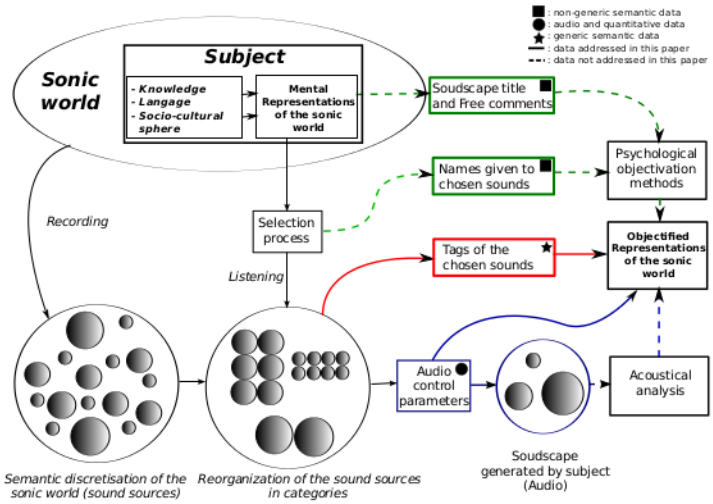


FIGURE 22 : TODO

4.3.3.1 *Données produites par le processus de simulation*

Ce protocole de simulation peut potentiellement produire un grand nombre de données. Ces dernières sont décrites à la figure 22. Nous les résumons dans La liste suivante :

- Données sémantiques objectives : La banque de données nous permet d'obtenir une information objective quant aux sources sonores présentes dans la scène. Les données sémantiques objectives sont les labels des classes sélectionnées.
- Données sémantiques subjectives : Il s'agit des noms donnés par le sujet 1) à la scène simulée, 2) aux classes de sons sélectionnées.
- Données quantitatives issues de la partition : Il s'agit de toutes les données relatives à la partition,i. e. pour chaque piste, le positionnement des samples et les paramètres (cf. section 4.2.2.2).
- Données quantitatives issues du signal : Il s'agit d'indicateurs acoustiques extraits du signal,e. g. le niveau sonore globale. Comme nous possédons les samples isolés utilisés pour la synthèse, il est possible de calculer ces descripteurs pour une classe, ou un ensemble de classes en particulier.

Le protocole nous permet de caractériser avec précision une scène simulée, sur la base de données sémantiques, subjectives ou objectives, ainsi que quantitatives. Considérant l'ensemble des données générées, les potentiels d'analyse sont vastes.

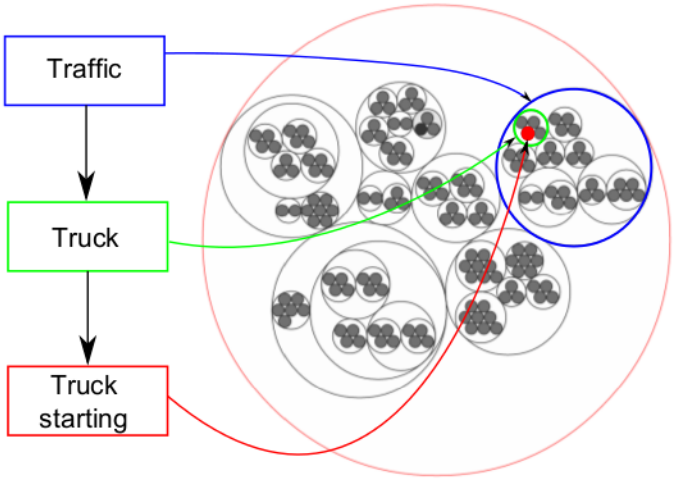


FIGURE 23 : L'interface de sélection aveugle de l'outil de simulation *Simscene*

4.3.4 *Interface de sélection aveugle des sons isolés*

Pour limiter l'influence de l'interface sur le sujet, il nous paraît nécessaire de libérer sa recherche de toute information textuelle. Nous proposons à l'utilisateur une interface graphique lui permettant d'explorer la banque de sons exclusivement à partir de l'écoute.

Visuellement, les classes du dernier niveau (les seules accessibles par le sujet) sont représentées par des cercles, et positionnées sur un plan. La disposition des cercles dans l'espace dépend de l'organisation hiérarchique de la base de données : les sous classes appartenant à la même classe sont proches les unes des autres, et ainsi de suite jusqu'à atteindre les classes des niveaux d'abstraction élevés.

La figure 23 présente l'interface pour la banque de données d'événements sonores. Cette organisation visuelle à été pensée afin de :

1. faciliter le parcours de la banque de données, les sons similaires (au sens des classes) étant proches les uns des autres. L'organisation hiérarchique se fonde, en effet, sur des principes cognitifs. Les classes ont été établies à partir de la littérature traitant des catégories de sources sonores (cf. section 4.2.2.1).

2. permettre aux sujets de rapidement appréhender toute l'étendue de la banque de données, i. e. l'ensemble des sons disponibles.

Chaque classe possède un son prototype. Ces sons ont été choisis par les expérimentateurs. Lorsqu'on " clique " sur un cercle, le prototype associé à la classe est joué. Le sujet parcours la banque de sons en cliquant sur les cercles.

Cette interface a fait l'objet d'une étude approfondie dont les résultats sont publiés dans (Lafay et al., 2016).

FIGURE 24 : L'outil de simulation *Simscene*

agrandir !

GL : TODO : Plus sur les résultats ?

4.3.5 *Interface de simulation : l'outil* Simscene

Simscene est un environnement de travail audio-numérique, supporté par navigateur internet, et développé dans le cadre du projet HOULE[1]. Il permet de simuler des paysages sonores à partir d'un corpus de sons. Il est prévu pour fonctionner via les navigateurs internet *Chrome* et *Firefox*. L'outil a été développé en javascript à l'aide des librairies *web-audio* [2] et *angular.js* [3]. L'interface de sélection (cf. section 4.3.4) a été développée à l'aide de la librairie *D3.js* (Bostock et al., 2011).

"library" -> "bibliothèque"

Web Audio est une technologie, pas une bibliothèque : c'est un standard qui fait partie intégrante de JavaScript

Simscene et ses fonctionnalités sont présentés en détail dans (Rossignol et al., 2015). Nous résumons ici ses fonctionnalités.

Le fonctionnement de *Simscene* se rapproche de celui d'un séquenceur audio. Chaque utilisateur choisit une classe de sons via l'interface de sélection (cf. section 4.3.4). Une fois la classe de sons sélectionnée, une piste audio, liée à cette classe, est créée. L'utilisateur peut alors modifier certaines propriétés de la piste via un groupe de paramètres de contrôle propre à chacune (cf. section 4.3.3). Des environnements de texte sont prévus afin de permettre à l'utilisateur de 1) nommer chaque piste, 2) donner un titre à la scène simulée et 3) commenter la scène simulée.

champs

L'interface propose un rendu graphique schématisé de la scène en cours de création (Cf. Figure 24). La piste est représentée par une bande possédant un axe temporel. Sur cette bande, chaque sample est représenté par un rectangle. L'espacement entre les rectangles est relatif à l'espacement entre les samples. De même, la hauteur des rectangles est proportionnelle au niveau sonore des samples. Dans le cas d'une piste de texture, un unique rectangle apparaît sur toute la longueur de la

1 Pour plus d'informations sur le Projet HOULE voir http://houle.ircam.fr/wordpress/

2 *web-audio* : Cf. http://www.w3.org/TR/webaudio/

3 *angular.js* : Cf. https://angularjs.org/

piste, un son de texture ne pouvant être entrecoupé de silences. Les caractéristiques des rectangles évoluent en fonction des changements de paramètres de la piste.

L'utilisateur a la possibilité, à tout moment, d'écouter la scène simulée.

L'outil de simulation est accessible via le lien suivant : http://www.irccyn.ec-nantes.fr/~lagrange/demonstrations/simScene.html

4.4 DU MODÈLE À LA SIMULATION : L'ANALYSE AUTOMATIQUE

5

APPLICATION DU MODÈLE MORPHOLOGIQUE À L'ÉTUDE PERCEPTIVE DES ENVIRONNEMENTS SONORES URBAINS

5.1 INTRODUCTION

Comme nous l'avons vu (Cf. section **??**), la recherche sur les paysages sonores a besoin d'outils permettant d'analyser les influences séparées des différentes sources sur les qualités affectives de l'environnement. La simulation offre des possibilités intéressantes, car elle nous permet d'obtenir des scènes sonores dont nous connaissons tous les paramètres structuraux, en particulier les caractéristiques distinctes des différentes sources.

Afin de montrer les potentialités inhérentes à l'utilisation de scènes simulées en analyse sensorielle, nous choisissons, comme cadre applicatif, le problème de l'agrément perçu dans les environnements sonores urbains.

Cette section présente les résultats d'une série de quatre expériences visant, chacune, à comprendre comment les différentes sources sonores qui composent une scène influent sur la perception de agrément. Toutes ses expériences s'appuient sur la simulation. La première est l'expérience de simulation à proprement parler i. e. où les sujets doivent créer les environnements. Les autres sont des épreuves de notation ou de tri classiques, utilisant les scènes simulées comme stimuli :

1. *Expérience de simulation* : Au cours de cette expérience, chaque sujet doit simuler deux environnements sonores urbains, le premier idéal/agréable et le deuxième non-idéale/désagréable, en utilisant l'outil et le protocole de simulation décrit à la section **??**.

2. *Évaluation de l'agrément* : Les sujets doivent évaluer l'agrément des scènes simulées à partir d'une échelle sémantique.

3. *Évaluation de l'agrément après modification des scènes* : Comme pour l'expérience précédente, les sujets doivent évaluer l'agrément des scènes simulées à partir d'une échelle sémantique. Cependant les scènes ont été modifiées, i. e. privées de certaines classes de sons identifiées comme ayant un impact sur l'agrément perçu.

4. *Catégorisation libre* : Les sujets doivent catégoriser les scènes sonores simulées. Au delà du problème initial de l'agrément, cette

pas très clair... tu veux dire "considérant une scène comme une source" ? Un peu surprenant.

dernière expérience nous amène à considérer logiquement l'influence de le la composition des scènes, comme sources sonores, sur les jugements de similarité.

Les expériences 1 et 2 vont de pair, et sont toutes deux décrites dans la section 5.2. Les expériences 3 et 4 sont elles décrites respectivement dans les sections 5.3 et 5.4

Tout au long de la présentation de ces expériences et de leurs résultats, nous poursuivrons deux objectifs :

- *Un objectif méthodologique* : Montrer les possibilités offertes par l'utilisation de scènes simulées dont nous connaissons précisément la partition **??** dans le cadre d'études sensorielles ayant trait à la perception des sons.
- *Un objectif applicatif* : Étudier quels sont les éléments qui participent à la perception de l'agrément dans les environnements sonores urbains.

5.2 L'IMPACT DE LA COMPOSITION DES SCÈNES SUR LA PERCEPTION DE L'AGRÉMENT

5.2.1 *Objectif*

L'objectif est d'étudier les influences spécifiques des différentes sources sonores qui composent les environnements urbains, sur la perception de l'agrément, en utilisant la simulation. Pour ce faire, nous planifions nos deux premières expériences (Cf. Figure 25) :

"comme précisé auparavant" -> un "nous" de moins :-P

- *L'expérience de simulation* :Nous l'avons dit, au cours de cette expérience, les sujets simulent les environnements qui serviront de stimuli pour les étapes suivantes. Chaque sujet compose deux scènes opposées, les premières répondant à la description idéales/agréables (i-scènes), les secondes répondant à la description non idéales/désagréables. Cette épreuve de simulation a fait l'objet d'une expérience pilote (Lafay, 2013; Lafay et al., 2014).
- *L'expérience d'évaluation* : A l'issue de la simulation, nous n'avons, de fait, qu'une connaissance binaire des propriétés affectives des scènes simulées : idéale (i) et non-idéale (ni). Cette seconde étape a pour but d'affiner notre connaissance sur l'agrément de chacune des scènes. Pour ce faire, nous demandons à un deuxième groupe de sujets d'évaluer, sur la base une échelle sémantique, l'agrément de chacune des scènes simulées. L'expérience d'évaluation sert deux buts :

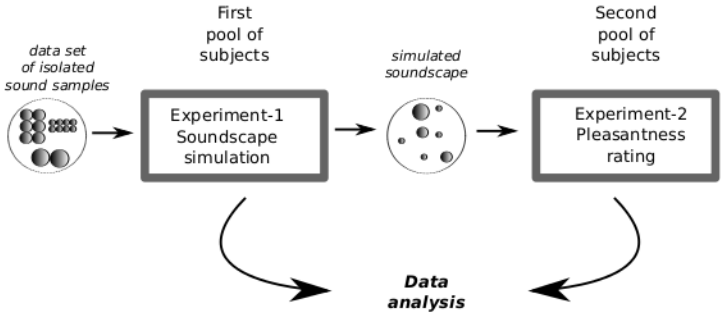


FIGURE 25 : Planification expérimentale des éxperiences de simulation et d'évaluation de l'agrément

1. Évaluer l'influence des différentes sources sur l'agrément perçu, de manière séparée, pour chaque type d'environnement (i ou ni).
2. Détecter la présence de cas extrêmes ou ambigus (*outlier*) dans les scènes simulées. Pour le reste de notre étude, GL : les qualités hédonique imposée (i et ni) servent de référence, de vérité terrain. Il nous faut donc garantir qu'il n'y ait pas d'ambiguïté entre les cas extrêmes des i- et ni-scènes, i. e. que la note d'agrément la plus basse des i-scènes reste supérieure à la note la plus haute des ni-scènes.

Notre analyse s'appuie sur les données produites par les deux expériences.

5.2.2 *Banque de données de sons isolés*

Dans cette section, nous présentons le processus de sélection et d'acquisition des sons utilisés comme matériau de base lors de la simulation des environnements sonores urbains. La banque de données est identique à celle utilisée dans le cadre de l'expérience pilote (Lafay, 2013; Lafay et al., 2014).

Pour plus de détails sur l'organisation interne de la banque de données, ainsi que sur l'interface graphique permettant de sélectionner ces dernières, se référer à la section **??**.

5.2.3 *Typologie des sources sonores présentes dans l'environnement urbain*

Afin de créer un corpus de sons isolés de référence, pour la simulation, nous avons réalisé une typologie des sons environnementaux urbains.

Pour ce faire, une étude bibliographique est effectuée afin d'identifier les sources et ambiances sonores les plus souvent citées dans la littérature. Cette étude porte sur 16 articles ou thèses. Chacun d'eux

traite de la manière dont nous discriminons les paysages sonores urbains. Il ressort que plusieurs approches sont possibles :

- 9 articles abordent le problème par une approche perceptive, soit en identifiant ou répertoriant des catégories de sources sonores, soit en étudiant l'impact de classes de sons spécifiques sur la perception de l'environnement : Defréville et al., 2004; Devergie, 2006; Dubois et al., 2006; Guastavino, 2003, 2006; Maffiolo, 1999; Niessen et al., 2010; Raimbault, 2002; Raimbault and Dubois, 2005
- 3 articles proposent une classification morpho-typologique, divisant l'environnement sonore urbain en "zones sonores" possédant une identité acoustique forte, selon la configuration et la pratique du site : Beaumont et al., 2004; Maffiolo, 1999; Polack et al., 2008
- 2 articles répertorient et classifient les sources sonores d'un point de vue expert : Brown et al., 2011; Leobon, 1986

La nature des classes est établie par rapport aux catégories perceptives, ou classes de sons, les plus souvent citées dans ces publications. A partir des éléments relevés, nous établissons deux taxonomies : une pour les événements (cf. Figure 26.a), une autre pour les textures (cf. Figure 26.b). Comme évoqué à la section **??**, la structure taxonomique de ces deux ensembles s'inspire grandement de l'axe vertical de l'organisation catégorielle proposée par E. Rosch (Cf. section **??**), i. e. plus le niveau d'abstraction de la classe est élevé, plus la description de la classe est précise, et plus les sources sonores incluses dans cette classe sont semblables (Cf. Figure 19). Pour les événements, nous considérons quatre niveaux d'abstraction allant des classes les plus globalisantes (niveau d'abstraction 0) aux classes les plus spécifiques (niveau d'abstraction 3). Pour les textures nous ne considérons que trois niveaux d'abstraction.

GL : Pour les événements, les regroupements se font en grande majorité par rapport à la source et sont d'ordre sémantique. Pour les textures nous considérons également la nature des lieux hébergeant ces dernières (e. g. *parc, rue*). La typologie des classes d'événements suit la nomenclature source-action introduit à la section **??**. En ce sens, elle est très similaire à cette autre typologie de sources sonores urbaines, effectuée postérieurement (Salamon et al., 2014).

5.2.4 *Acquisition des sons isolés*

Sur la base des typologies précédemment établies, 483 sons ont été collectés, dont 381 événements et 102 textures.

Parmi les événements :

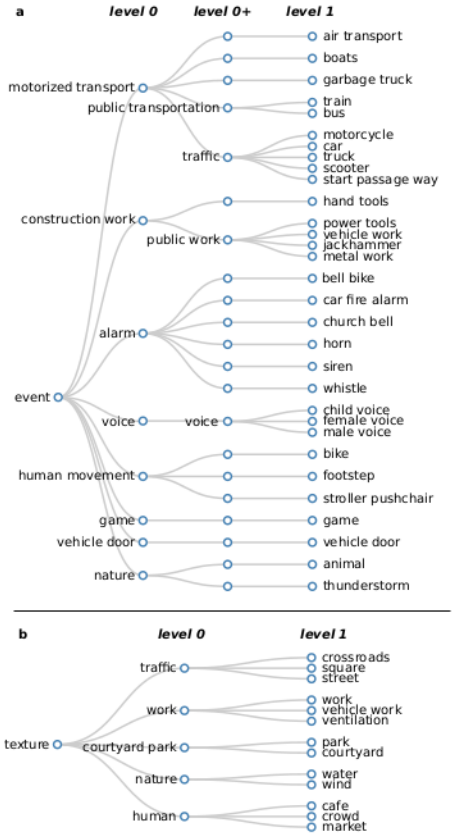


FIGURE 26 : Taxonomies des classes de sons utilisées pour la simulation des environnements sonores urbains pour (a) les événements sonores et (b) les textures sonores. Nous présentons ici uniquement les niveaux d'abstraction 0 et 1. Un niveau intermédiare, nommé 0+, et utilisé pour l'analyse, est également introduit.

- 260 sont issus d'enregistrements
- 89 sont issus de la banque de sons *SoundIdeas*[1]
- 32 sont issus de la banque de sons *Universal SoundBank*[2]

Parmi les textures :

- 72 sont issues d'enregistrements
- 23 sont issues de la banque de sons *SoundIdeas*
- 7 sont issues de la banque de sons *Universal SoundBank*

Tous les enregistrements ont été effectués à l'aide d'un micro canon *AT8035*[3] relié à un enregistreur *ZOOM H4n*[4]. L'utilisation du

1 Pour plus de détails sur *SoundIdeas* voir :http://www.sound-ideas.com/
2 Pour plus de détails sur *Universal SoundBank* voir : http://www.universal-soundbank.com/
3 Cf. http://eu.audio-technica.com/fr/products/product.asp?catID=1&subID=6&prodID=1845
4 Cf. http://www.zoom.co.jp/english/products/h4n/

brouhaha ?

micro canon nous permet d'isoler les événements sonores du magma urbain. Pour les textures, il nous permet d'éviter les événements sonores proches du preneur de son. Nous pouvons ainsi pointer des "zones sonores", en nous tenant à une certaine distance de ces dernières afin de capter uniquement le brouhaha émanant de la zone ciblée.

Tous les sons ont été normalisés au même niveau RMS [5] de −12 dB (FS) [6].

5.2.5 *Planification expérimentale*

5.2.5.1 *Épreuve de simulation*

Procédure

Les sujets doivent simuler deux environnements sonores urbains, chacune des scènes devant durer 1 minute. Pour ces simulations, les sujets doivent se conformer aux consignes suivantes :

- Première simulation : Simuler un paysage sonore **urbain plausible** qui selon vous est idéal (où vous aimeriez vivre).
- Deuxième simulation : Simuler un paysage sonore **urbain plausible** qui selon vous est non-idéal (où vous n'aimeriez pas vivre).

Tous les sujets commencent par simuler l'environnement idéal. Les sujets ne prennent connaissance de la deuxième consigne qu'à la fin de la première simulation.

Les sujets son totalement libres dans le choix des sons, et des paramètres (pour plus de détails sur les paramètres se référer à la section **??**). Ils doivent cependant se soumettre à deux contraintes :

- le sujet doit prendre le point de vue d'un auditeur fixe
- le paysage sonore doit être réaliste au sens de physiquement plausible. Autrement dit, le sujet à tout à fait le droit de placer 10 chiens dans son paysage sonore, mais il n'a pas le droit de placer 1 chien aboyant toutes les 10 millisecondes.

un

Chaque processus de simulation comprend deux parties :

1. la réalisation de la simulation : cette étape peut, elle même, se décomposer en trois actions (Cf. section **??**) :

5 Le niveau RMS, de l'anglais *Root Mean Square* qui désigne la valeur efficace d'un signal. Formellement, le niveau RMS x_{RMS} d'un signal $x = (x_1, x_2, \ldots, x_n)$ s'obtient en calculant la moyenne quadratique de ce dernier $x_{RMS} = \sqrt{\frac{1}{n}\sum_i x_i^2}$

6 dB (FS) est le sigle anglais désignant une valeur en décibels relative à la pleine échelle (*relative to Full Scale*), i. e. le rapport entre le niveau du signal et sa valeur maximale. Dans notre cas, ce niveau pleine échelle est de 1 Volt.

Index	Tâche	Durée (min)
1	Présentation de l'expérience Lecture de la consigne	10
2	Tutoriel (Réalisation d'une scène test)	20
3	Première simulation : scène idéale	40
4	Commentaire de la scène idéale	15
3	Deuxième simulation : scène non-idéale	40
4	Commentaire de la scène non-idéale	15
5	Critique de l'interface de simulation et de l'interface de sélection	10

TABLE 2 : Résumé des étapes de l'expérience de simulation

- sélectionner les classes de sons
- nommer les classes de sons sélectionnées
- paramétrer les pistes (Cf. section **??**) relatives aux classes de sons sélectionnées

2. la réalisation d'un commentaire libre du paysage sonore simulé

En complément, et une fois les deux scènes sonores réalisées, le sujet est invité à

- indiquer les sources sonores manquantes
- commenter l'ergonomie du logiciel de simulation
- commenter l'ergonomie de l'interface de sélection

je sais ce que tu veux dire, mais si je ne savais pas je ne comprendrais pas

Avant de commencer la première simulation, un petit tutoriel de 20 minutes est proposé aux sujets, afin qu'ils se familiarisent avec le logiciel de simulation, et la banque de données. Le tableau 2 résume les étapes de l'expérience ainsi que leurs durées respectives. L'expérience est prévue pour durer 2h30.

inutile

Apparatus Dispositif expérimental

Tous les sujets passent l'expérience sur des machines identiques (GL : description des machines). L'audio est présenté en stéréophonie, par le biais de casques audio. Pendant le tutoriel, les sujets doivent ajuster le niveau sonore à un volume confortable. Ils ne peuvent le modifier par la suite.

Tous les sujets réalisent l'expérience simultanément. Ils sont répartis de manière égale dans trois pièces identiques, toutes possédant un environnement calme. Ils n'ont pas le droit de s'adresser la parole pendant l'expérience.

Trois expérimentateurs, un dans chaque pièce, sont présents durant la totalité de l'expérience, afin de contrôler le bon déroulement de cette dernière, et de répondre aux éventuelles questions des sujets.

Participants

44 étudiants (14 femmes) de L'École Centrale de Nantes ont participé à l'expérience. Ils ont tous sensiblement le même âge (moyenne : 21.6, écart-type : 2). Tous les sujets ont vécu dans la même ville (Nantes), au minimum pendant les deux dernières années précédant l'expérience. Tous les sujets sont Nantais, et ont vécu dans cette ville durant les deux années, au moins, qui ont précédé l'expérience.

Sur les 44 sujets, 40 réalisent l'expérience avec succès, produisant au final 80 scènes sonores simulées, dont 40 scènes idéales, et 40 scènes non idéales. 4 sont éliminés pour non respect et/ou incompréhension des consignes, d'une part, dépassement du temps, d'autre part.

5.2.5.2 *Épreuve d'évaluation de l'agrément*

Procédure

orelles

En raison de contraintes temps, les sujets n'évaluent que des séquences de 30 secondes des scènes simulées, chacune de ces séquences commençant à la seconde 15, et finissant à le seconde 45, de la scène évaluée.

L'évaluation s'effectue sur une échelle sémantique bipolaire de 7 points allant de -3 (non-idéale/très désagréable) à +3 (idéale/très agréable). Avant de noter une scène, les sujets doivent obligatoirement écouter les 20 premières secondes de cette dernière. Après la notation, ils sont libres de passer à la scène suivante.

de calibrer ?

Pour chaque sujet, les scènes sont présentées dans un ordre aléatoire. Les 10 premières scènes permettent au sujet d'ajuster ses notes. Elles sont obligatoirement composées de 5 scènes idéales et de 5 non-idéales. Ces 10 premières scènes sont rejouées à la fin de l'expérience, et seules les notes données à la deuxième occurrence sont prises en compte.

Apparatus

pas précisé avant, pourquoi ?

Tous les sujets passent l'expérience sur des machines identiques (GL : description des machines). L'audio est présenté en stéréophonie, par le biais de casques audio semi-ouvert *Beyer-Dynamic DT 990 Pro*. Toutes les scènes sonores ont été re-simulées sur la base des partitions obtenues lors de l'expérience de simulation. Le niveau sonore de sortie est identique pour tous les sujets.

Tous les sujets réalisent l'expérience simultanément, dans un environnement calme. Ils n'ont pas le droit de s'adresser la parole pendant l'expérience.

Un expérimentateur est présent durant la totalité de l'expérience, afin de contrôler le bon déroulement de cette dernière, et de répondre aux éventuelles questions des sujets.

Participants

10 étudiants (2 femmes) de L'École Centrale de Nantes ont participé à l'expérience. Aucun d'entre eux n'a réalisé l'expérience de simulation. Tous les sujets ont sensiblement le même âge (moyenne : 23.1, écart-type : 1.8). Tous les sujets sont Nantais, et vivent dans cette ville depuis deux ans ou plus.

Tous les sujets ont réalisé l'expérience avec succès.

5.2.6 *Données et méthodes d'analyses*

5.2.6.1 *Nature des données analysées*

A partir des données produites par l'épreuve de simulation, nous analysons :

- les partitions des scènes simulées
- les signaux des scènes simulées
- les commentaires sur les sons manquants et l'ergonomie des interfaces de simulation et de sélection

Chaque scène est décrite par un groupe de descripteurs. C'est sur la base de ces descripteurs que nous pratiquons l'analyse. Un résumé des descripteurs, ainsi que des acronymes les désignant est présenté dans le Tableau 3. Afin de rester cohérent avec l'épreuve d'évaluation, les descripteurs issus des partitions ou des signaux des scènes ne sont pas calculés sur la durée totale de celles-ci, mais sur une version réduite de 30 secondes (Cf. section 5.2.5.2).

Pour chaque scène sonore, trois types de descripteurs sont considérés :

- *Perceptif* : Il s'agit de l'agrément perçu de la scène simulée, évalué sur une échelle sémantique 7 points. Nous notons A l'agrément moyen d'une scène, obtenu en moyennant les notes de tous les sujets. Considérant le faible nombre de sujets, nous faisons le choix, dans cette étude, de ne pas normaliser les notes d'agrément.
- *Sémantique* : Il s'agit d'un vecteur booléen noté $S = (x_1, x_2, \ldots, x_n)$ indiquant les classes de sons présentes dans la scène. Chaque

point x de ce vecteur correspond à une classe de sons particulière : $x = 1$ si la classe est présente dans la scène, et $x = 0$ autrement. La dimension n des vecteurs dépend du niveau d'abstraction considéré, e. g. pour le niveau d'abstraction 1, qui comprend 44 classes de sons, cette dimension sera de $n = 44$.

- *Structurel* : Les descripteurs structurels sont calculés à partir des partitions et des signaux des scènes simulées. Trois descripteurs structurels sont envisagés :
  - *Diversité* (DIV) : Il s'agit d'un scalaire représentant la diversité des classes sonores utilisées pour simuler une scène. Nous calculons DIV en comptant le nombre de classes de sons distinctes utilisées pour une simulation. Ce nombre dépend du niveau d'abstraction considéré. Considérons les deux sous classes du niveau d'abstraction 2 *passage de voiture* et *démarrage de voiture*, toutes deux appartenant à la classe *voiture* du niveau d'abstraction 1. Nous comptons deux classes pour la diversité des niveaux d'abstraction 2 et 1, et seulement une pour les niveaux d'abstraction 0 et 1.

    Par exemple, considérant..., nous comptons...
  - *Densité* (D) : Il s'agit d'un scalaire représentant le nombre de sources sonores présentes en moyenne. Pour obtenir D, nous calculons le logarithme du nombre d'éléments sonores par fenêtre de 125 millisecondes (sans recouvrement), et moyennons au cours du temps. Le calcul de D peut inclure toutes les sources sonores de la scène, ou seulement une partie. Dans ce cas, les fenêtres ne contenant pas de sources sonores ne sont pas prises en compte. Nous notons $D(E)$ et $D(T)$ les densités calculées en considérant séparément les sources d'événements et de textures sonores.
  - *Niveau Sonore* (L) : Pour représenter le niveau sonore, nous nous inspirons de la mesure L_{Aeq}. Dans notre cas, il s'agit d'un scalaire, calculé sur le signal en volts, et non en pression, et donné en décibels en prenant un référentiel de 1 Volt. Le niveau est obtenu en calculant, toutes les secondes, la moyenne quadratique du signal, et en moyennant sur la durée de la scène. Un filtrage de type A est opéré avant le calcul des moyennes quadratiques. D'autre descripteurs, inspirés eux aussi de descripteurs acoustiques classiques (L_{Amin}, L_{Amax}, L_{A10-90}), et utilisant un opérateur autre que la moyenne (minimum, maximum, les 10-90ème quantiles) pour intégrer les fenêtres de 1 seconde, ont été testés. Mais, ces derniers présentant tous une corrélation élevée avec L ($r_{pearson} \geqslant 0.76$, $p < 0.01$), nous conservons le scalaire ci-devant mentionné comme unique descripteur objectif du niveau sonore.

Descripteurs	Acronymes	Descripteurs	Acronymes
Densité	D	Niveau (textures)	L(T)
Densité (événements)	D(E)	Diversité	DIV
Densité (textures)	D(T)	Diversité (événements)	DIV(E)
Niveau	L	Diversité (textures)	DIV(T)
Niveau (événements)	L(E)	Agrément moyen (par scène)	A

Termes	Acronymes
Idéal/agréable	i
non-idéale/désagréable	ni
Scène idéale/agréable	i-scène
Scène non-idéale/désagréable	ni-scène

TABLE 3 : TODO

5.2.6.2 *Méthodologie et Outils statistiques*

Afin d'évaluer l'impact spécifique des différentes sources sonores sur l'agrément perçu, nous soumettons nos travaux aux six tests/études de significativité présentés ci-après :

- *Étude qualitative* : Afin de vérifier la validité écologique de 1) la banque de données et 2) l'interface de sélection, nous réalisons une étude qualitative des critiques ergonomiques effectuées par les sujets.
- *Étude comparative entre les descripteurs structurels* : Afin d'évaluer si la distinction affective imposée entre les i- et ni-scènes impacte de manière significative la nature des scènes, i.e. si il existe des différences significatives entre les descripteurs structurels et/ou l'agrément perçu, nous évaluons cette significativité à partir d'un test de Student à deux échantillons appariés (Cf. Annexe A). s'il
- *Étude de l'influence des descripteurs structurels sur l'agrément perçu* : Afin d'évaluer l'impact potentiel des descripteurs structurels sur l'agrément perçu, nous étudions l'existence de corrélation linéaire entre ces deux types de descripteurs. Pour mesurer la corrélation, nous utilisons le coefficient de Pearson (Cf. Annexe A).

Nous adoptons ici une méthodologie couramment utilisée dans l'approche dimensionnelle.

- *Étude comparative entre les descripteurs sémantiques* : Afin d'apprécier si la distinction affective imposée a eu un impact sur la composition des scènes en terme de sources sonores, ou, pour être plus précis, s'il existe des classes de sons qui ont été particulièrement utilisées pour simuler un type d'environnement, nous utilisons le V-test. Nous vérifions si la présence d'une classe de sons est typique d'un environnement (i ou ni). Le test est effectué pour chaque niveau d'abstraction, et séparément pour les classes d'événements et de textures. Pour chaque classe j et chaque type d'environnements k ($k = i, ni$), la valeur V_{jk} du V-test se calcule comme suit :

$$V_{jk} = \frac{c_{jk} - c_k \frac{c_j}{c}}{\sqrt{c_k \frac{c - c_k}{c-1} \frac{c_j}{c} (1 - \frac{c_j}{c})}}$$

où c est le...

avec c le nombre de classes utilisées, c_k le nombre de classes utilisés pour un type d'environnements k, c_j le nombre de classes j utilisées, et c_{jk} le nombre de classes j utilisées pour un type d'environnements k. Le V-test teste l'hypothèse nulle que la proportion $\frac{c_{jk}}{c}$ ne diffère pas significativement de la proportion $\frac{c_{jk}}{c_k}$. Si pour un environnement k et une classe j l'hypothèse est rejetée, la classe j est alors typique de l'environnement k. Les classes typiques sont nommées **les marqueurs sonores**.

- *Étude des espaces de représentations induits par les descripteurs sémantiques* : Afin d'étudier si une représentation basée uniquement sur la présence ou l'absence des classes de sons permet de séparer les deux types d'environnements, nous considérons l'espace induit par les descripteurs sémantiques S. S étant un vecteur booléen, nous calculons les distances entre les scènes à partir de la distance de *Hamming*. Considérant les deux vecteurs $S_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$ et $S_2 = (x_{2,1}, x_{2,2}, \ldots, x_{2,n})$ de dimension n, avec $x = 0, 1$, la distance de *Hamming* d_{ham} mesure le pourcentage de coordonnés qui diffèrent entre les deux vecteurs.

Nom propre, pas de raison de mettre en italiques

$$d_{ham}(S_1, S_2) = \frac{1}{n} \sum_{i=1}^{n} (x_{1,i} \bigoplus x_{2,i})$$

où $\bigoplus$ désigne l'opérateur du *ou-exclusif*. Plus la composition des deux scènes est similaire, et plus ces deux scènes seront proches. L'utilisation de la distance de *Hamming* permet de prendre en

compte de manière égale les classes présentes et absentes. Pour mesurer la capacité intrinsèque de l'espace à séparer les i- et ni-scènes, nous utilisons une métrique de *clustering* nommée précision au rang k (P@k). GL : La P@k mesure la précision obtenue après que k items ait été retrouvés. Formellement, pour chaque scène s_i, nous calculons le rapport entre le nombre de scènes s_j, prises parmi les k plus proche(s) voisin(s) de s_i et partageant le même label que s_i, sur le nombre d'items à retrouver (k). La P@k est alors la moyenne des rapports pour tous les items.

ont

- *Étude de l'influence spécifique des marqueurs sonores sur l'agrément perçu* : Afin d'évaluer les contributions spécifiques de certaines sources sonores, nous évaluons une nouvelle fois l'impact potentiel des descripteurs structurels sur l'agrément perçu, mais en ne tenant compte, cette fois, que des marqueurs sonores pour calculer ces descripteurs.

Excepté le V-test, tous les tests de significativité sont effectués avec un seuil critique $\alpha = 0.05$. Pour le V-test, étant donné que nous testons beaucoup de classes, une correction de Bonferroni GL : ref est appliquée. Concernant les valeurs p, dans le cas où la valeur $p \geqslant 0.05$, nous indiquons sa valeur. Dans le cas ou $0.01 \leqslant p < 0.05$, nous indiquons seulement $p < 0.05$. Dans le dernier cas nous indiquons $p < 0.01$.

Concernant l'interprétation du coefficient de corrélation de Pearson adoptée dans ce document, nous invitons le lecteur à se référer à l'Annexe A.3.

5.2.7 *Validité écologique de l'expérience*

5.2.7.1 *Diversité de la banque de sons*

Nous voulons vérifier que la diversité des classes de sons proposées est suffisante, pour pouvoir simuler un environnent sonore. Nous analysons les commentaires des sujets sur la banque de données. 63% d'entre eux indiquent avoir été, au moins une fois, dans l'incapacité de trouver un son, avec un maximum de 4 sons par sujet. Parmi les sons manquants relevés, nous identifions 26 classes de sons dont :

- 16 sont bien présentes dans la banque de données, l'incapacité des sujets à les trouver n'étant donc pas imputable à la diversité de la base.
- 1 fait référence à des sons de musique, que nous avons choisi délibérément d'occulter.
- 9 sont effectivement absentes.

Concernant ces dernières, nous observons qu'il s'agit de classes très spécifiques (e. g. *voiture de sport* ou *voix d'adolescent*), et qui peuvent

être remplacées par des classes similaires (e. g. *voiture* ou *voix d'enfant* ou *voix d'adulte*). Nous en concluons que la diversité proposée par la banque de sons est suffisante dans le cadre de notre étude.

5.2.7.2 *Ergonomie de l'interface de sélection*

Nous voulons vérifier l'efficience de l'interface de sélection. Nous analysons les retours des sujets. 32.5% d'entre eux indiquent spontanément que l'interface est un moyen " simple et efficace " de sélectionner des sons sans l'aide de texte. 57.5% ne font pas mention de difficultés particulières, 10% signalent enfin avoir rencontré des difficultés avec l'interface, sans toutefois que la simulation en ait été affectée.

Nous en concluons que l'interface de sélection sans texte ne perturbe pas les sujets outre mesure. Un même constat avait été tiré de l'expérience pilote (Lafay, 2013; Lafay et al., 2014).

GL : TODO : vérifier une dernière fois

5.2.7.3 *Ergonomie de l'interface de simulation*

5.2.8 *Vérification de l'agrément des scènes simulées*

même remarque sur \mathcal

Nous analysons ici l'agrément perçu des 80 scènes sonores simulées. La Figure 27a affiche l'agrément moyen A pour les i- et ni-scènes.

Dans un premier temps, et afin de garantir la cohérence de nos données, nous voulons nous assurer qu'aucune ni-scène n'ait un A supérieur à celui d'une i-scène. Quatre des scènes ne respectent pas la contrainte. Elles et leurs correspondantes i ou ni sont retirées. 36 i-scènes et 36 ni-scènes restent dans le champ de l'analyse.

Dans un deuxième temps, nous voulons tester si les sujets ont bien perçu une différence d'agrément entre les i- et ni-scènes. Pour ce faire, nous observons l'agrément moyen de chaque sujet, calculé séparément pour chaque type d'environnement (Cf. Figure 27b). Il apparaît que les i-scènes ont bien été perçues comme significativement plus agréables ($p < 0.01$) que les ni-scènes.

5.2.9 *Étude comparative entre les descripteurs structurels*

En premier lieu, nous nous concentrons sur le niveau sonore. Les figures 28a, 28b et 28c affichent les distributions des niveaux L, $L(E)$ et $L(T)$. Il existe bien une différence de niveaux significative entre les i- et ni-scènes ($L : p < 0.01$), avec un écart moyen de -7 dB. Cette différence affecte aussi bien les événements ($L(E) : p < 0.01$, écart moyen : -7 dB) que les textures ($L(T) : p < 0.01$, écart moyen : -6 dB).

Nous vérifions, sans surprise, que le niveau des sources sonores est bien un indicateur d'agrément, les ni-scènes ayant tendance à être plus fortes. Nous constatons encre que cette différence de niveaux

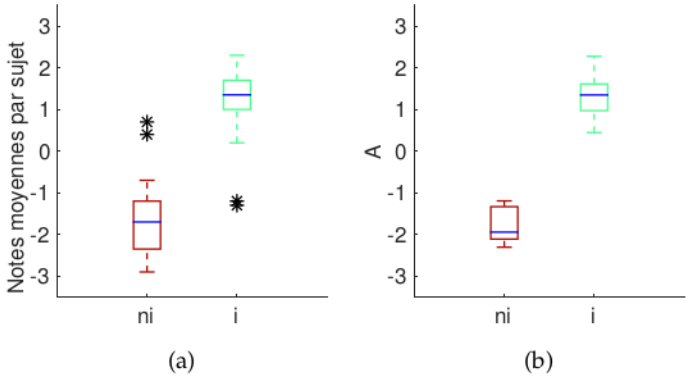


FIGURE 27 : TODO

s'observe de manière égale pour les événements et les textures sonores.

Il apparaît que ce sont les événements qui impactent le plus le niveau global des scènes, l'écart entre L et L(E) n'étant que de 1 dB pour les i-scènes et les ni-scènes. Cette observation fait écho aux résultats obtenus par Kuwano et *al.*(Kuwano et al., 2003). Au cours de leur expérience, les auteurs demandent à leurs sujets d'évaluer une série d'environnements sonores d'abord de manière globale, ensuite, d'en évaluer le niveau aux instants où chacun identifie une source sonore. L'étude montre qu'il n'y a pas de différences significatives entre les jugements globaux et les moyennes des jugements instantanés. Pour en revenir à nous, c'est comme si nos propres sujets avaient inconsciemment tenu compte de cette réalité perceptive lors de la simulation, en faisant porter le niveau sonore global par des sons courts et bien identifiés, i. e. les événements.

Nous observons enfin que le niveau seul ne permet pas de clairement faire la distinction entre les différents types d'environnements. En effet, 20% des i-scènes ont un niveau supérieur au niveau minimal des ni-scènes, alors qu'il n'y a pas de recouvrement, si l'on considère l'agrément perçu A.

En second lieu, nous nous penchons sur les densités de sources sonores. Les Figures 29a et 29b affichent les distributions de D et D(E). Que l'on prenne en compte toutes les sources, ou uniquement les événements, la densité est significativement plus élevée pour les ni-scènes (D : $p < 0.01$, D(E) : $p < 0.01$). Nous observons un écart moyen de +0.36 pour D (soit en moyenne 2.3 sources sonores par fenêtre de plus pour les ni-scènes), et de +0.32 pour D(E) (soit en moyenne 2.1 sources sonores par fenêtre de plus pour les ni-scènes). Si ces écarts sont très similaires, c'est que la densité des textures D(T) ne varie pas de manière significative entre les i- et ni-scènes (D(T) : $p < 0.08$), l'écart moyen étant de +0.17 (soit en moyenne 0.7 sources

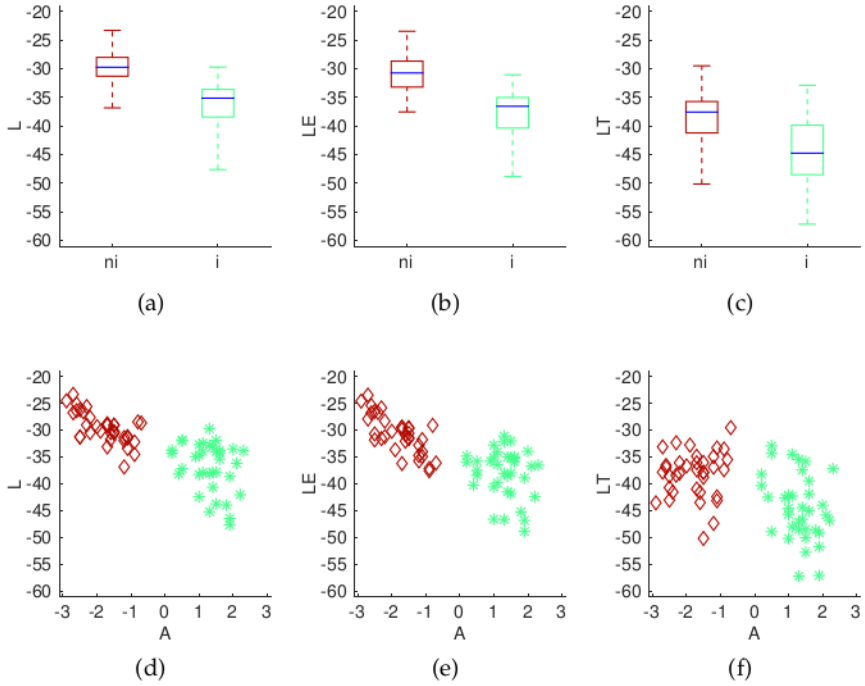


FIGURE 28 : TODO

sonores par fenêtre de plus pour les ni-scènes), et l'écart médian étant quant à lui nul.

Nous constatons ici que la densité peut être un indicateur de qualité, si l'on considère uniquement les événements sonores. Comme pour les niveaux sonores, la densité ne permet pas de clairement séparer les i- et ni-scènes, 43% des i-scènes ayant un D(E) supérieur à la densité d'événement minimale des ni-scènes.

En dernier lieu, nous nous intéressons à la diversité. Nous affichons sur la figure 30 DIV(E) et DiV(T), en séparant les différents niveaux d'abstractions. Excepté pour le niveau d'abstraction 0, la diversité de classes d'événements sonores est plus élevée pour les ni-scènes (DIV(E) niveaux 1,2 et 3 : $p < 0.01$), avec en moyenne 2 classes présentes en plus. Aucune différence significative n'est observée pour les textures.

Les tendances globales observées montrent, d'une part, qu'un environnement sonore non-idéal est plus fort, plus dense et composé d'une plus grande variété d'événements sonores qu'un environnement sonore idéal. Elles montrent, d'autre part, que ce sont les caractéristiques des événements, plus que celles des textures, qui semblent porter la distinction entre les i- et ni-scènes. Cependant, aucun des descripteurs ne permet, à lui seul, de faire une distinction nette entre les deux types d'environnements, distinction pourtant perçue de manière non ambiguë par les sujets.

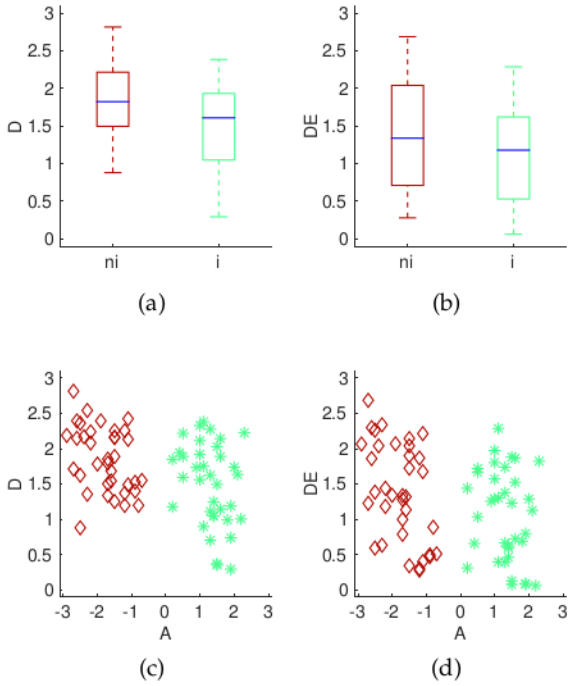


FIGURE 29 : TODO

5.2.10 *Influence des descripteurs structurels sur l'agrément perçu*

Nous analysons, dans cette section, les relations fines qui peuvent exister entre les descripteurs structurels d'une part et l'agrément perçu d'autre part. Contrairement à la section précédente, où la qualité affective des scènes est représentée de manière binaire (i *vs.* ni), nous considérons, ici, l'agrément moyen A comme descripteur perceptif. Il s'agit d'étudier l'existence de potentielles corrélations entre les descripteurs structurels et A. Les coefficients de corrélation linéaire calculés entre A *vs.* L, $L(E)$, $L(T)$, D, $D(E)$ et $DIV(E)$ sont présentés dans le tableau 4. Les relations entre A et les descripteurs structurels sont illustrées par les figures 28d, 28e et 28f pour les niveaux sonores, et les figures 29c et 29d pour les densités.

Concernant L, on observe une forte corrélation négative ($r = -0.76$, $p < 0.01$) avec A, indiquant que plus le niveau sonore est élevé, plus la scène est désagréable. Cependant, la figure 28d suggère que cette relation ne s'opère pas de la même manière pour les i- et ni-scènes. En effet, la corrélation entre L et A pour les ni-scènes reste élevée ($r = -0.67$, $p < 0.01$), mais est inexistante pour les i-scènes.

Cette corrélation élevée, considérant l'ensemble des scènes, résulte du fait que les i-scènes ont tendance à être moins fortes que les ni-scènes, donnant ainsi l'illusion de prolonger la corrélation négative observée pour les ni-scènes.

Nous en concluons que L :

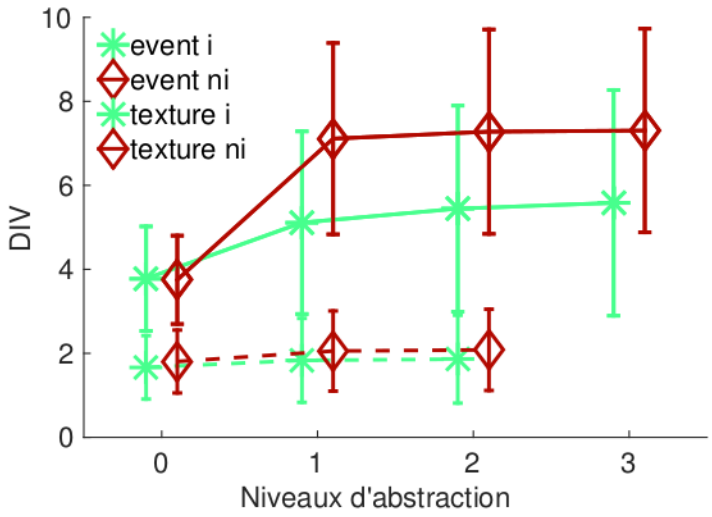


FIGURE 30 : TODO

- permet bien de faire la distinction entre les i- et ni-scènes,
- permet de finement caractériser l'agrément perçu des ni-scènes,
- n'est pas un indicateur pertinent de l'agrément perçu pour des environnements a priori agréables.

Les mêmes observations sont faîtes concernant $L(E)$ (cf. reffig :soundlevele). Pour $L(T)$ (cf. reffig :soundlevelf), bien que, à considérer l'ensemble des scènes, on observe une corrélation modérée, cela n'est pas vérifié quand on regarde séparément les i-scènes ($r = -0.33$, $p = 0.05$) et les ni-scènes ($r = 0.09$, $p = 0.06$). Là encore on peut penser que la corrélation négative observée pour l'ensemble des scènes est un artefact, résultant du fait que le niveau des textures des i-scènes a tendance à être plus bas que celui des ni-scènes. Ainsi, si les événements sonores conservent une certaine capacité de prédiction de l'agrément pour les ni-scènes, le niveau des textures n'apporte lui que peu d'informations, quel que soit l'environnement.

Considérant l'ensemble des scènes, nous observons une corrélation négative faible pour D ($r = -0.41$, $p < 0.01$) et $D(E)$ ($r = -0.34$, $p < 0.01$). Là encore, une relation du même acabit est observée pour les ni-scènes, mais aucune corrélation n'est observée pour les i-scènes. La densité de sources sonores semble donc avoir un faible impact sur l'agrément perçu si l'on considère les ni-scènes, mais, comme pour les niveaux, la densité ne semble pas avoir d'impact pour les i-scènes.

En ce qui concerne la diversité des classes d'événements, une corrélation négative faible est obtenue pour les niveaux d'abstraction 1, 2 et 3 en tenant compte de l'ensemble des scènes. Si l'on considère les i- et ni-scènes séparément, aucune corrélation significative n'est trouvée. Les conclusions sont similaires à celles faites pour $L(T)$: la diversité permet uniquement de faire la distinction entre les deux

	ensemble	i-scènes	ni-scènes
L	**-0.76** ($p < 0.01$)	-0.32 ($p = 0.06$)	**-0.67** ($p < 0.01$)
L(E)	**-0.75** ($p < 0.01$)	-0.20 ($p = 0.24$)	**-0.75** ($p < 0.01$)
L(T)	**-0.51** ($p < 0.01$)	-0.33 ($p = 0.05$)	0.09 ($p = 0.6$)
D	**-0.41** ($p < 0.01$)	-0.31 ($p = 0.07$)	**-0.37** ($p = 0.03$)
D(E)	**-0.34** ($p < 0.01$)	-0.22 ($p = 0.21$)	**-0.47** ($p < 0.01$)
DIV(E) 0	-0.07 ($p = 0.57$)	-0.25 ($p = 0.15$)	-0.22 ($p = 0.23$)
DIV(E) 1	**-0.46** ($p < 0.01$)	-0.25 ($p = 0.14$)	-0.18 ($p = 0.30$)
DIV(E) 2	**-0.40** ($p < 0.01$)	-0.21 ($p = 0.22$)	-0.18 ($p = 0.30$)
DIV(E) 3	**-0.36** ($p < 0.01$)	-0.18 ($p = 0.30$)	-0.18 ($p = 0.30$)

TABLE 4 : Coefficients de corrélation linéaire calculés entre l'agrément perçu moyen A *vs.* TODO

types d'environnements, mais ne permet pas de caractériser précisément l'agrément perçu.

En résumer, en présence d'un environnement désagréable, les niveaux sonores, et en particulier ceux des événements, ainsi que, dans une moindre mesure, la densité de sources présentes, ont un impact négatif sur l'agrément. En présence d'un environnement agréable, aucun des descripteurs structurels considérés ici ne semble influer sur la perception de l'agrément.

Ces premiers résultats pourraient montrer qu'il existe deux modes de perception, mobilisant chacun des descripteurs indépendants, modes qui s'activent en fonction de la nature de l'environnement en présence (i ou ni).

Le fait qu'aucun des descripteurs globaux ne permettent de caractériser l'agrément des i-scènes peut nous amener à penser que toutes les sources sonores ne contribuent pas de manière égale à la perception de l'agrément, mais, que seules les caractéristiques de certaines d'entre elles ont une réelle influence. Afin approfondir ce point, nous analysons, dans la section suivante, les scènes d'un point de vue sémantique, i. e. en nous intéressant à la nature des sources qui les composent.

5.2.11 *Étude comparative entre les descripteurs sémantiques*

5.2.11.1 *Analyse qualitative*

Nous analysons la composition des scènes en comptant le nombre de sujets ayant utilisé une classe de sons pour simuler un type d'environnements. Les résultats sont présentés à la figure 31a pour les événements et à la figure 31b pour les textures. Par souci d'espace, nous

choisissons un niveau d'abstraction intermédiaire entre le niveau 0 et 1, noté 0+, pour représenter les classes (cf. Figure 26).

Nous observons une différence notable dans le choix des classes entre les i- et ni-scènes. La répartition des classes est très proche de celles obtenues dans une étude similaire sur les environnements sonores urbains idéaux (Guastavino, 2006), i. e. les classes suggérant la présence humaine et la nature sont très présentes dans les i-scènes, a contrario les classes désignant des sons mécaniques et de travaux sont principalement utilisées pour les ni-scènes.

Ces résultats confirment un fait déjà observé, la nature sémantique des sources sonores joue un rôle prédominant dans l'appréciation de l'environnement (Dubois et al., 2006; Raimbault and Dubois, 2005).

Nous notons quelques différences avec (Guastavino, 2006): les résultats obtenus par Guastavino et *al.*montrent que les sons de *transports publics* sont caractéristiques des environnements sonores urbains idéaux. Les auteurs attribuent cela au fait que la perception de l'agrément est, entre autre, soumise à un contexte socio-culturel. Dans notre représentation du monde, les sons de transports publics sont positivement connotés, et ont ainsi tendance à être mieux acceptés que les sons de véhicules privés.

Dans une certaine mesure, nos résultats contredisent ce fait. La figure 31a montre, en effet, que les classes d'événements de *transports publics* (*bus* et *train*, cf. Figure reffig :soundsourcec) ont été utilisées par les sujets, pour des i-scènes, dans 28% des cas, et pour des ni-scènes, dans 42% des cas. Les résultats ne remettent pas en question le fait que les sons de *transports publics* soient bien acceptés : 25% des sujets ont utilisés la classe *bus* pour les i-scènes, un chiffre comparable à celui de la classe *Vélo*, et bien supérieur à celui de toute autre classe de véhicules privés. Cependant les classes *transports publics* sont également bien présentes dans les ni-scènes, plus que les classes *voiture* ou *camion* par exemple. La classe *transports publics* ne peut donc pas être considérée comme typique d'un environnement sonore urbain idéal.

Cette différence peut s'expliquer par la nature des deux protocoles expérimentaux utilisés. Comme nous l'avons fait, Guastavino et *al.*demandent à ses sujets de décrire un environnement en se basant sur leurs mémoires. Mais, contrairement à nous, les sujets de Guastavino et *al.*ne disposent pas de supports sonores. Le fait que nos sujets soient confrontés à la réalité acoustique des sons pour recréer leurs environnements peut avoir pour effet de diminuer l'effet du contexte socio-culturel. D'autres études utilisant des sons comme stimuli montrent que la classe *bus* peut avoir un effet négatif sur l'appréciation de l'environnement (Lavandier and Defréville, 2006).

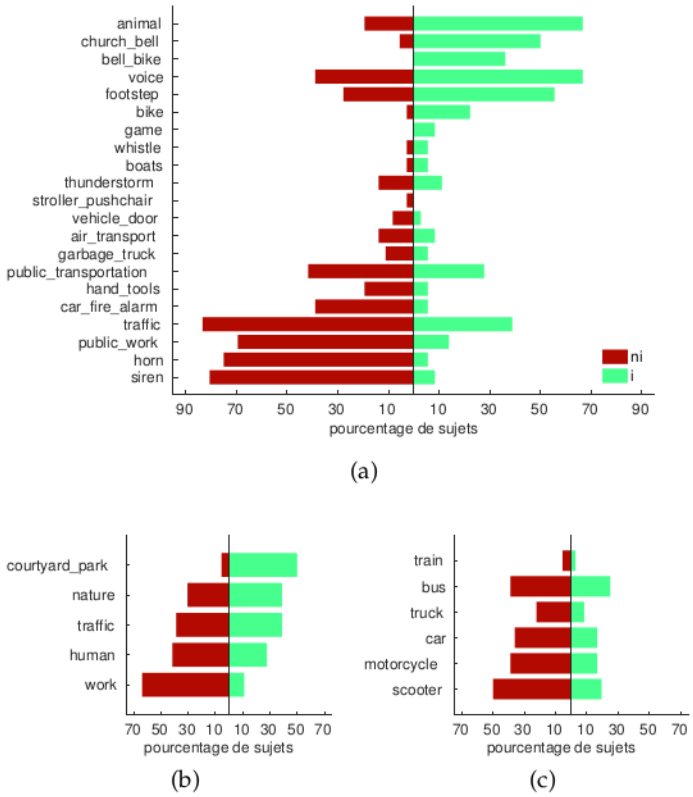


FIGURE 31 : TODO

5.2.11.2 *Marqueurs sonores*

Nous avons mis en évidence que, qualitativement, la composition des sources sonores des scènes diffère selon les types d'environnements (i ou ni). Nous essayons de voir maintenant si, parmi ces classes, certaines, appelées marqueurs sonores, sont typiques d'un environnement en particulier. Pour ce faire, nous utilisons le V-test (cf. section 5.2.6.2), en considérant séparément chaque niveau d'abstraction. Les résultats sont présentés dans le tableau 5.

on ne les appelle "marqueurs sonores" qu'après avoir montré qu'elles étaient typiques...

Concernant les événements sonores, 9 marqueurs sont identifiés sur l'ensemble des niveaux d'abstraction. Comme la figure 31 nous le laissait présager, les classes relatives à l'activité humaine (*pas homme béton,sonnette vélo*), et à la nature (*animaux, oiseaux,chants d'oiseaux*) sont des marqueurs de i-scènes. Nous notons également la présence de la classe *cloche* dans les marqueurs d'un environnement idéal. Ce fait est possiblement dû au background socio-culturel des sujets, dans leur grande majorité des citoyens européens. En effet, selon Schafer, un son reconnu par un individu comme faisant partie intégrante de son environnement est bien accepté. Les marqueurs de ni-scènes sont des classes faisant référence à des sons de travaux (*travaux*), ou suggérant un trafic dense (*klaxon*, *sirène*).

Niveau d'abstraction	Marqueurs sonores événements i-scènes	ni-scènes
0		construction work (3.78)
1	church bell (4.5) bell bike (4.3) animal (4.2)	horn (3.9) siren (3.9)
2	birds (4.8) church bell (4.4) bell bike (4.2)	horn (4.0) siren (4.0)
3	birds singing (4.8) church bell (4.3) bell bike (4.2) male footsteps concrete (3.6)	horn (4.1) siren (4.0)

	Marqueurs sonores textures i-scènes	ni-scènes
0	courtyard/park (4.1)	construction work (3.9)
1	park (3.65)	crossroads (3.6) vehicle work (3.3)
2	park (3.64)	crossroads (3.56)

TABLE 5 : Classes d'événements identifiées comme étant des marqueurs sonores. Dans chaque cellule, les marqueurs sont ordonnés par ordre décroissant de valeur V.

Concernant les textures sonores, 5 marqueurs sont identifiés. Pour les i-scènes, il s'agit de classes faisant référence à des ambiances amorphes, calmes, (*cour-intérieur/parc* et *parc*). Pour les ni-scènes, il s'agit, comme pour les événements, de classes faisant référence à des bruits de travaux (*travaux* et *véhicule de travaux*), ainsi que d'une classe faisant référence au trafic (*carrefour*).

incorr.

Bien que l'ensemble des marqueurs identifiés soient intuitifs, nous notons aucune des classes d'événements faisant directement référence aux bruits de véhicules motorisés dans leur nombre, exception faite du marqueur de texture (*carrefour*). Pour représenter un trafic désagréable, les sujets ont porté leur choix sur les classes *klaxon* et *sirène*. On peut supposer que les sons isolés de véhicules sont compris comme faisant partie intégrante de l'environnement urbain, et ne sont ainsi pas particulièrement associés à un environnement désagréable.

GL : ici analyse des caractéristiques des classes trafics
GL : ici reprendre les conclusions de (Lavandier and Defréville, 2006) et (Ricciardi et al., 2015)

5.2.12 *Étude des espaces de représentations induits par les descripteurs sémantiques*

Dans cette partie, nous évaluons la capacité d'une représentation sémantique à séparer les deux types d'environnements. Pour ce faire, nous calculons une précision au rang 5 (p@5) sur l'espace induit par les descripteurs sémantiques S, et ce pour chaque niveau d'abstraction (cf. section 5.2.6.2). Les vecteurs S sont construits en utilisant toutes les classes (ET), les classes d'événements (E), les classes de textures (T), les classes d'événements ne considérant que les marqueurs sonores (E_m), les classes d'événements ne considérant pas les marqueurs sonores $E_{w/o,m}$. Les résultats sont affichés sur la figure 32.

En ce qui concerne ET, la p@5 est de 76% pour le niveau d'abstraction 0, et reste supérieure à 86% à partir du niveau d'abstraction 1. Ces résultats confirment qu'il est possible de clairement distinguer les deux types d'environnements en se basant seulement sur la présence ou l'absence des classes de sons. Nous notons également que, plus le niveau d'abstraction est élevé, plus la capacité de séparer les environnements est importante. En d'autres termes, plus nous sommes précis dans notre description de la composition des scènes, plus nous sommes à même d'établir une distinction claire entre les i- et ni-scènes.

En considérant séparément E et T, il apparaît que 1) la p@5 obtenue avec E est similaire à celle obtenue avec ET, et 2) que la p@5 obtenue avec T est systématiquement inférieure d'environ 10 à 15% à celle de E. Ces résultats indiquent que l'information sémantique permettant de séparer les deux environnements est principalement portée par les événements. Ces résultats font, par ailleurs, écho aux travaux de (Maffiolo, 1999), qui montrent que nous analysons de manière descriptive (en identifiant les sources) les scènes événementielles,i. e. composées d'événements sonores (cf. section 3.4.5.2).

Enfin, il apparaît que la p@5 obtenue avec E_m est similaire, voire supérieure à celles de E et ET, et ce bien qu'une information partielle soit utilisée dans ce cas pour décrire les scènes. La dimension des vecteurs de description S pour E_m est en effet inférieure à la dimension des vecteurs S pour E, qui est elle même inférieure à celle obtenue dans le cas où toutes les classes sont utilisées (ET). De plus, Dans le cas où les marqueurs ne sont pas pris en compte pour la description ($E_{w/o,m}$), les résultats chutent, passant même en dessous de ceux obtenus en ne considérant que les textures. Ce qui confirme que la majorité de l'information sémantique permettant de faire le distinction entre les i- et ni-scènes est incluses dans les marqueurs. Cela

En résumé nous déduisons de cette analyse les points suivants

1. Contrairement à ce que nous avions constaté avec les descripteurs structurels, une description sémantique de la composition des scènes en terme de présence/absence de sources sonores

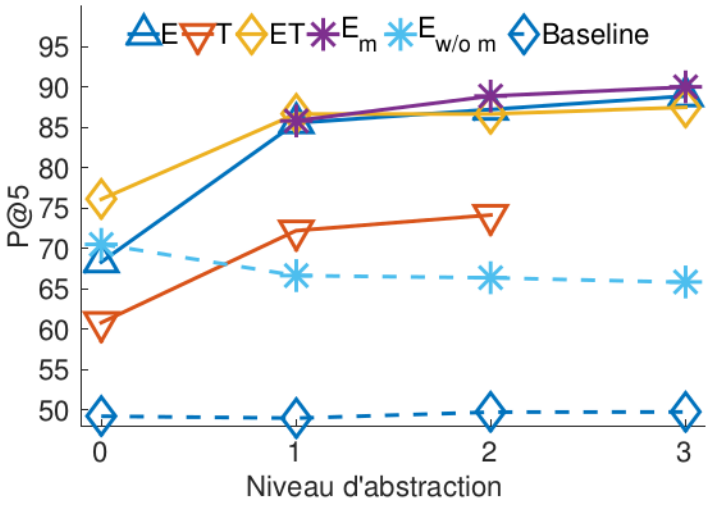


FIGURE 32 : TODO

permet de bien séparer les deux types d'environnements (i ou ni).

2. L'information sémantique est majoritairement portée par les classes d'événements sonores

3. Parmi les classes d'événements, seule une partie, i. e. les marqueurs sonores, sont nécessaires afin de faire la distinction entre les i- et ni-scènes

un peu "relâché"

Maintenant que nous avons isolé les classes typiques des i- et ni-scènes, et que nous avons vérifié que la distinction entre ces environnements dépendait de la présence de ces classes, il nous reste à voir si une description structurelle des scènes, basée uniquement sur ces marqueurs sonores, permet de caractériser l'agrément perçu, mieux qu'une description structurelle globale.

GL : analyse marqueur texture ?
GL : HCA sur S, + ANOVA sur A

5.2.13 *L'influence spécifique des marqueurs sonores sur l'agrément perçu*

Comme pour la section 5.2.10, nous évaluons les corrélations entre A et les descripteurs structurels. Pour cette section, les descripteurs structurels sont calculés en tenant compte des marqueurs sonores précédemment identifiés. Nous définissons X_m le descripteurs X calculé en ne prenant en compte que les sons des marqueurs. A l'inverse, nous définissons X_b (b : pour "bruit") le descripteurs X calculé en prenant en compte toutes les classes de sons excepté les marqueurs. Lorsque le descripteur caractérise une i-scène (idem pour une ni-

	i-scenes	ni-scenes
L_m	0.03 ($p = 0.88$)	**-0.63** ($p < 0.01$)
$L(E)_m$	0.08 ($p = 0.66$)	**-0.60** ($p < 0.01$)
$L(T)_m$	-0.11 ($p = 0.66$)	-0.00 ($p = 0.98$)
L_b	**-0.52** ($p < 0.01$)	-0.29 ($p = 0.09$)
$L(E)_b$	**-0.51** ($p < 0.01$)	-0.16 ($p = 0.36$)
$L(T)_b$	-0.32 ($p = 0.05$)	**-0.75** ($p < 0.01$)
$L_m - L_b$	**0.67** ($p < 0.01$)	-0.24 ($p = 0.15$)
$L(E)_m - L(E)_b$	**0.66** ($p < 0.01$)	-0.31 ($p = 0.06$)
$L(T)_m - L(T)_b$	0.16 ($p = 0.54$)	**0.39** ($p < 0.05$)
D_m	0.03 ($p = 0.85$)	-0.32 ($p = 0.05$)
$D(E)_m$	0.09 ($p = 0.62$)	**-0.47** ($p < 0.01$)

TABLE 6 : Coefficients de corrélation linéaire calculés entre l'agrément perçu moyen A *vs.* TODO.

scène), nous ne considérons, pour le calcul, que les marqueurs identifiés pour les i-scènes (ou pour les ni-scènes), que nous nommons i-marqueurs (ou ni-marqueurs). Les résultats sont affichés sur le tableau 6.

Considérons dans un premier temps les densités. Les résultats pour D_m et $D(E)_m$ sont similaires à ceux observés précédemment pour D et D(E), a l'exception de $D(E)_m$ qui ne présente plus une corrélation significative pour les ni-scènes. Ces résultats tendent à confirmer que la densité est un indicateur d'agrément de faible importance, qu'on la considère globalement, ou en prenant en compte les contributions séparées de différentes sources.

GL : TODO : Rajouter D_b

Concernant les niveaux sonores (Cf. Figures 34a, 34b, 34c, 34d, 34e et 34f), là encore les mêmes tendances sont observées entre L_m, $L(E)_m$ et $L(T)_m$, d'une part, et L, L(E) et L(T), d'autre part. Que l'on considère uniquement les marqueurs, ou l'ensemble des classes, il s'avère que :

1. Il existe une différence significative entre les niveaux des i- et ni-scènes (L_m, $L(E)_m$ et $L(T)_m$: $p < 0.01$)

2. Le niveau sonore des scènes est majoritairement porté par les événements sonores.

3. Le niveau sonore des événements a une influence sur la perception de l'agrément pour les ni-scènes, mais pas pour les i-scènes.

4. Le niveau sonore des textures ne joue aucun rôle dans la perception de l'agrément

En conclusion, le niveau des ni-marqueurs a une influence négative sur l'agrément pour les ni-scènes, en revanche le niveau des i-marqueurs n'impacte pas l'agrément perçu pour les i-scènes.

En considérant maintenant les classes non marqueurs, nous remarquons, sur les i-scènes, une corrélation négative modérée pour L_b ($r = -52$, $p < 0.01$) et $L(E)_b$ ($r = -51$, $p < 0.01$). C'est la première fois qu'un indicateur objectif nous permet de caractériser l'agrément des environnements agréables. Il s'avère que le niveau des classes de sons n'étant pas typique d'un environnement agréable a un impact négatif sur l'agrément.

préciser

Par ailleurs, alors que $L(T)$ ne présentait pas de corrélation pour les ni-scènes, une corrélation négative forte est observée pour $L(T)_b$ ($r - 0.75$, $p < 0.01$). Ce fait indique que des classes de textures, utilisées pour simuler aussi bien les i-scènes que les ni-scènes, n'affectent pas l'agrément perçu de la même manière. Pour une même classe *foule*, le niveau perçu dans un cadre idéal n'affectera pas la perception de l'agrément, alors que dans un cadre non-idéal, il impactera négativement le ressenti GL : chiffre.

Pour finir, nous considérons un dernier groupe de descripteurs, nommément $L_m - L_b$, $L(E)_m - L(E)_b$ et $L(T)_m - L(T)_b$ (Cf. Figures 34g, 34h, 34i, 34j, 34k et 34l). Ces descripteurs expriment la différence entre les niveaux des marqueurs, et ceux des autres classes de sons. Ils traduisent l'émergence des marqueurs par rapport à la mixture sonore.

Pour les i-scènes, une corrélation forte et positive est observée pour $L_m - L_b$ ($r = 0.67$, $p < 0.01$) et $L(E)_m - L(E)_b$ ($r = 0.66$, $p < 0.01$). Pour les ni-scènes, seule une légère corrélation positive est observée pour $L_m - L_b$ ($r = -0.39$, $p < 0.05$). Dans le cas des i-scènes, ce n'est donc pas le niveau absolu des marqueurs qui importe, mais leur niveau relatif, par rapport aux autres sons qui composent la scène. On observe donc pour les environnements idéaux un double mécanisme perceptif :

- Plus le niveau absolu des sons n'étant pas des i-marqueurs est élevé, plus l'agrément est faible
- Plus le niveau relatif des i-marqueurs, par rapport aux autres sons, est élevé, plus l'agrément est élevé

Pour les ni-scènes, le fait que nous observions des corrélations pour L_m et $L(E)_m$, et aucune pour $L_m - L_b$ et $L(E)_m - L(E)_b$, montre que c'est bien le niveau absolu qui importe.

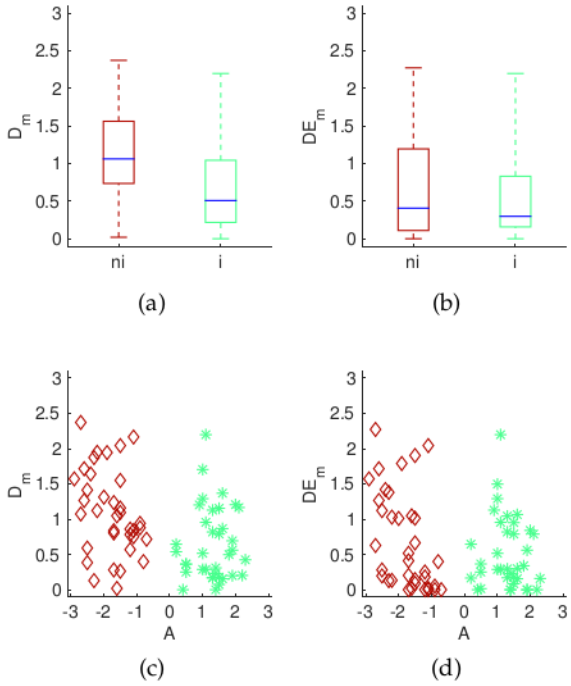


FIGURE 33 : TODO

5.2.14 *Discussions*

Introduire pour faire comprendre que c'est un bilan

Nous identifions 6 indicateurs structurels globaux permettant de distinguer, de manière globale, les environnements sonore idéaux et non-idéaux.

- niveau sonore : calculé sur tous les sons L, les événements L(E) et les textures L(T)
- densité : calculé de manière globale (D) et sur les événements D(E).
- diversité : calculé uniquement sur les événements E

Parmi, ces indicateurs structurels, seuls L et LE permettent de prédire l'agrément. Nous notons cependant que cette prédiction ne vaut que pour les ni-scènes.

Nous observons qu'une description sémantique des scènes, basée sur la présence/absence des classes de sons, permet de bien prédire la nature de l'environnement. Par ailleurs, il apparaît qu'il est possible d'obtenir une prédiction similaire, voire meilleure, en ne considérant qu'un sous groupe de classes d'événements,i. e. les marqueurs sonores.

Parmi les descripteurs structurels spécifiques, calculés en tenant compte des marqueurs sonores, plusieurs permettent maintenant de faire la distinction entre les i-scènes et ni-scènes :

- X

Parmi ces descripteurs, 5 sont capables de prédire l'agrément de manière fine :

- $L(E)_m - L(E)_b$ et $L_m - L_b$ pour les i-scènes
-

De cette analyse, nous retenons les points suivant :

- *Distinguer les i- et ni-scènes* : Les descripteurs sémantiques, ainsi que certains descripteurs structurels globaux, permettent de faire la distinction entre les i-scènes et les ni-scènes. La description sémantique semble être plu performante.
- *Événements ou textures* : Que ce soit pour les descripteurs sémantiques ou structurels, c'est majoritairement les événements qui permettent de distinguer les deux types d'environnements, les textures n'apportant au mieux qu'une information limitée.
- *Prédire l'agrément* : Si l'on considère une description fine de l'agrément, il semble que la manière de percevoir la qualité de l'environnement diffère en fonction de la nature de ce dernier (i ou ni). Il n'apparaît pas envisageable de considérer un même jeu de descripteurs pour prédire à la fois l'agrément des i-scènes et l'agrément des ni-scènes. Pour les ni-scènes, c'est le niveau global (L et $L(E)$), la densité globale (D et $D(E)$), ou encore le niveau des marqueurs sonores (L_m et L), qui impactent négativement l'agrément. On note ici que prendre en compte les contributions de différentes sources n'améliore pas la capacité de prédiction de l'agrément, par rapport à une analyse holistique de l'environnement. Pour les i-scènes, par contre, prédire l'agrément requiert d'étudier de manière séparée les caractéristiques des marqueurs sonores et du reste des sons. Ainsi le niveau des marqueurs relatifs au bruit est positivement corrélé à l'agrément, alors que le niveau du bruit est, lui, négativement corrélé.

L'existence de deux modes de perceptions, mobilisant différents types de descripteurs, et dépendant de la nature du stimuli, est un phénomène qui a déjà été observé pour la perception des textures (cf. section 3.5). Le cerveau adapte sa manière de traiter l'information (résumé statistique pour les textures, description fine pour les événements) suite à une prise de décision antérieure quant à la nature du stimuli (à savoir "est-ce un événement ou une texture ?"). De la même manière, les indicateurs actifs dans le jugement de l'agrément dépendent eux aussi d'une identification préalable de la nature hédonique globale de l'environnement (idéale ou non idéale).

Ces résultats peuvent potentiellement influer sur les stratégies à adopter pour améliorer la qualité de l'environnement sonore :

- Dans le cadre de scènes non-idéales, il s'agit de diminuer le niveau sonore, soit de manière globale, soit en agissant sur certaines sources (*sirène*, *klaxon*).
- Dans le cadre de scènes idéales, il s'agit 1) d'identifier les sons agréables, i. e. les marqueurs sonores, 2) de baisser le niveau des autres sons, 3) voire, en restant dans la limite du raisonnable, d'augmenter le niveau des marqueurs par rapport aux autres sons.

Nous montrons que les descripteurs à utiliser dépendent de la nature de l'environnement, et que cette nature est elle même dépendante de la composition sémantique, i. e. des sources sonores présentes. Dans une certaine mesure, nous pouvons donc dire que les descripteurs dépendent des sources sonores présentes. Mais nous observons également que le type de descripteurs à utiliser pour une même source varie en fonction de la nature de l'environnement GL : TODO : développer sur le contexte environnemental pour l'agrément, reprendre l'exemple de *foule*.

GL : TODO : Reprendre la conclusion de l'article et Proposer un modèle perceptif sur la base du modèle prédictif GL : TODO, l'utilisation de trafic reprendre les conclusions de (Lavandier and Defréville, 2006) GL : TODO, les résultats (2 modes d'obs) concordent avec les observations faites pas (Ricciardi et al., 2015) (cf. section **??**).

5.3 agir sur l'agrément perçu en modifiant la composition sémantique

5.3.1 *Objectif de l'expérience*

5.3.2 *Planification expérimentale*

5.3.3 *Données et méthodes d'analyses*

5.3.4 *Discussions*

5.4 l'impact de la composition sur les processus de catégorisation des scènes

5.4.1 *Objectif de l'expérience*

5.4.2 *Planification expérimentale*

5.4.3 *Données et méthodes d'analyses*

5.4.4 *Discussions*

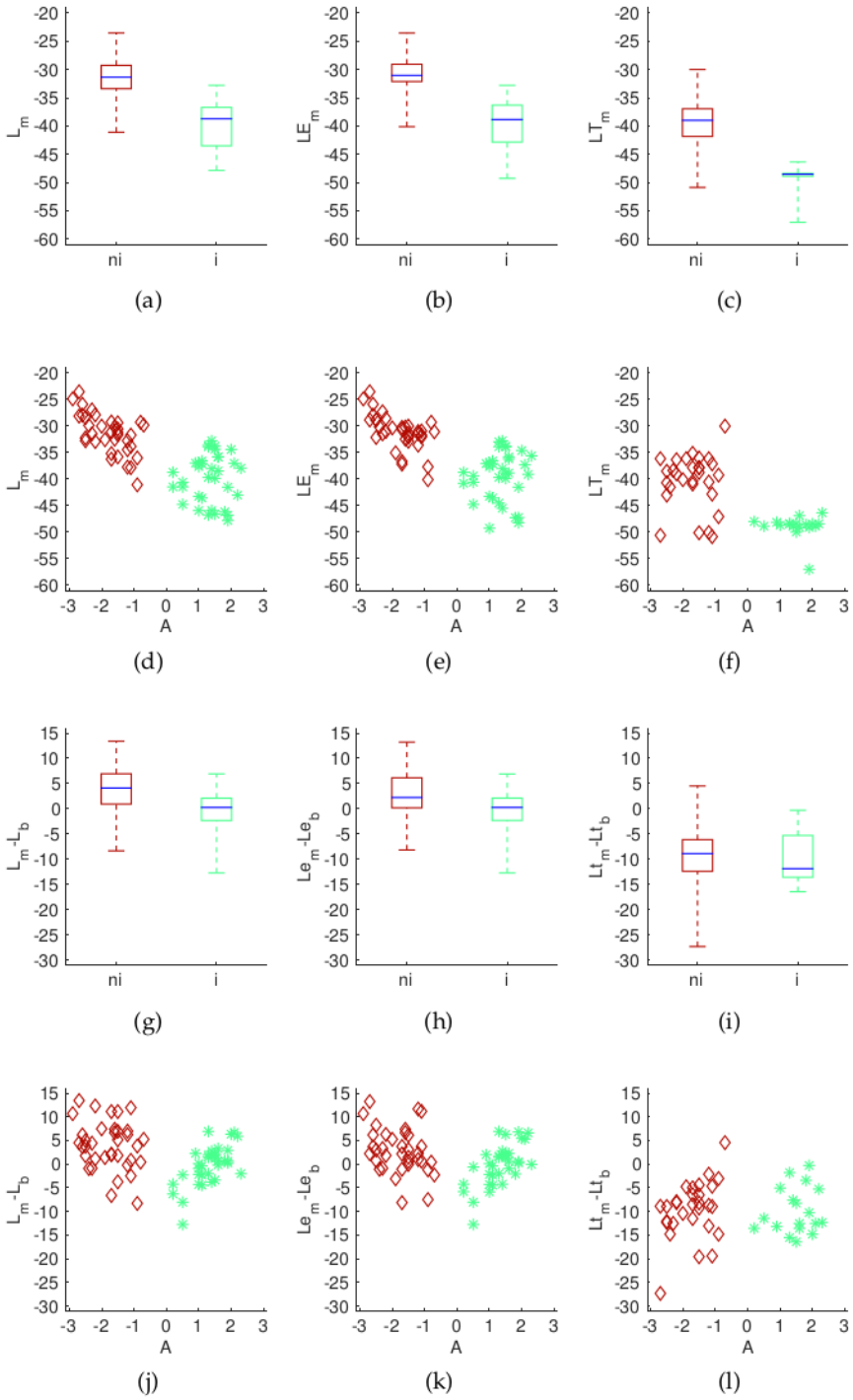


FIGURE 34 : TODO

Troisième partie

ANALYSE AUTOMATIQUE

preamble text here.

6

L'ANALYSE AUTOMATIQUE DES SCÈNES SONORES ENVIRONNEMENTALES, UN ÉTAT DE L'ART

[2 septembre 2016 at 4:43 – classicthesis version 4.2]

7

APPLICATION DU MODÈLE MORPHOLOGIQUE À L'ÉTUDE DES PERFORMANCES DES ALGORITHMES D'ANALYSE AUTOMATIQUE DES SCÈNES SONORES ENVIRONNEMENTALES

7.1 DE L'UTILISATION DE SCÈNES SIMULÉES EN ANALYSE AUTOMATIQUE

8

APPLICATION DU MODÈLE MORPHOLOGIQUE À L'ANALYSE AUTOMATIQUE DES ENVIRONNEMENTS SONORES

Quatrième partie

CONCLUSIONS ET PERSPECTIVES

preamble text here.

9 CONCLUSIONS

10 PERSPECTIVES

Cinquième partie

APPENDICE

A

OUTILS D'ANALYSE STATISTIQUE UNI-VARIÉE

http://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/

A.1 TEST PARAMÉTRIQUES À DEUX POPULATIONS

A.2 TEST PARAMÉTRIQUES À PLUS DE DEUX POPULATIONS

A.3 MESURES DE CORRÉLATION PARAMÉTRIQUE

A.4 RÉGRESSION LINÉAIRE MULTIPLE

B

OUTILS D'ANALYSE DIMENSIONNELLE

B.1 ANALYSE DISCRIMINANTE

B.2 ANALYSE PAR COMPOSANTE PRINCIPALE

B.3 POSITIONNEMENT MULTIDIMENSIONNEL

C

SÉQUENCE D'ÉVÉNEMENTS OU TEXTURE SONORE : L'INFLUENCE DE LA PÉRIODE D'ATTENTION.

C.1 LA THÉORIE DE LA DÉTECTION DU SIGNAL

C.2 OBJECTIF DE L'EXPÉRIENCE

C.3 PLANIFICATION EXPÉRIMENTALE

C.4 RÉSULTATS

BIBLIOGRAPHY

Adams, Mags D, Neil S Bruce, William J Davies, Rebecca Cain, Paul Jennings, Angus Carlyle, Peter Cusack, Ken Hume, and C Plack (2008). "Soundwalking as a methodology for understanding soundscapes." In: *Proceedings of the Institute of Acoustics*. Vol. 30. 2.

Agus, Trevor R, Simon J Thorpe, and Daniel Pressnitzer (2010). "Rapid formation of robust auditory memories : insights from noise." In: *Neuron* 66.4, pp. 610–618.

Aletta, Francesco, Jian Kang, and Östen Axelsson (2016). "Soundscape descriptors and a conceptual framework for developing predictive soundscape models." In: *Landscape and Urban Planning* 149, pp. 65–74.

Anderson, John R (1991). "The adaptive nature of human categorization." In: *Psychological Review* 98.3, p. 409.

Axelsson, Osten, Birgitta Berglund, and Mats E Nilsson (2005). "Soundscape assessment." In: *The Journal of the Acoustical Society of America* 117.4, pp. 2591–2592.

Axelsson, Östen, Mats E Nilsson, and Birgitta Berglund (2010). "A principal components model of soundscape perceptiona)." In: *The Journal of the Acoustical Society of America* 128.5, pp. 2836–2846.

Ballas, James A and James H Howard (1987). "Interpreting the language of environmental sounds." In: *Environment and behavior* 19.1, pp. 91–114.

Barsalou, Lawrence W (1999). "Perceptions of perceptual symbols." In: *Behavioral and brain sciences* 22.04, pp. 637–660.

— (2010). "Grounded cognition : Past, present, and future." In: *Topics in cognitive science* 2.4, pp. 716–724.

Beaumont, Jacques, Stéphen Lesaux, Benjamin Robin, Jean-Dominique Polack, Cristina Pronello, Christine Arras, and Laurent Droin (2004). "Pertinence des descripteurs d'ambiance sonore urbaine." In: *Acoustique et techniques*.

Bendavid, R. and M. Chasles-Parot (2014). *Les Français et les Nuisances Sonores (French and Noise Nuisances)*. Tech. rep. Paris, France: Institut français d'opinion publique (IFOP), p. 24.

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer (2011). "D3 : Data-Driven Documents." In: *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

Botteldooren, Dick, Bert De Coensel, and Tom De Muer (2006). "The temporal structure of urban soundscapes." In: *Journal of sound and vibration* 292.1, pp. 105–123.

Bregman, Albert S (1994). *Auditory scene analysis : The perceptual organization of sound*. MIT press.

Brocolini, Laurent, Catherine Lavandier, Catherine Marquis-Favre, Mathias Quoy, and Mathieu Lavandier (2012). "Prediction and explanation of sound quality indicators by multiple linear regressions and artificial neural networks." In: *Proc. IOA/CFA congress, Acoustics*.

Brown, AL, Jian Kang, and Truls Gjestland (2011). "Towards standardization in soundscape preference assessment." In: *Applied Acoustics* 72.6, pp. 387–392.

Bruce, Neil S and William J Davies (2014). "The effects of expectation on the perception of soundscapes." In: *Applied Acoustics* 85, pp. 1–11.

Bruce, Neil S, William J Davies, and Mags D Adams (2009). "Development of a soundscape simulator tool." In: *proceedings of Internoise 2009*.

Cain, Rebecca, Paul Jennings, and John Poxon (2013). "The development and application of the emotional dimensions of a soundscape." In: *Applied Acoustics* 74.2, pp. 232–239.

Carlyon, Robert P (2004). "How the brain separates sounds." In: *Trends in cognitive sciences* 8.10, pp. 465–471.

Carlyon, Robert P, John Deeks, Dennis Norris, and Sally Butterfield (2002). "The continuity illusion and vowel identification." In: *Acta Acustica United with Acustica* 88.3, pp. 408–415.

Cusack, Rhodri, John Decks, Genevieve Aikman, and Robert P Carlyon (2004). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis." In: *Journal of Experimental Psychology : Human Perception and Performance* 30.4, p. 643.

Dannenbring, Gary L (1976). "Perceived auditory continuity with alternately rising and falling frequency transitions." In: *Canadian Journal of Psychology* 30.2, p. 99.

Davies, WJ et al. (2009). "The positive soundscape project : a synthesis of results from many disciplines." In: *Internoise 2009*.

Davies, William J, Mags D Adams, Neil S Bruce, Rebecca Cain, Angus Carlyle, Peter Cusack, Deborah A Hall, Ken I Hume, Amy Irwin, Paul Jennings, et al. (2013). "Perception of soundscapes : An interdisciplinary approach." In: *Applied acoustics* 74.2, pp. 224–231.

Davies, William J, Neil S Bruce, and Jesse E Murphy (2014). "Soundscape reproduction and synthesis." In: *Acta Acustica United with Acustica* 100.2, pp. 285–292.

De Coensel, Bert and Dick Botteldooren (2006). "The quiet rural soundscape and how to characterize it." In: *Acta Acustica united with Acustica* 92.6, pp. 887–897.

De Coensel, Bert, Michiel Boes, Damiano Oldoni, and Dick Botteldooren (2013). "Characterizing the soundscape of tranquil urban spaces." In: *Proceedings of Meetings on Acoustics*. Vol. 19. 1. Acoustical Society of America, p. 040052.

Defréville, Boris, Catherine Lavandier, and Marc Laniray (2004). "Activity of urban sound sources." In: *Proceedings of the 18th International Congress in Acoustics*. Kyoto.

Delaitre, Pauline, Catherine Lavandier, Caroline Cance, and Jean Pruvost (2012). "What is the Definition for the French Word calme in the European Directive Related to "Quiet Areas"? A Lexicographic Study from the 16th Century Until Today." In: *Acta Acustica united with Acustica* 98.5, pp. 734–740.

Devergie, Aymeric (2006). "Relations entre Perception Globale et Composition de S´equences Sonores." MA thesis. IRCAM, Paris VI UPMC.

Dubois, Danièle (1991). *Sémantique et cognition : catégories, prototypes, typicalité*. Diffusion, Presses du CNRS.

— (2000). "Categories as acts of meaning : The case of categories in olfaction and audition." In: *Cognitive science quarterly* 1.1, pp. 35–68.

Dubois, Danièle, Catherine Guastavino, and Manon Raimbault (2006). "A cognitive approach to urban soundscapes : Using verbal data to access everyday life auditory categories." In: *Acta acustica united with acustica* 92.6, pp. 865–874.

Finney, Nathaniel and Jordi Janer (2010). "Soundscape generation for virtual environments using community-provided audio databases." In: *W3C Workshop : Augmented Reality on the Web*.

Fried, Lisbeth S and Keith J Holyoak (1984). "Induction of category distributions : A framework for classification learning." In: *Journal of Experimental Psychology : Learning, Memory, and Cognition* 10.2, pp. 234–257.

Galbrun, Laurent and Tahrir Ali (2012). "Perceptual assessment of water sounds for road traffic noise masking." In: *Proceedings of Meetings on Acoustics*. Acoustical Society of America.

García Pérez, Igone, Itziar Aspuru Soloaga, Karmele Herranz-Pascual, and Ibone García-Borreguero (2012). "Validation of an indicator for the assessment of the environmental sound in urban places." In: *EURONOISE*. Prague.

Gaver, William W (1993a). "How do we hear in the world? Explorations in ecological acoustics." In: *Ecological psychology* 5.4, pp. 285–313.

— (1993b). "What in the world do we hear? : An ecological approach to auditory event perception." In: *Ecological psychology* 5.1, pp. 1–29.

Gibson, James J (1978). "The ecological approach to the visual perception of pictures." In: *Leonardo* 11.3, pp. 227–235.

Gibson, James Jerome (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gille, Laure-Anne and Catherine Marquis-Favre (2016). "Dose-effect relationships for annoyance due to road traffic noise : Multi-level

regression and consideration of noise sensitivity." In: *The Journal of the Acoustical Society of America* 139.4, pp. 2070–2070.

Gille, Laure-Anne, Catherine Marquis-Favre, and Achim Klein (2016a). "Noise Annoyance Due To Urban Road Traffic with Powered-Two-Wheelers : Quiet Periods, Order and Number of Vehicles." In: *Acta Acustica united with Acustica* 102.3, pp. 474–487.

Gille, Laure-Anne, Catherine Marquis-Favre, and Julien Morel (2016b). "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." In: *Environment International* 94, pp. 83–94.

Goldstone, Robert L and Lawrence W Barsalou (1998). "Reuniting perception and conception." In: *Cognition* 65.2, pp. 231–262.

Goldstone, Robert L and Alan Kersten (2003). "Concepts and categorization." In: *Handbook of psychology*.

Guastavino, Catherine (2003). "Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain, (*Semantic and acoustic study of lowfrequency noises perception in urban sound environment*)." PhD thesis. Paris, France: Université Paris VI.

— (2006). "The ideal urban soundscape : Investigating the sound quality of French cities." In: *Acta Acustica united with Acustica* 92.6, pp. 945–951.

— (2007). "Categorization of environmental sounds." In: *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 61.1, p. 54.

Guastavino, Catherine and Pascale Cheminée (2003). "Une approche psycholinguistique de la perception des basses fréquences : Conceptualisations en langue, représentations cognitives et validité écologique." In: *Psychologie française* 48.4, pp. 91–101.

Guastavino, Catherine and Brian FG Katz (2004). "Perceptual evaluation of multi-dimensional spatial audio reproduction." In: *The Journal of the Acoustical Society of America* 116.2, pp. 1105–1115.

Guastavino, Catherine, Brian FG Katz, Jean-Dominique Polack, Daniel J Levitin, and Daniele Dubois (2005). "Ecological validity of soundscape reproduction." In: *Acta Acustica united with Acustica* 91.2, pp. 333–341.

Guillén, José Domingo and Isabel López Barrio (2007). "Importance of personal, attitudinal and contextual variables in the assessment of pleasantness of the urban sound environment." In: *Proceedings of 19 th INTERNATIONAL CONGRESS ON ACOUSTICS*. Madrid.

Guyot, F., M. Castellengo, and B. Fabre (1997). "Catégorisation et Cognition : De la Perception au Discours." In: Paris, France: Édition Kimé. Chap. A study of the categorization of an everyday sound set, pp. 41–58.

Gygi, Brian and Valeriy Shafiro (2011). "The incongruency advantage for environmental sounds presented in natural auditory scenes." In: *Journal of Experimental Psychology : Human Perception and Performance* 37.2, p. 551.

Gygi, Brian, Gary R Kidd, and Charles S Watson (2007). "Similarity and categorization of environmental sounds." In: *Perception & psychophysics* 69.6, pp. 839–855.

Hall, Deborah A, Amy Irwin, Mark Edmondson-Jones, Scott Phillips, and John EW Poxon (2013). "An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes." In: *Applied Acoustics* 74.2, pp. 248–254.

Hong, Joo Young and Jin Yong Jeon (2013). "Designing sound and visual components for enhancement of urban soundscapes." In: *The Journal of the Acoustical Society of America* 134.3, pp. 2026–2036.

Houdé, Olivier, Daniel Kayser, Olivier Koenig, Joëlle Proust, and François Rastier (1998). *Vocabulaire de sciences cognitives*. Paris : Presses Universitaires de France.

Houix, Olivier, Guillaume Lemaitre, Nicolas Misdariis, Patrick Susini, and Isabel Urdapilleta (2012). "A lexical analysis of environmental sound categories." In: *Journal of Experimental Psychology : Applied* 18.1, pp. 52–80. (Visited on 04/09/2013).

Hume, Ken and Mujthaba Ahtamad (2013). "Physiological responses to and subjective estimates of soundscape elements." In: *Applied Acoustics* 74.2, pp. 275–281.

Jeon, Jin Yong, Pyoung Jik Lee, Joo Young Hong, and Densil Cabrera (2011). "Non-auditory factors affecting urban soundscape evaluation." In: *The Journal of the Acoustical Society of America* 130.6, pp. 3761–3770.

Jeon, Jin Yong, Joo Young Hong, and Pyoung Jik Lee (2013). "Soundwalk approach to identify urban soundscapes individually." In: *The Journal of the Acoustical Society of America* 134.1, pp. 803–812.

Kang, Jian (2006). *Urban sound environment*. CRC Press.

Kang, Jian and M Zhang (2010). "Semantic differential analysis of the soundscape in urban open public spaces." In: *Building and environment* 45.1, pp. 150–157.

Kardous, Chucri A and Peter B Shaw (2014). "Evaluation of smartphone sound measurement applicationsa)." In: *The Journal of the Acoustical Society of America* 135.4, EL186–EL192.

Klein, Achim, Catherine Marquis-Favre, Rheinard Weber, and Arnaud Trollé (2015). "Spectral and modulation indices for annoyance-relevant features of urban road single-vehicle pass-by noises." In: *The Journal of the Acoustical Society of America* 137.3, pp. 1238–1250.

Kuwano, Sonoko, Seiichiro Namba, Tohru Kato, and Jürgen Hellbrück (2003). "Memory of the loudness of sounds in relation to overall impression." In: *Acoustics Science and Technics* 4.24.

Lafay, Grégoire (2013). "Caractérisation sémantique des scènes sonores environnementales : Étude paramétrique et perceptive d'un paradigme de synthèse séquentielle par corpus." MA thesis. IRCAM, Paris VI UPMC.

Lafay, Grégoire, Mathias Rossignol, Nicolas Misdariis, Mathieu Lagrange, and Jean-François Petiot (2014). "A new experimental approach for urban soundscape characterization based on sound manipulation : A pilot study." In: *International Symposium on Musical Acoustics*.

Lafay, Grégoire, Nicolas Misdariis, Mathieu Lagrange, and Mathias Rossignol (2016). "Semantic browsing of sound databases without keywords." In: *Journal of the Audio Engineering Society*.

Lavandier, Catherine and Boris Defréville (2006). "The contribution of sound source characteristics in the assessment of urban soundscapes." In: *Acta Acustica united with Acustica* 92.6, pp. 912–921.

Lecointre, Guillaume and Hervé Le Guyader (2006). *The tree of life : a phylogenetic classification*. Vol. 20. Harvard University Press.

Lemaitre, Guillaume, Olivier Houix, Nicolas Misdariis, and Patrick Susini (2010). "Listener expertise and sound identification influence the categorization of environmental sounds." In: *Journal of Experimental Psychology : Applied* 16.1, p. 16.

Leobon, A. (1986). "Analyse psycho-acoustique du paysage sonore urbain, (*Psychoacoustic analysis of urban soundscape*)." PhD thesis. Strasbourg, France: Université Louis Pasteur.

Maffiolo, Valérie (1997). *Méthodes d'approche de l'environnement sonore urbain*. Tech. rep. Paris, France: Mairie de Paris, Direction de la protection de l'environnement, SPAAS.

— (1999). "De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain, (*Semantic and acoustical characterisation of the sound quality of urban environment*)." PhD thesis. Le Mans, France: Université du Mans.

Marcell, Michael M, Diane Borella, Michael Greene, Elizabeth Kerr, and Summer Rogers (2000). "Confrontation naming of environmental sounds." In: *Journal of clinical and experimental neuropsychology* 22.6, pp. 830–864.

Marquis-Favre, C, E Premat, D Aubrée, and M Vallet (2005a). "Noise and its effects : A review on qualitative aspects of sound. Part I : Notions and acoustic ratings." In: *Acta acustica united with acustica* 91.4, pp. 613–625.

Marquis-Favre, C, E Premat, and D Aubrée (2005b). "Noise and its effects : A review on qualitative aspects of sound. Part II : Noise and annoyance." In: *Acta acustica united with acustica* 91.4, pp. 626–642.

Marquis-Favre, Catherine and Julien Morel (2015). "A simulated environment experiment on annoyance due to combined road traffic

and industrial noises." In: *International journal of environmental research and public health* 12.7, pp. 8413–8433.

McAdams, Stephen and Emmanuel Bigand (1994). *Penser les sons : psychologie cognitive de l'audition*. Presses Univ. de France.

McDermott, Josh H and Eero P Simoncelli (2011). "Sound texture perception via statistics of the auditory periphery : evidence from sound synthesis." In: *Neuron* 71.5, pp. 926–940.

McDermott, Josh H, Michael Schemitsch, and Eero P Simoncelli (2013). "Summary statistics in auditory perception." In: *Nature neuroscience* 16.4, pp. 493–498.

Medin, Douglas L and Marguerite M Schaffer (1978). "Context theory of classification learning." In: *Psychological review* 85.3, p. 207.

Memoli, Gianluca, Alan Bloomfield, and Max Dixon (2008). "Soundscape characterization in selected areas of Central London." In: *Proceedings of Meetings on Acoustics*. Paris.

Meng, Qi, Jian Kang, and Hong Jin (2013). "Field study on the influence of spatial and environmental characteristics on the evaluation of subjective loudness and acoustic comfort in underground shopping streets." In: *Applied Acoustics* 74.8, pp. 1001–1009.

Misra, Ananya, Perry R Cook, and Ge Wang (2006). "A new paradigm for sound design." In: *Proceedings of the International Conference on Digital Audio Effects (DAFx-06)*. Citeseer, pp. 319–324.

Misra, Ananya, Ge Wang, and Perry Cook (2007). "Musical Tapestry : Re-composing Natural Sound." In: *Journal of New Music Research* 36.4, pp. 241–250.

Moore, Brian CJ (1973). "Frequency difference limens for short-duration tones." In: *The Journal of the Acoustical Society of America* 54.3, pp. 610–619.

Morel, Julien, Catherine Marquis-Favre, and L-A Gille (2016). "Noise annoyance assessment of various urban road vehicle pass-by noises in isolation and combined with industrial noise : A laboratory study." In: *Applied Acoustics* 101, pp. 47–57.

Neisser, Ulric (1967). *Cognitive psychology*. (Reprinted as *Cognitive psychology : Classic edition*. Psychology Press, 2014). New York: Appleton-Century-Crofts.

— (1976). *Cognition and reality : principles and implications of cognitive psychology*. WH Freeman.

Nelken, Israel and Alain de Cheveigné (2013). "An ear for statistics." In: *Nature neuroscience* 16.4, pp. 381–382.

Nielbo, Frederik L, Daniel Steele, and Catherine Guastavino (2013). "Investigating soundscape affordances through activity appropriateness." In: *Proceedings of Meetings on Acoustics*. Vol. 19. 1. Acoustical Society of America, p. 040059.

Niessen, Maria E, Leendert Van Maanen, and Tjeerd C Andringa (2008). "Disambiguating sound through context." In: *International Journal of Semantic Computing* 2.03, pp. 327–341.

Niessen, Maria, Caroline Cance, and Danile Dubois (2010). "Categories for soundscape : toward a hybrid classification." In: *Inter-Noise and Noise-Con Congress and Conference Proceedings*. Vol. 2010. 5. Institute of Noise Control Engineering, pp. 5816–5829.

Nilsson, M, Dick Botteldooren, and Bert De Coensel (2007). "Acoustic indicators of soundscape quality and noise annoyance in outdoor urban areas." In: *Proceedings of the 19th International Congress on Acoustics*.

Nilsson, Mats E (2007). "Soundscape quality in urban open spaces." In: *Inter-Noise*. Istanbul, Turkey.

Nilsson, Mats E and Birgitta Berglund (2006). "Soundscape quality in suburban green areas and city parks." In: *Acta Acustica united with Acustica* 92.6, pp. 903–911.

Ozcevik, Asli and Zerhan Yuksel Can (2012). "A laboratory study on the evaluation of soundscape." In: *Proceedings of Meetings on Acoustics*. Acoustical Society of America.

Payne, Sarah R (2013). "The production of a perceived restorativeness soundscape scale." In: *Applied Acoustics* 74.2, pp. 255–263.

Pheasant, RJ, GR Watts, and KV Horoshenkov (2009). "Validation of a tranquillity rating prediction tool." In: *Acta Acustica united with Acustica* 95.6, pp. 1024–1031.

Pheasant, Robert, Kirill Horoshenkov, Greg Watts, and Brendan Barrett (2008). "The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places ?" In: *The Journal of the Acoustical Society of America* 123.3, pp. 1446–1457.

Poeppel, David (2003). "The analysis of speech in different temporal integration windows : cerebral lateralization as asymmetric sampling in time." In: *Speech communication* 41.1, pp. 245–255.

Polack, Jean-Dominique, Jacuqes Beaumont, Christine Arras, Mikael Zekri, and Benjamin Robin (2008). "Perceptive relevance of soundscape descriptors : a morpho-typological approach." In: *Journal of the Acoustical Society of America* 123.5, p. 3810.

Raimbault, Manon (2002). "Simulation des ambiances sonores urbaines : intégration des aspects qualitatifs,*Urban soundscape simulation : focusing on qualitative aspect*)." PhD thesis. Nantes, France: Unversité de Nantes - Ecole polytechnique de Nantes.

— (2006). "Qualitative judgements of urban soundscapes : Questionning questionnaires and semantic scales." In: *Acta acustica united with acustica* 92.6, pp. 929–937.

Raimbault, Manon and Daniele Dubois (2005). "Urban soundscapes : Experiences and knowledge." In: *Cities* 22.5, pp. 339–350.

Reed, Stephen K (1972). "Pattern recognition and categorization." In: *Cognitive psychology* 3.3, pp. 382–407.

Ribeiro, Carlos, Celine Anselme, Fanny Mietlicki, Bruno Vincent, Raphael Da Silva, and Piotr Gaudibert (2013). "At the heart of Har-

monica project : the Common Noise Index (CNI)." In: *Proceedings of 42nd International Congress on Noise Control Engineering, Internoise,(2013, Innsbruck, Austria)*.

Ricciardi, Paola, Pauline Delaitre, Catherine Lavandier, Francesca Torchia, and Pierre Aumond (2015). "Sound quality indicators for urban places in Paris cross-validated by Milan data." In: *The Journal of the Acoustical Society of America* 138.4, pp. 2337–2348.

Rosch, Eleanor (1975). "Cognitive representations of semantic categories." In: *Journal of experimental psychology : General* 104.3, p. 192.

Rosch, Eleanor and Barbara B Lloyd (1974). *Huamn communication : Theoretical perspectives*. Halsted Press, New York.

— (1978). *Cognition and categorization*. Hillsdale, New Jersey.

Rosch, Eleanor and C B Mervis (1975). "Family resemblances : Studies in the internal structure of categories." In: *Cognitive Psychology* 7, pp. 573–605.

Rosch, Eleanor, Carol Simpson, and R Scott Miller (1976). "Structural bases of typicality effects." In: *Journal of Experimental Psychology : Human perception and performance* 2.4, p. 491.

Rossignol, Mathias, Grégoire Lafay, Mathieu Lagrange, and Nicolas Misdariis (2015). "SimScene : a web-based acoustic scenes simulator." In: *1st Web Audio Conference (WAC)*.

Rychtáriková, Monika and Gerrit Vermeir (2013). "Soundscape categorization on the basis of objective acoustical parameters." In: *Applied Acoustics* 74.2, pp. 240–247.

Saint-Arnaud, Nicolas (1995). "Classification of sound textures." MA thesis. Massachusetts Institute of Technology.

Salamon, J., C. Jacoby, and J. P. Bello (2014). "A Dataset and Taxonomy for Urban Sound Research." In: *22st ACM International Conference on Multimedia (ACM-MM'14)*. Orlando, FL, USA.

Schafer, R.M. (1969). *The New Soudscape : A Handbook for the Modern Music Teacher*. Ontario : Berandol Music Limited.

— (1977). *The Tuning of the World*. Borzoi book. (Reprinted as *Our Sonic Environment and the Soundscape : The Tuning of the World*. Destiny Books, 1994). New York: Knopf.

Schirosa, Mattia, Jordi Janer, Stefan Kersten, and Gerard Roma (2010). "A system for soundscape generation, composition and streaming." In: *XVII CIM-Colloquium of Musical Informatics*.

Schulte-Fortkamp, Brigitte (2013). "Soundscape-focusing on resources." In: *Proceedings of Meetings on Acoustics*. Vol. 19. 1. Acoustical Society of America, p. 040117.

Schulte-Fortkamp, Brigitte and Andre Fiebig (2006). "Soundscape analysis in a residential area : An evaluation of noise and people's mind." In: *Acta Acustica united with Acustica* 92.6, pp. 875–880.

Schulte-Fortkamp, Brigitte and Jian Kang (2010). "Soundscape research in networking across countries : COST Action TD0804." In: *The Journal of the Acoustical Society of America* 127.3, pp. 1801–1801.

Schulte-Fortkamp, Brigitte, Bennett M Brooks, and Wade R Bray (2007). "Soundscape : An Approach to Rely on Human Perception and Expertise in the Post-Modern Community Noise Era." In: *Acoustics Today* 3.1, pp. 7–15.

Schwartz, Jean-Luc, Nicolas Grimault, Jean-Michel Hupé, Brian CJ Moore, and Daniel Pressnitzer (2012). "Multistability in perception : binding sensory modalities an overview." In: *Philosophical Transactions of the Royal Society of London B : Biological Sciences* 367.1591, pp. 896–905.

Schwarz, Diemo (2011). "State of the art in sound texture synthesis." In: *Digital Audio Effects (DAFx)*, pp. 1–1.

Snyder, Joel S and Claude Alain (2007). "Toward a neurophysiological theory of auditory stream segregation." In: *Psychological bulletin* 133.5, p. 780.

Southworth, Michael (1969). "The sonic environment of cities." In: *Environment and behavior* 1.1, p. 49.

Standardization, International Organization for (2013). *ISO 12913-1 :2014 acoustics-soundscape-part 1 : definition and conceptual framework*. Tech. rep. Genève: ISO.

Szeremeta, Bani and Paulo Henrique Trombetta Zannin (2009). "Analysis and evaluation of soundscapes in public parks through interviews and measurement of noise." In: *Science of the total environment* 407.24, pp. 6143–6149.

Tae Hong Park, Johnathan Turner, Michael Musick, Jun Hee Lee, Christopher Jacoby, Charlie Mydlarz, and Justin Salamon (2014). "Sensing Urban Soundscapes." In: *EDBT/ICDT Workshops*, pp. 375–382.

Torija, Antonio J, Diego P Ruiz, and AF Ramos-Ridao (2013). "Application of a methodology for categorizing and differentiating urban soundscapes using acoustical descriptors and semantic-differential attributes." In: *The Journal of the Acoustical Society of America* 134.1, pp. 791–802.

Trollé, Arnaud, Catherine Marquis-Favre, and Étienne Parizet (2015). "Perception and Annoyance Due to Vibrations in Dwellings Generated From Ground Transportation : A Review." In: *Journal of Low Frequency Noise, Vibration and Active Control* 34.4, pp. 413–457.

Truax, Barry (1978). *Handbook for acoustic ecology*. (originally published by the world soundscape project). simon fraser university and ARC Publications.

Tse, Man Sze, Chi Kwan Chau, Yat Sze Choy, Wai Keung Tsui, Chak Ngai Chan, and Shiu Keung Tang (2012). "Perception of urban park soundscape." In: *The Journal of the Acoustical Society of America* 131.4, pp. 2762–2771.

Tversky, Amos (1977). "Features of similarity." In: *Psychological review* 84.4, p. 327.

Valle, Andrea, Mattia Schirosa, and Vincenzo Lombardo (2009). "A framework for soundscape analysis and re-synthesis." In: *Proceedings of the SMC*, pp. 13–18.

Winkler, Istvan, Susan L Denham, and Israel Nelken (2009). "Modeling the auditory scene : predictive regularity representations and perceptual objects." In: *Trends in cognitive sciences* 13.12, pp. 532–540.

Yabe, Hirooki, Mari Tervaniemi, Janne Sinkkonen, Minna Huotilainen, Risto J Ilmoniemi, and Risto Näätänen (1998). "Temporal window of integration of auditory information in the human brain." In: *Psychophysiology* 35.5, pp. 615–619.

Yang, Ming and Jian Kang (2013). "Psychoacoustical evaluation of natural and urban sounds in soundscapes." In: *The Journal of the Acoustical Society of America* 134.1, pp. 840–851.

Yang, Wei and Jian Kang (2005). "Acoustic comfort evaluation in urban open public spaces." In: *Applied acoustics* 66.2, pp. 211–229.

Yost, William A (1994). *Fundamentals of hearing : An introduction*. Academic Press.

Yu, Lei and Jian Kang (2009). "Modeling subjective evaluation of soundscape quality in urban open spaces : An artificial neural network approach." In: *The Journal of the Acoustical Society of America* 126.3, pp. 1163–1174.

— (2010). "Factors influencing the sound preference in urban open spaces." In: *Applied Acoustics* 71.7, pp. 622–633.

DECLARATION

Put your declaration here.

France, Décembre 2016

Grégoire Lafay

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX :

https://bitbucket.org/amiede/classicthesis/