RAPPORT DE STAGE

Master II ATIAM

Université Pierre & Marie Curie, IRCAM, TelecomParisTech

Travail réalisé par

GRÉGOIRE LAFAY

Caractérisation sémantique des scènes sonores environnementales
-
Étude paramétrique et perceptive d'un paradigme de synthèse séquentielle par corpus

---

Equipe Perception et Design Sonores

I.R.C.A.M.

Encadrement

NICOLAS MISDARIIS & MATHIEU LAGRANGE

Année universitaire 2012 - 2013

RÉSUMÉ

L'objet du présent rapport porte sur la perception des environnements sonores. Nous proposons une nouvelle approche expérimentale, basée sur un paradigme de synthèse séquentielle par corpus, ayant pour but de caractériser des paysages sonores urbains types, à partir de données sémantiques et numériques. Pour ce faire, nous nous servons de l'environnement audio-numérique *SceneSynth*, développé dans le cadre du projet HOULE, ainsi que d'une interface dédiée à l'exploration de banques de sons, conçue pour les besoins de l'étude. Nous présentons les fondements théoriques sur lesquels s'appuie notre protocole expérimental et testons sa viabilité via une expérience pilote.

ABSTRACT

This report is about sound environments perception. We propose a new experimental approach to characterize urban soudscapes on the basis of numeric and semantic data. This approach is based on an synthesis process. In order to achieve this, We use a audio-digital environment called *SceneSynth*, developed in the framework of the HOULE project, and a web audio interface for sound data mining developed for the purposes of this study. We present the theories on which relies our experimental methodology and test its viability via a preliminary test.

REMERCIEMENTS

Je remercie très chaleureusement mes deux responsables, Nicolas Misdariis et Mathieu Lagrange, de m'avoir permis de réaliser ce stage. Malgré le caractère vaste du sujet, ils m'ont laissé prendre des initiatives, quitte à en modifier les objectifs, tout en me réfrénant quand mes idées devenaient trop fantasques. Cette expérience a été pour moi des plus enrichissante. Le caractère pluridisciplinaire du sujet m'a permis de m'ouvrir à nombre de problématiques passionnantes qui m'étaient alors inconnues, problématiques dont j'approfondirai l'étude l'année prochaine dans le cadre d'une thèse.

Je remercie les membres de l'équipe "Perception et Design Sonores" de l'IRCAM, Patrick Susini, Olivier Houix, Jean-julien Aucouturier et Mondher Ayari pour leurs conseils dispensés dans le cadre des nombreuses réunions d'équipe qui ponctuent agréablement la vie du laboratoire.

Je remercie mes collègues, stagiaires et doctorants de PdS, Emmanuel Ponsot, Sara Adhitya, Maxime Carron, Edgar Hemery, et Anne Laure Verneil pour les nombreuses discussions et l'excellente ambiance dans laquelle s'est déroulé le stage.

Je remercie bien sur mes collègues du master ATIAM, nombreux cette année à avoir réalisé leur stage à l'IRCAM, pour la bonne humeur dans laquelle s'est passée cette année. J'en profite également ici pour remercier toute l'équipe pédagogique de l'ATIAM de proposer un master aussi riche et diversifié en terme de matières abordées.

Enfin, j'aimerais adresser ici des remerciements tout particuliers à Mathias Rossignol, développeur du logiciel *SceneSynth*, pour la précieuse aide qu'il m'a fournie, tant sur des points théoriques que sur des points de développement. Je le remercie, malgré le décalage horaire, d'avoir toujours pris le temps de répondre rapidement à mes questions, posées via de très nombreux mails, pas toujours très bien organisés. Le stage n'aurait pas pu être mené à bien sans son engagement.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION GÉNÉRALE

L'objet du présent rapport porte sur la perception des environnements sonores. Il s'agit de proposer une nouvelle méthode permettant de caractériser, sur la base de paramètres sémantiques comme numériques, des paysages sonores urbains types.

Nous proposons une approche expérimentale originale, qui procède par la synthèse de scènes sonores urbaines. Pour ce faire, nous disposons de l'application web *SceneSynth*, développée dans le cadre du projet HOULE. *SceneSynth* est un outil de synthèse qui, à la manière d'un séquenceur audio, permet de créer des environnements sonores complexes, à partir d'un corpus de sons.

L'objectif du stage est de mettre au point un protocole expérimental viable permettant, via *SceneSynth*, d'objectiver et d'informer, des scènes sonores urbaines types, à partir d'éléments constitutifs desdites scènes sonores. Il s'agit alors de tester ce protocole, par le biais d'une expérience pilote, visant à examiner les résultats qu'il produit dans le cadre de pratiques-utilisateur et à apprécier les fonctionnalités de *SceneSynth*.

Afin de satisfaire à cet objectif, et au vu des problématiques rencontrées, nous avons procédé en 5 étapes :

1. l'étude bibliographique de la littérature traitant de l'approche perceptive et de l'approche cognitive en matière de caractérisation des environnements sonores, ce dans la perspective de poser les bases théoriques de notre propre démarche.
2. la constitution d'un corpus sonore de référence, sur la base d'une typologie également établie à partir d'une étude bibliographique. Ce corpus est structuré suivant des paramètres perceptifs.
3. le développement d'une interface web destinée à l'exploration du corpus sonore.
4. l'évaluation des performances de cette interface via une expérience de type *crowdsourcing*.
5. la mise en place de l'expérience pilote de synthèse, réalisée en laboratoire. Cette étape s'est elle-même subdivisée en 3 parties :
    a) l'intégration du corpus sonore à l'outil *SceneSynth*
    b) l'intégration de l'interface d'exploration à l'outil *SceneSynth*
    c) la préparation et la réalisation de l'expérience pilote de synthèse séquentielle par corpus

Ainsi, ce stage a comporté trois phases

- une phase "étude bibliographique"

- une phase "développement web"
- une phase "expérimentation"

Ce rapport se découpe suivant trois parties.

Dans la première, nous commençons par présenter les enjeux scientifiques liées à l'étude des environnements sonores urbains. Nous exposons les résultats de la recherche bibliographique menée sur l'étude de la perception des environnements sonores et les méthodes expérimentales mises en œuvre. Nous explicitons les courants et théories psychologiques sur lesquels se fondent ces méthodes ainsi que notre approche.

Dans la deuxième, nous présentons l'expérience de synthèse (hypothèse, paradigme, fondement théorique, aspect technique) et la situons par rapport une expérience de même type, réalisée dans le même domaine. Nous détaillons l'étape de création du corpus de sons ainsi que celle de l'interface d'exploration de ce dernier. Nous terminons en exposant le protocole expérimental de l'expérience pilote de synthèse et ses résultats.

Dans la troisième et dernière partie, nous discutons les différents points traités lors du stage et ouvrons sur les perspectives qu'offre un formalisme de synthèse appliqué à l'étude des paysages sonores.

Première partie

ÉTAT DE L'ART : PERCEVOIR UN ENVIRONNEMENT SONORE URBAIN

1

L'ENVIRONNEMENT SONORE URBAIN

Introduction

Dans cette partie nous introduirons les enjeux de l'étude de l'environnement sonore urbain, avant d'aborder la notion de "paysage sonore".

1.1 POURQUOI ÉTUDIER L'ENVIRONNEMENT SONORE URBAIN ?

La ville a toujours été un environnement bruyant, et ce quelles que soient les époques. Ce qui a par contre évolué, c'est la perception de ce bruit. C'est dans les années 80 que l'association bruit/pollution se fait la plus forte. Le bruit est alors considéré comme une dégradation globale de la qualité de vie (voir Raimbault, 2002). En réponse, une législation "anti-bruit" se met en place. Celle-ci prévoit de combattre les nuisances sonores en réduisant les niveaux d'intensité.

Mais le problème persiste, et pour cause, le bruit demeure un phénomène subjectif, autrement dit dépendant de "l'appréciation l'auditeur". Le bruit est affaire de contexte. Il peut rassurer, prévenir, divertir, comme gêner ou agacer. Corriger l'environnement sonore uniquement suivant des paramètres acoustiques, par définition objectifs (par exemple le niveau sonore), ne suffit donc pas. Rappelons d'ailleurs que ville agréable, ne rime pas avec ville silencieuse. En ce sens de nouveaux concepts d'"ambiances" ou de "paysages" sonores sont introduits. Ils envisagent alors la "nuisance sonore" d'une manière plus large, prenant en compte les aspects qualitatifs et sémantiques des phénomènes acoustiques. L'analyse de ces phénomènes réclame une méthodologie s'éloignant par définition de celle plus traditionnelle de la psychophysique.

Le bruit passe alors d'"un objet physique à un objet cognitif" (voir Guastavino, 2003). Il ne s'agit plus de savoir à partir de quand le bruit n'est plus nuisant, mais pourquoi tel bruit est perçu comme gênant par tel individu.

1.2 LA NOTION DE PAYSAGE SONORE

La notion de paysage sonore 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 se pose Schafer est alors :

> 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 ?

Le paysage sonore se définit comme "l'environnement sonore d'un sujet déterminé" (Nadrigny, 2010). La définition se veut très générale. Tout environnement peut être considéré comme un paysage sonore si tant est qu'on lui associe un ensemble de sons entendus par un sujet donné. Notons qu'il, le sujet, se réfère à des environnements existants aussi bien qu'à des structures abstraites comme un enregistrement musical.

En ce qui concerne l'amélioration de la qualité de l'environnement, Schafer explicite la nécessité de ne plus considérer seulement le "bruit", mais également sa perception par les individus qui le subissent. L'approche étant ainsi centrée sur le sujet, l'analyse des paysages sonores fait appel à une méthodologie issue de la psychologie et de la sociologie, et permettant ainsi de décrire les sons en respectant leur contexte d'écoute.

Depuis trente ans qu'elle existe, cette approche a permis de développer une base de descripteurs qualitatifs et acoustiques grâce auxquels nous jugeons mieux, et sommes mieux à même d'améliorer notre environnement sonore. Un des enjeux présent 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).

2

L'HOMME FACE AU MONDE SONORE : UNE APPROCHE COGNITIVE

Introduction

La perception du monde sonore qui nous entoure est un phénomène complexe et 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, à la seule lecture d'une partition de musique, le musicien entrainée est capable d'entendre la musique comme si elle était jouée.

Une des voies d'exploration des processus régissant la perception au sens large de notre environnement est de passer par la psychologie cognitive. Dans ce chapitre, nous décrivons les différentes étapes composant la chaîne de traitement de l'information sensorielle, puis nous introduisons l'approche cognitive appliquée à l'audition, et enfin nous introduisons certaines notions jugées utiles pour notre étude sur les environnements sonores urbains.

2.1 L'ÉTUDE ACOUSTIQUE DE LA PERCEPTION DU MONDE

2.1.1 *Percevoir l'environnement sonore*

La perception désigne l'ensemble des processus de traitement de l'information sensorielle. Ces processus nous permettent, par l'interprétation des données reçues en continu par nos organes, de construire une représentation interne du monde qui nous entoure (voir Houix, 2003).

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 :

- 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. Cela a été mis en évidence par la *Gestalt theorie*. Un exemple parlant concret, emprunté au domaine de la vision, est celui du phénomène dit de bi-stabilité. La faculté, chez un sujet, de voir dans une même image tantôt un canard, tantôt un lapin, autrement dit, de tirer d'un même stimulus deux analyses différentes, mais jamais simultanément.

Un autre exemple, cette fois dans le domaine de l'audition, nous semble illustrer le caractère dual de la perception. Il est donné par McAdams et Bigand (voir McAdams and Bigand, 1994). Nous les citons ici.

> "...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 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..."

Dans une approche cognitive (cf. 2.2), la perception revêt un caractère d'interprétation. Cette interprétation s'exerce à différents niveaux depuis l'analyse sensorielle du stimulus jusqu'à son analyse sémantique (voir Houdé et al., 1998).

2.1.2 *La chaîne de traitement*

L'étude de la perception des environnements sonores requiert une connaissance du processus de traitement de l'information auditive

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, transforme l'information acoustique en une 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 1 extraite du livre de McAdams et Bigand (voir McAdams and Bigand, 1994) nous donne un aperçu des principales fonctionnalités du système de traitement auditif.

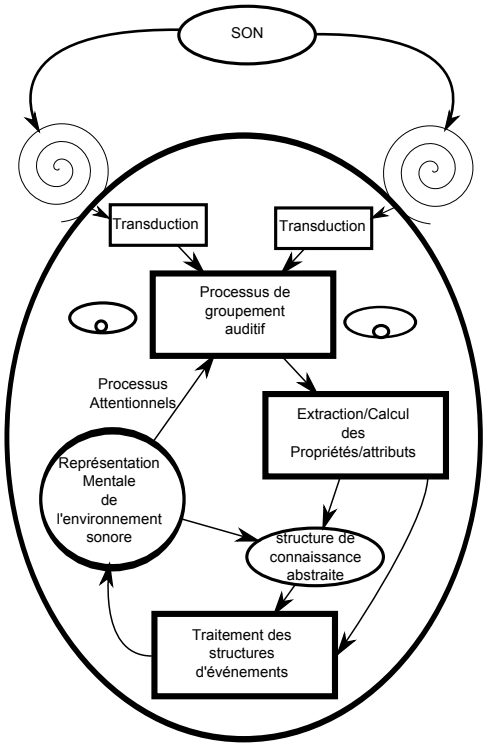

FIGURE 1: Principaux processus de traitement de l'information auditive et leurs interactions.

Lors de l'étape de *transduction*, les vibrations sonores parvenant au tympan sont analysées puis traduites en impulsions nerveuses transmises au cerveau. Ces impulsions rendent compte des attributs spectraux et temporels de l'onde.

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 Elhilali (2004). Contrairement à ce que pensaient les Grecs, nous ne possédons pas de "canaux" séparés pour chaque objet sonore présent dans l'environnement. 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 s'agit de déterminer, combien d'objets sonores sont présents, d'où viennent t-ils et quel est leur sens. Les processus à l'origine de ces associations sont regroupés sous le terme : analyse de scènes auditives (*Auditory scene analysis*)

Prenons pour exemple les chorals de Bach. C'est le *processus de groupement auditif* qui nous permet, sur la base des paramètres spectro-temporels du signal, de distinguer les quatre voix basse, ténor, alto et soprano. Par contre c'est à partir d'une analyse des paramètres 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 analyse perceptive intervient pendant la phase dite d'*extraction/calcul des propriétés/attributs*. C'est elle qui permet la "perception globale de la structure et de l'organisation temporelle des séquences sonores".

Une fois interprété perceptivement, le stimulus fait appel à des processus psychologiques et neuro-psychologiques afin de donner une signification à l'ensemble de la scène sonore, en établissant une relation sémantique entre les différentes sources identifiées.

2.2 LA PSYCHOLOGIE COGNITIVE

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 qui se fonde sur « l'étude des comportements objectivement observables de l'être humain », négligeant, de fait, le rôle de la conscience. La psychologie cognitive, au contraire elle, 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

2.2.1 *Une approche globale*

Nous partons d'une définition de la cognition proposée par Neisser[1] (voir Neisser, 1976) :

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

Le terme cognition renvoie à la notion de connaissance. Dans un sens plus précis, il désigne les conditions qui permettent l'acquisition et le développement d'une connaissance du monde.

L'approche cognitiviste, dans l'étude de la perception auditive, s'éloigne de celle plus traditionnelle de la psychoacoustique[2]. Tandis que la psychoacoustique émet l'hypo-

1. Ulric Neisser est considéré comme un des pères du cognitivisme notamment grâce à son livre *Cognitive Psychology*. Il a par la suite beaucoup critiqué la direction prise par le mouvement, lui reprochant son recourt excessif aux travaux en laboratoire au détriment des conditions in situ.

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

thèse d'une relation directe entre un stimulus et la réponse de l'individu à ce dernier, la psychologie cognitive soutient qu'à un stimulus, l'homme donne des réponses entièrement corrélées au contexte, à l'expérience, aux interactions multi-sensorielles (voir Maffiolo, 1997). Ces réponses tiennent 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)(voir Maffiolo, 1999).

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 se 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. (voir Maffiolo, 1999).

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 (voir McAdams and Bigand, 1994).

2.2.2 *Paradigme de la psychologie cognitive*

Comme nous l'avons vu, le cognitivisme ne conçoit pas l'individu comme une "boîte noire", mais envisage ce dernier comme un système de traitement de l'information. Le cognitivisme, fait ainsi l'analogie entre le fonctionnement humain et le fonctionnement de l'ordinateur (voir Maffiolo, 1999).

Il ne défend pas l'hypothèse d'un comportement linéaire entre un stimulus externe et la réponse du sujet. Il admet au contraire que le sujet adopte une stratégie, dans le but d'optimiser son comportement face au stimulus. Cette stratégie dépend de la nature du stimulus, du contexte ainsi que des connaissances a priori du sujet.

Maffiolo dans sa thèse *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain* (voir Maffiolo, 1999) propose un résumé des présupposés sur lesquels repose le cognitivisme (cf. figure : 2) :

- le monde est discrétisé en dimensions ou propriétés issues de la physique, considérées comme vraies
- ces dimensions ou propriétés peuvent être mesurées objectivement par des instruments, rendant ainsi compte de "la réalité"
- le sujet intègre de manière séquentielle ces dimensions ou propriétés en fonction du contexte
- l'évaluation subjective du sujet est mesurée comme un décalage par rapport à la mesure objective considérée comme vraie

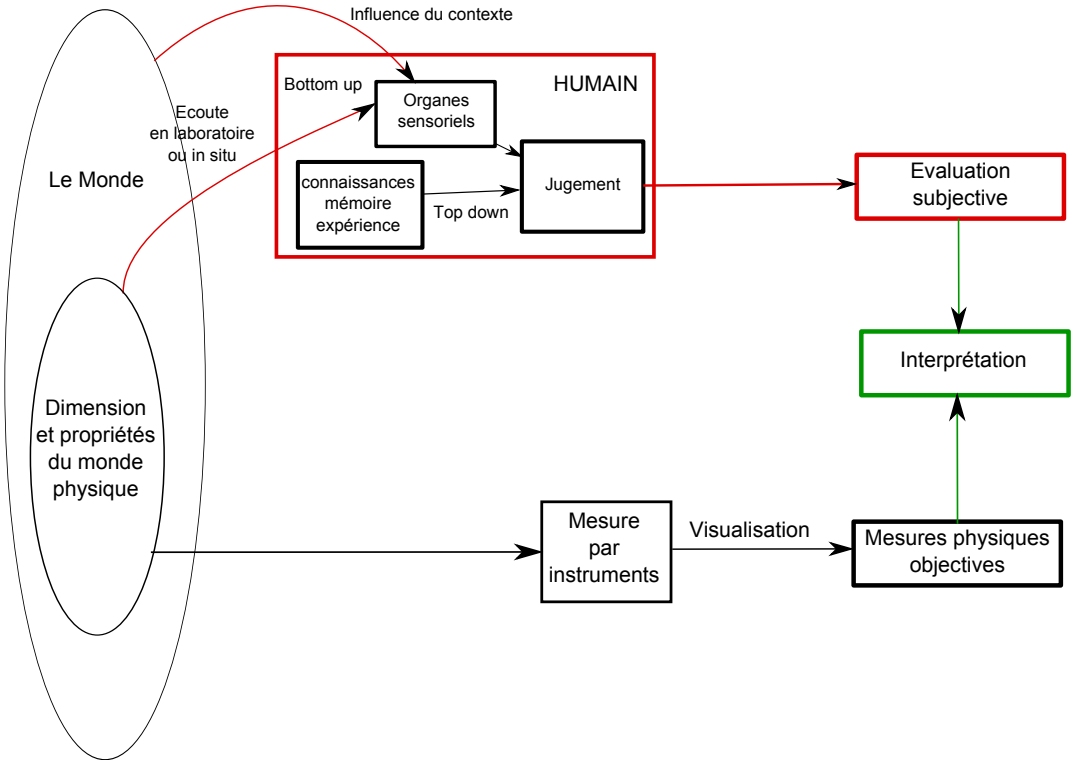

FIGURE 2: Paradigme du cognitivisme, d'après Maffiolo (1999)

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.

Il faut bien ici distinguer d'une part les approches cognitivistes, qui s'intéressent plus particulièrement aux processus de type *botttom up* relatifs au traitement de l'information perçue , des approches dites cognitives, lesquelles interrogent, avant tout, les processus de type *top-down* liés à la mémoire du sujet ainsi qu'au contexte. (voir Guastavino, 2003).

2.2.3 *Entre Perception et cognition : les représentations internes*

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 les systèmes cognitifs s'appuient sur des représentations mentales des réalités externes, par essence amodales.

Une définition des représentations mentales est donnée par Michel Denis dans 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 (voir McAdams and Bigand, 1994), 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.

La dichotomie entre perception et cognition a été vivement critiquée. 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 (voir Barsalou, 2010). Il tente ainsi de réunir les processus perceptifs et cognitifs (voir Goldstone and Barsalou, 1998; Barsalou, 1999)

La figure 3 illustre les deux approches :

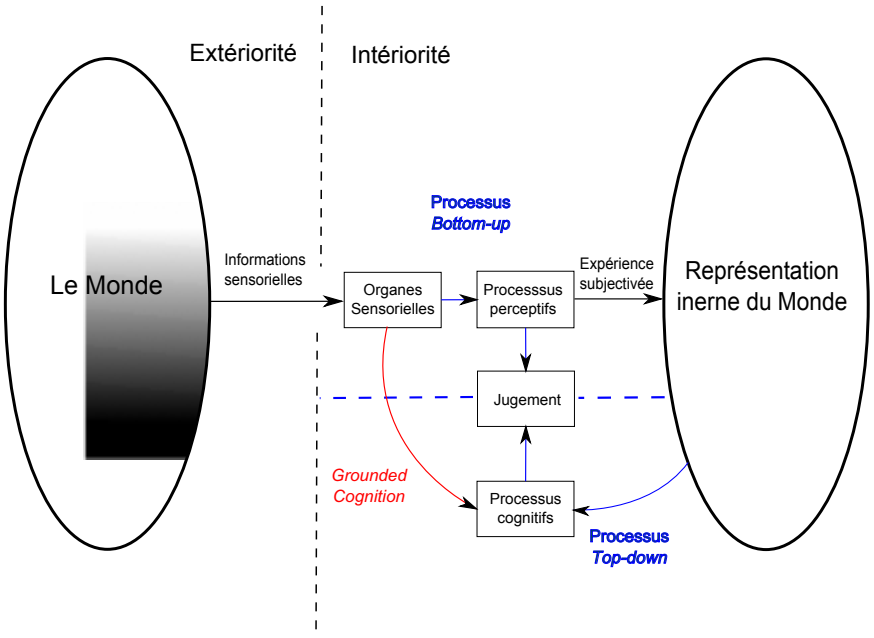

FIGURE 3: Processus cognitifs et perceptifs

2.2.4 *Une approche sémantique : La catégorisation*

La catégorisation peut être vue comme un processus mental nous permettant de discrétiser le monde extérieur, sur la base d'une représentation interne elle même partitionnée en catégories. Il s'agit de la conceptualisation du maillon de la chaîne d'analyse, faisant le lien entre d'une part le stimuli traité, et, d'autre part les connaissances en mémoire du sujet.

Il existe plusieurs théories sur la manière dont les catégories sont structurées et la façon dont elles opèrent sur les objets sonores perçus. Au nombre de celles-ci, la "théorie classique", la "théorie prototypique" et la "théorie des exemplaires". Nous exposons ici les grandes lignes de la théorie prototypique formalisée par Eleonor Rosch.

Selon Rosch, "la tâche fondamentale de tout organisme est de segmenter l'environnement en classifications à partir desquelles des stimuli non identiques peuvent être traités comme équivalents" (Rosch (1978) cité par Maffiolo (1999)).

Cette "classification" se fait par rapport à des catégories cognitives, ces dernières étant formées suivant l'expérience personnelle de chaque individu. D'après Rosch, (voir Dubois, 1993) la catégorisation repose sur deux principes psychologiques :

1. **Le principe d'économie** cognitive qui consiste à extraire le maximum d'informations d'un stimulus avec le moins d'effort possible.

2. **Le principe de la structure du monde perçu** qui postule "qu'à la différence des ensembles de stimuli utilisés dans les tâches traditionnelles de formation de concept, en laboratoire, le monde perçu n'est pas un ensemble totalement non structuré d'attributs de co-occurrences équiprobables." (Rosch and Lloyd (1978) cité par Dubois (1993)) En d'autres termes, le monde ne peut se réduire à des paramètres dimensionnés, fixés et indépendants, manipulables dans le cadre d'études en laboratoire. Au contraire, les objets dont il est composé sont 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"). Ainsi, nous pouvons déduire la nature d'un objet en observant les corrélations et discontinuités présentes au niveau des attributs entre l'objet en question, et les prototypes des catégories en mémoire.

La théorie postule que les catégories sont structurées en interne autour de prototypes. Un prototype étant l'élément le plus représentatif des objets de la catégorie, et donc possédant les attributs assumés typiques de celle-ci.

Il est à noter que les frontières entre les différentes catégories ne sont pas figées et peuvent se recouvrir. La typicité d'un élément d'une catégorie est donc fonction de son degré d'appartenance à celle-ci, ainsi que de son indépendance vis à vis des autres catégories.

L'appartenance d'un objet à une catégorie dépend de la ressemblance qu'entretient ce dernier avec le prototype. Le degré de similarité entre un objet et le prototype se juge

de manière globale, suivant l'ensemble des propriétés attribuées à ces derniers, et non en comparant séparément chacune d'elles.

Outre cette structure intra-catégorielle (dite aussi dimension horizontale), la théorie postule également l'existence d'une organisation inter-catégorielle (dite aussi dimension verticale). Elle se décline en trois niveau hiérarchisés : Le niveau superordonné, le niveau intermédiaire (ou de base), et le niveau subordonné (cf. figure 4).

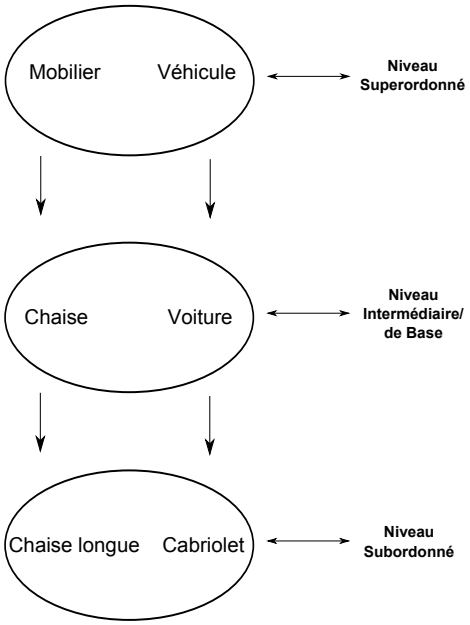

FIGURE 4: Théorie prototypique : structure inter-catégorielle

Selon Rosch, le niveau de base est celui sur lequel fonctionne les individus appartenant à la même sphère socioculturelle, au même groupe social.

2.2.5 *Une approche écologique*

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

Cette approche reconnaît que la réponse à un stimulus dépend et de l'information perçue (processus *bottom-up*), et de la connaissance du monde (processus *top-down*), autrement dit l'environnement quotidien et le contexte habituel d'écoute du stimulus.

Afin de garantir une validité écologique, l'approche écologique requiert, dans le cadre d'études perceptives sonores, de prendre en compte cet environnement particulier dans lequel gravitent le sujet et le stimulus. Elle s'oppose ainsi aux méthodes expérimentales traditionnelles, celles de la psychophysique en particulier, en postulant que les expériences en laboratoire décontextualisent le sujet et sa réponse. (voir Guastavino, 2003)

2.3 NOTIONS LIÉES À L'ÉTUDE DES ENVIRONNEMENTS SONORES

Dans cette partie nous introduisons trois notions jugées importantes dans le cadre de notre étude perceptive/cognitive des environnements sonores.

2.3.1 *Traitement holistique ou analytique*

On fait l'hypothèse de l'existence de deux processus psychologiques de traitement distincts, opérant en fonction de la nature du stimulus :

- Le traitement analytique
- Le traitement holistique

Le premier intervient quand des dimensions sont identifiables et séparables à l'intérieur d'un stimulus. Celui-ci est alors divisé en objet ayant une identité dimensionnelle propre. Quand cette division ne peut s'effectuer, le stimulus est alors perçu comme un tout unitaire. Le traitement est alors holistique.

Maffiolo, dans sa thèse (voir Maffiolo, 1999), a mis en évidence l'existence d'un couplage fort entre les processus de traitement (analytique/holistique) et la nature (évènementielle/amorphe) des sons. Elle montre que les séquences sonores dites évènementielles, par définition séparables, requièrent une écoute analytique. Décrites par les sujets en termes de sources sonores, d'évènements et d'activités, elles bénéficient d'un traitement sémantique de type *top-down*. Les séquences amorphes, quant à elles, requièrent une écoute holistique. Décrites au moyen de descripteurs acoustiques, elles bénéficient d'un traitement sémantique de type *bottum up*.

A noter la différence entre séquence sonore événementielle et séquence sonore amorphe est robuste aux changements d'intensité.

Cette même idée de couplage entre traitement analytique/holistique et nature événementielle/amorphe du stimulus est présente dans d'autres études sur les environnements sonores urbains : Raimbault (2006), Guastavino (2006), Dubois et al. (2006).

2.3.2 *Attention et saillance*

L'"Attention" est la capacité de notre système auditif à se focaliser sur des composantes spécifiques de notre environnement sonore en ignorant le reste. Ces composantes forment un groupe appelé "flux auditif". Le flux audio dépend de plusieurs indices (Gestalt theorie, acoustique) (voir Elhilali, 2004). En fonction du contexte de la scène, certains flux ont tendance à attirer plus facilement notre "attention". Un des paramètres pouvant susciter l'attention est la saillance.

La saillance d'un flux audio peut se voir comme l'impact potentiel d'un stimulus sur notre perception et notre comportement. Cette saillance est fonction du contexte d'écoute de la scène sonore. L'attention et la saillance ont une influence dans l'identification des sources. Cette identification, et l'attribution de "sens" qui en découle, est une étape primordiale dans le processus de création de l'image mentale d'un environnement, à partir de la perception de son empreinte sonore. Ainsi, un élément saillant est facilement identifiable. A l'inverse, les sources d'un fond sonore (background) par définition peu saillant (i.e attirant potentiellement moins l'attention) seront moins discernables. (voir Elhilali et al., 2009).

De Coensel et Botteldooren proposent plusieurs modèles permettant de simuler l'attention (voir Botteldooren et al. (2006), Botteldooren and De Coensel (2009), De Coensel et al. (2010) et De Coensel and Botteldooren (2010)). Les modèles calculent une "carte de saillance" décrivant l'évolution de la saillance d'une scène en fonction du temps. Les deux chercheurs partent du principe que le cerveau ne peut pas traiter toutes les informations en même temps. Il sélectionne l'information utile. Ces modèles ne prennent pas en compte les traitements de type *top-down* et se concentrent sur les processus *bottom-up* relatifs à l'analyse du signal des stimuli.

Deuxième partie

L'EXPÉRIENCE DE SYNTHÈSE PAR CORPUS

3

L'APPROCHE PAR LA SYNTHÈSE

Introduction

Dans ce chapitre nous parlons de l'expérience de synthèse séquentielle par corpus. Il ne s'agit pas ici d'exposer le protocole expérimental exact mais d'introduire notre approche en explicitant les théories et hypothèses sur lesquelles elle s'appuie. Nous situons par ailleurs notre expérience par rapport aux autres recherches sur la perception des environnements sonores.

3.1 RÉSUMÉ DE L'APPROCHE ADOPTÉE

Nous cherchons à déchiffrer les représentations mentales des environnements sonores propres aux grandes villes. Pour cela nous demandons au sujet de reconstituer un paysage sonore complexe via un processus de synthèse séquentielle par corpus.

Nous adoptons une approche expérimentale. Nous mettons à la disposition du sujet un corpus composé de sons environnementaux urbains. Ce corpus est divisé en deux grandes parties. L'une composée d'événements, sonores et l'autre composée de textures sonores (cf. section 4.1 pour plus de détails). Les évènements sonores sont des sons isolables et saillants (cf. section 2.3.2). Les textures sonores sont des sons potentiellement infinis en durée, des sons pour lesquels on peut difficilement identifier une source.

Le but de l'expérience est de synthétiser une scène sonore. Le sujet réalise le paysage en piochant des sons à l'intérieur du corpus via un le logiciel de synthèse séquentielle *SceneSynth*. A la manière d'un séquenceur audio, il place les sons sélectionnés sur des "pistes audio" graduées temporellement. Nous proposons au sujet plusieurs contrôleurs permettant de paramétrer ces dernières. A chaque piste audio correspond un son sélectionné.

Afin de ne pas influer sur la sélection des échantillons sonores, le sujet explore la banque de sons uniquement à l'écoute, grâce à une interface nommée *Speed Sound Finding*, interface développée dans le cadre du stage (cf. chapitre 5).

Les représentations mentales n'étant pas directement accessibles, nous passons par une procédure d'objectivation psychologique. Cette objectivation tient compte de plusieurs données que nous récupérons pendant l'expérience.

Ces données sont de deux sortes :

- les données verbales

- les données numériques

Les données verbales sont composées :

1. du nom attribué par le sujet au paysage sonore synthétisé
2. du nom attribué par le sujet à chaque son sélectionné
3. du nom "réel" ou "tag" de chaque échantillon sélectionné (source enregistrée, lieu de l'enregistrement)
4. d'une description du paysage sonore synthétisé sous la forme d'un questionnaire libre

Pour les données numériques, nous considérons les valeurs des paramètres de contrôle appliqués par le sujet à chaque piste audio.

A partir de ces données, nous cherchons à identifier des invariants structuraux afin de faire ressortir des catégories explicitant le tissu sonore urbain, catégories informées par des données sémantiques et numériques.

3.2 HYPOTHÈSES ET FONDEMENT THÉORIQUE

3.2.1 *Approche cognitive*

Notre expérience se veut une étude perceptive basée sur des théories issues de la psychologie cognitive (cf. section 2.2). Elle participe de l'étude des processus mentaux mis en œuvre dans l'analyse de stimuli. Elle cherche à mettre en évidence les ressources sur lesquelles s'appuient ces processus.

Nous admettons l'existence d'une représentation interne du monde, propre à chaque individu, et jouant un rôle primordial dans l'analyse de l'information perçue (cf. section 2.2.3).

A travers notre expérience nous recherchons un moyen original d'objectiver et d'informer ces représentations, à partir de données à la fois textuelles et numériques.

3.2.2 *Catégorisation*

Conceptualisant le lien entre perception et connaissance, nous admettons que notre représentation interne du sonore se structure autour de catégories perceptives. Nous adoptons une approche prototypique de la théorie de la catégorisation, basée sur une condition d'appartenance par similarité (cf. chapitre 2.2.4). Nous tenons que chaque catégorie est organisée en interne, autour de prototypes représentatifs des caractéristiques communes aux éléments de cette dernière.

Notre expérience a pour but de faire ressortir les principales catégories nous permettant de discriminer le tissu sonore urbain. Notons que ces catégories n'auront de sens perceptif que dans la mesure où elles seront identifiées par des sujets partageant la même sphère socioculturelle, ou, du moins, le même socle de connaissances acquises.

Notre expérience ne peut cependant pas être vue comme une expérience de catégorisation à proprement parler. Nous ne demandons pas au sujet de classifier des stimuli. Nous déduisons des catégories, des sons sélectionnés par lui pour synthétiser la scène.

3.2.3 *Dimension écologique de notre approche*

Ce sont les avantages inhérents à notre protocole qui nous permettent de satisfaire, d'une manière nouvelle, les conditions écologiques (cf. section 2.2.5). Ces avantages peuvent se résumer en trois points :

1. Le processus de synthèse permet, par définition, de re-contextualiser le sujet dans son environnement naturel.
2. Nous proposons au sujet un corpus de sons de taille importante au regard de ceux utilisés habituellement dans ce genre d'expérience (voir Houix et al., 2012). Cela nous permet de rendre compte, autant que possible, de la diversité sonore de tels environnements, et, par la même, de s'affranchir de la capacité de mémoire du sujet.
3. Nous demandons au sujet de nous faire remonter les sons qu'il n'aurait pas trouvés, absents du corpus proposé.

Nous montrerons, lors de l'analyse des résultats de notre expérience pilote, que ce protocole permet de garantir une validité écologique forte.

Notons que dans le cadre de ce stage, l'expérience à l'origine pour une diffusion en *crowdsourcing*, s'est finalement déroulée en laboratoire. S'agissant d'un test pilote, nous avons souhaité garder un contrôle sur l'expérience, notamment au niveau de la présentation des consignes.

3.3 ORIGINALITÉ DE L'APPROCHE PAR LA SYNTHÈSE : COMPARAISON AVEC LES TRAVAUX DE MAFFIOLO

Afin d'expliciter l'originalité de notre approche, nous nous proposons ici de comparer notre protocole expérimental à celui de Maffiolo lors de sa thèse (voir Maffiolo, 1999) (cf. section 2.2.2). Comme elle, nous cherchons à comprendre comment sont structurées les représentations mentales du milieu sonore urbain, en utilisant les outils fournis par la psychologie cognitive. Les figures 5 6 illustrent les deux approches.

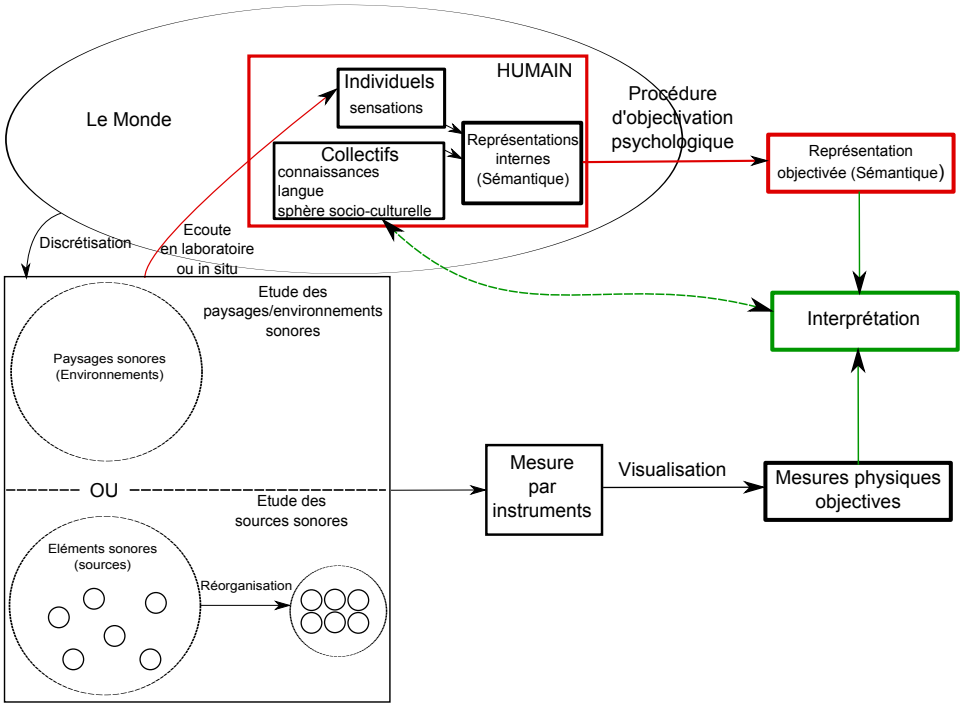

FIGURE 5: Protocole expérimental mis en place par Maffiolo , d'après Maffiolo (1999)

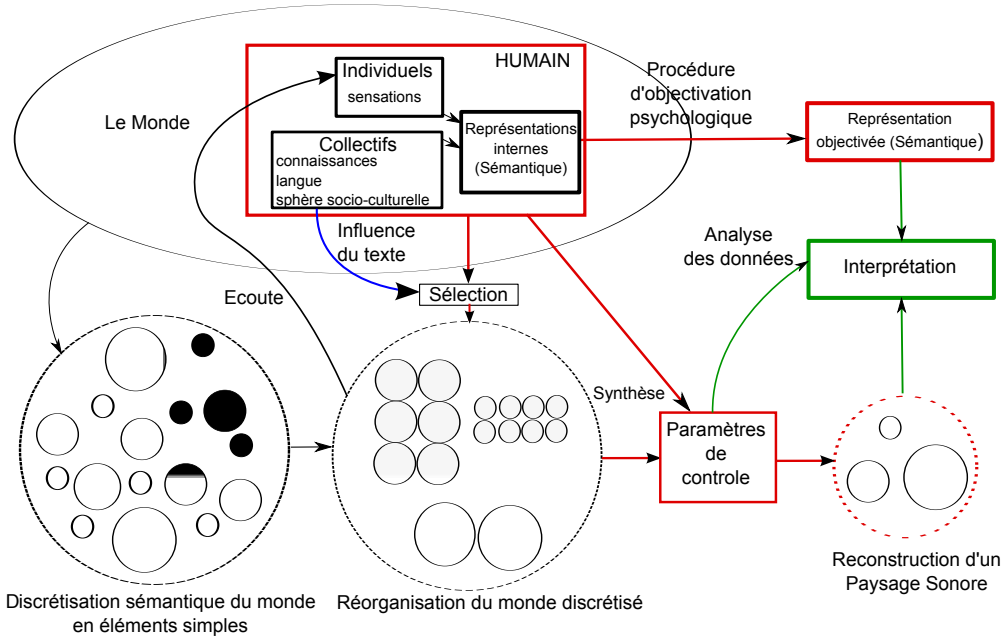

FIGURE 6: Protocole expérimental de l'expérience de synthèse par corpus

3.3.1 *Constitution d'un corpus sonore : vérité de terrain*

Afin de préciser la nature des représentations mentales, Maffiolo effectue une étude préliminaire sur la base de questionnaires graphiques et verbaux. Elle cherche à identifier et caractériser, par le biais d'une double objectivation psychologique, "les terrains urbains ayant des qualités acoustiques particulières". Ses travaux portent sur les ambiances sonores de Paris. Elle envisage ainsi l'"ambiance sonore" d'un lieu comme l'objet unitaire de son étude. Pour notre part, nous nous plaçons à un niveau plus précis, en considérant comme unité de base les éléments sonores composant ces ambiances.

De l'étude précédente, Maffiolo identifie quatre lieux ayant une empreinte sonore caractéristique. Des enregistrements de ces derniers, elle extrait 16 séquences d'environ 15 secondes (4 par lieux). Ces 16 séquences sont réparties en deux groupes : les séquences évènementielles, composées d'événements sonores discriminables, et les séquences amorphes, composées d'un bruit de fond d'où n'émerge aucun évènement faisant sens. Nous conservons là une distinction similaire au sein de nos propres enregistrements, d'un côté nous considérons les événements sonores, c'est à dire les éléments constituant les scènes événementielles, et de l'autre les textures sonores dont notre définition reprend celle donnée pour les scènes amorphes (cf. section 2.3.1).

Dans le travail de Maffiolo, les systèmes d'acquisition et de diffusion des sons ont fait l'objet d'une étude à part entière, ceci afin de garantir un rendu écologique. L'effet décontextualisant de l'écoute en laboratoire peut en effet "mutiler la perception normale de l'individu en situation". Cette attention particulière se retrouve aussi dans les travaux de Vogel et Guastavino (voir Guastavino (2003) et Vogel (1999)).

En ce qui nous concerne, nous n'avons pas poussé ce point plus avant. Nous considérons que le caractère écologique de notre approche est intrinsèque au processus de synthèse. En revanche, nous accordons un soin tout particulier à la structuration et à la présentation du corpus sonore afin de contrôler son influence sur le sujet (cf. chapitres 4 et 5).

3.3.2 *Procédure expérimentale : méthode d'objectivation*

Afin de tester ses hypothèses relatives au traitement cognitif des ambiances urbaines, Maffiolo utilise plusieurs procédés expérimentaux dont elle compare les résultats, considérant que "c'est la cohérence des informations tirées de diverses approches qui reste, en dernière analyse, le meilleur garant" de résultats fiables. Nos objectifs, fixés en début de stage, n'étant pas aussi ambitieux, nous ne comparons notre méthode qu'à une partie restreinte du programme expérimental déployé par Maffiolo.

Afin d'appréhender l'organisation des représentations mentales, Maffiolo passe par un procédé de catégorisation libre. Il s'agit de demander au sujet de "trier" les sons sans contrainte quant au nombre de classes, suivant une consigne donnée (principes d'inten-

sité et d'agrément). Le sujet doit par la suite, nommer et décrire chacune des classes. Le traitement des résultats est alors double. D'un côté une analyse de dissimilarité sur les regroupements (analyse arborée et multidimensionnelle). De l'autre une analyse linguistique des données verbalisées.

Les catégories mises en évidence par ce procédé portent sur des scènes sonores complexes. La robustesse du groupement catégoriel obtenu est appréciée en suivant la variation de deux paramètres expérimentaux :

- la variation de la consigne de classification (intensité, agrément) pour un même groupe de stimuli
- la variation de la nature physique des stimuli (intensité) pour une même consigne de classification

Nous adoptons une méthode inverse à celle de la catégorisation. Nous proposons au sujet de reconstituer un paysage sonore complexe à partir d'éléments simples, structurés suivant un organigramme et des classes issus principalement de la littérature. Nous lui demandons de nommer le paysage sonore constitué ainsi que les échantillons utilisés. Nous recherchons alors des recoupements à l'intérieur de ces verbalisations afin d'identifier :

- des catégories de scènes sonores urbaines
- des "classes d'éléments" au sein de ces mêmes catégories

Notre procédé permet ainsi d'étudier la manière dont sont composées les grandes catégories de scènes urbaines et d'y expliciter des classes de sons présents. Nous informons ainsi sémantiquement les scènes obtenues.

3.3.3 *Analyse des données*

Maffiolo met en évidence les catégories en faisant ressortir des proximités suivant les diverses classifications des sujets. Ces similarités sont obtenues suivant des modèles spatiaux (dans son cas une analyse multidimensionnelle de proximité : *MultiDimensional Scaling*), suivant aussi des modèles de réseaux (représentation arborée). Elle vérifie alors si les regroupements obtenus sont d'ordre sémantique ou acoustique, et s'ils correspondent aux différentes mesures physiques des séquences. Ces mesures sont considérées comme objectives, renvoyant à une réalité du monde.

En ce qui nous concerne, nous considérons comme "vérité-terrain", les tags sémantiques de nos échantillons. Ceux-ci faisant majoritairement référence à la source présente, s'agissant des évènements (pas, voiture), ou au lieu d'enregistrement, s'agissant des textures (place, parc). Nous comparons ces tags aux noms donnés par les sujets, et cherchons des invariants parmi ces données. Nous observons aussi la présence d'une certaine stabilité au niveau des paramètres numériques appliqués (cf. 3.4).

3.4 L'ENVIRONNEMENT AUDIO-NUMÉRIQUE SCENESYNTH

SceneSynth est un environnement de travail audio-numérique développé dans le cadre du projet HOULE [1], et permettant de synthétiser des paysages sonores à partir d'un corpus de sons. Il est prévu pour fonctionner via le navigateur internet Chrome [2].

3.4.1 *La base de données*

SceneSynth est relié à une application web permettant de gérer sa base de données. C'est grâce à cette application que l'on peut ajouter des sons au corpus de synthèse. *SceneSynth* fait la différence entre deux types de sons : les évènements sonores d'une part et les textures sonores d'autre part. Nous renvoyons le lecteur à la section 4.1 pour les détails concernant ces notions.

Les sons que nous ajoutons à la base de données sont appelés "sons-sources". Il s'agit de séquences sonores pouvant comporter indifféremment des événements et des textures. Une fois dans la base, nous avons la possibilité de découper des "fragments" de ces "sons sources". Chacun de ces fragments est alors soit un évènement sonore, soit une texture. Ces fragments sont ensuite regroupés en "collections" suivant leur valeur sémantique. Les collections sont ensuite structurées en classes hiérarchisées, selon une organisation là encore sémantique.

Prenons un exemple. Considérons une séquence d'un enregistrement de rue comportant des passages de voitures et de scooters. La séquence est considérée comme un "son-source". C'est elle que nous chargeons dans la base de données. Dans cette séquence nous venons découper des fragments, via l'interface web de gestion de la banque de sons. Chacun de ces fragments correspond dans notre cas, à un évènement sonore de "passage de voiture" ou de "passage de scooter". Tous les fragments de "passage de voiture" et de "passage de scooter" sont regroupés respectivement dans les collections *passage-voiture* et *passage-scooter*. Ces deux collections sont alors elles même groupées dans la classe *Circulation*.

Les collections ne regroupent que des fragments extraits d'un même enregistrement, ceci afin de garantir un rendu réaliste.

A partir de la structure hiérarchique ainsi créée, nous construisons l'interface d'exploration du corpus *Speed Sound Finding* (cf. chapitre 5). Une telle fonctionnalité nous permet de conserver un lien dynamique entre la banque de son, *Speed Sound Finding* et *SceneSynth* (cf. figure 7). Nous précisons que deux instances de *Speed Sound Finding* sont créées, l'une pour les évènements et l'autre pour les textures.

1. Pour plus d'informations sur le Projet HOULE voir http://houle.ircam.fr/wordpress/
2. voir http://www.google.fr/intl/fr/chrome/

Nous soulignons ici que l'utilisateur de *SceneSynth* a accès, via *Speed Sound Finding*, non pas à des "sons" mais à des "collections de sons" (*passage-voiture* et *passage-scooter*), chacune de ces collections étant représentée par un prototype, choisis arbitrairement comme le son le plus représentatif de la collection.

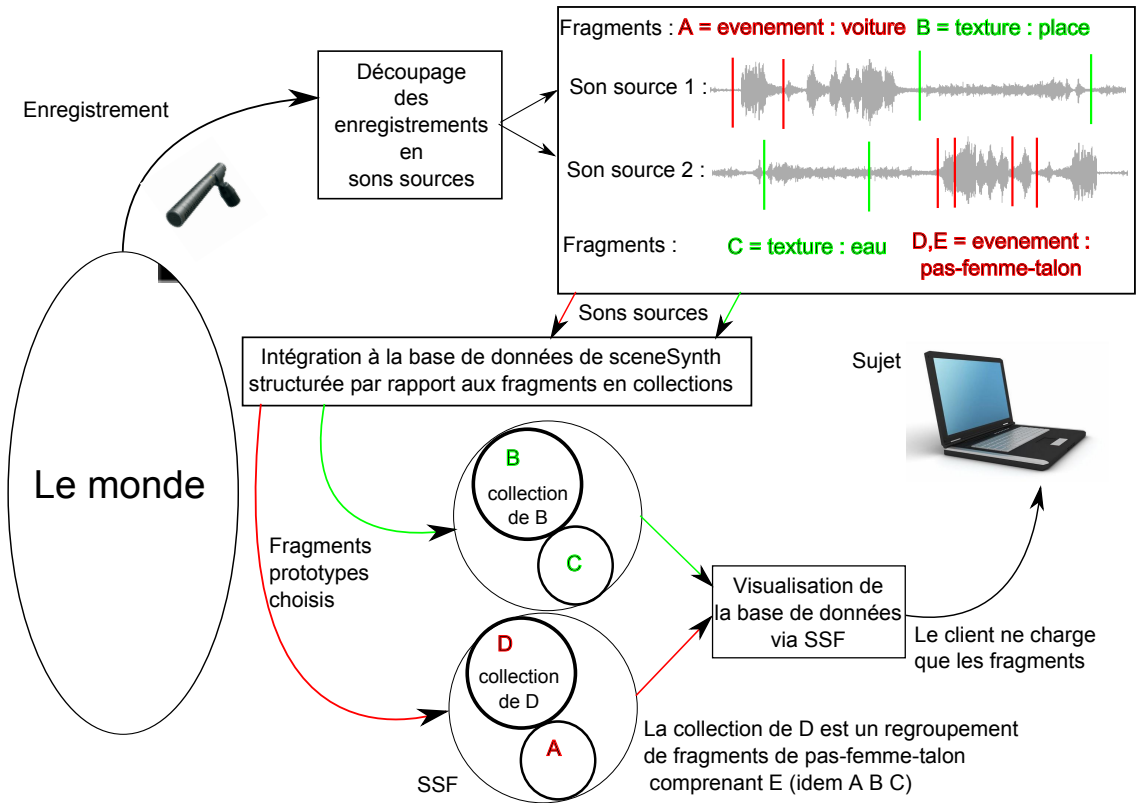

FIGURE 7: SceneSynth : lien avec la base de données

3.4.2 *La fonctionnalité de synthèse*

Le fonctionnement de *SceneSynth* se rapproche de celui d'un séquenceur audio. Chaque qu'un utilisateur choisit une collection de sons, en désignant son prototype, une piste audio est créée. Il peut alors modifier certaines propriétés du son via un groupe de paramètres de contrôle propre à chaque piste audio.

Lorsqu'il sélectionne une collection d'évènements sonores, une série d'objets apparaît sur la piste (cf. figure 8). Chacun de ces objets correspond à un fragment de la collection cooptée. Si l'utilisateur choisit une collection de textures sonores, un unique objet apparaît sur toute la longueur de la piste (cf. figure 8). Cet objet peut être composé soit d'un fragment de texture suffisamment long, soit de plusieurs fragments enchainés suivant un système de boucle.

Les paramètres disponibles dépendent de la nature événement/texture de la collection choisie. Pour les événements sonores, ces paramètres sont au nombre de quatre :

- **Niveau sonore : moyenne** : la moyenne des niveaux sonores en dB de tous les fragments.
- **Niveau sonore : Écart type** : l'écart type entre les niveaux sonores des fragments.
- **Position dans la scène (début/fin** : la position en seconde de départ/fin du premier/dernier fragment.
- **Intervalle entre les fragments : moyenne** : la moyenne en seconde des intervalles de temps séparant les différents fragments.
- **Intervalle entre les fragments : écart type** : l'écart type entre les intervalles séparant les différents fragments.
- **Fondu global In/out** : une transition progressive des niveaux sonores de début et de fin, réalisée sur l'ensemble des fragments.
- **Fondu In/out** : un fondu symétrique propre au niveau sonore de chaque fragment.

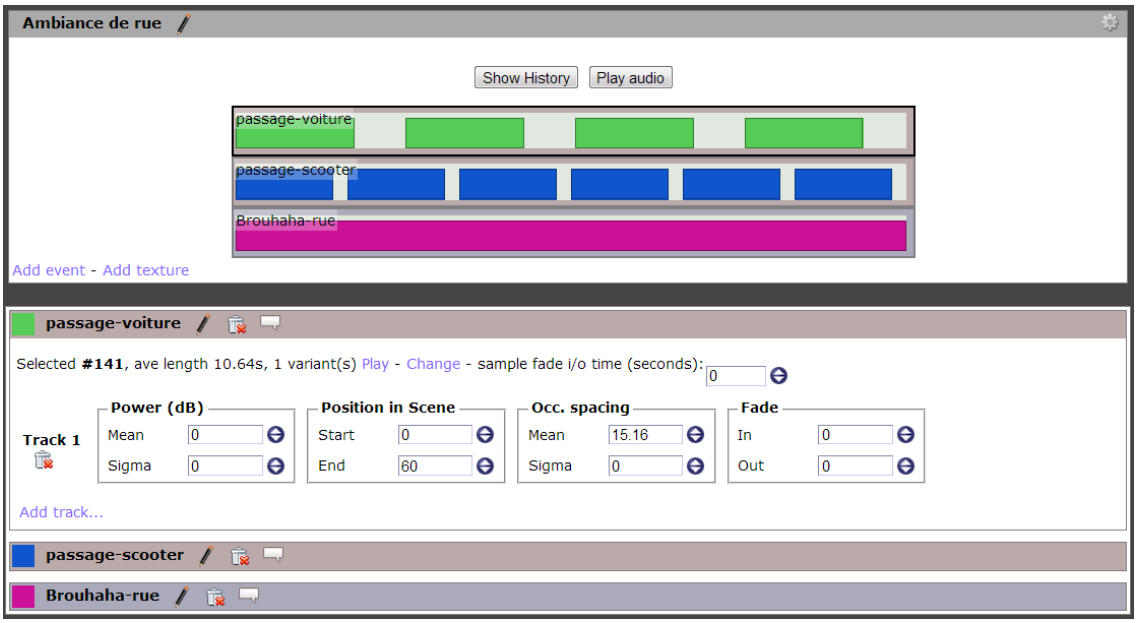

FIGURE 8: SceneSynth : interface de la fonctionnalité de synthèse

Pour les textures, l'utilisateur n'a pas accès à l'"intervalle entre les fragments", la texture occupant l'intégralité de la piste sonore.

L'application est accessible à l'adresse suivante : `http://217.70.189.118/soundthings/SceneSynth-SSF/`. Elle requiert l'utilisation du navigateur Chrome pour fonctionner. La figure 8 donne un aperçu des points précédemment évoqués.

4

CRÉATION D'UN CORPUS DE SONS ENVIRONNEMENTAUX URBAINS

Introduction

Dans cette partie nous décrivons l'étape de constructions du corpus sonore.

4.1 LES NOTIONS DE *texture* ET *évènement*

Essayer d'établir une typologie des éléments constituant un paysage sonore urbain revient à se demander comment décomposer ledit paysage sonore.
Nos éléments sonores sont répartis en deux grandes catégories :

- les textures
- les événements

Cette séparation est intrinsèque à la fonctionnalité de synthèse du logiciel SceneSynth, les deux catégories bénéficiant d'un traitement différent (cf. section 3.4).

Nous établissons :

1. qu'une texture sonore possède les qualités suivantes :
   - Elle est composée soit d'un son continu, soit de plusieurs sons répétés de manière périodique, provenant de sources acoustiques similaires.
   - Elle est temporellement homogène.
   - Elle favorise un traitement holistique de la part de l'auditeur, ainsi qu'une analyse acoustique (en fonction de paramètres physiques comme l'intensité) et non sémantique.
   - Il est difficile d'en identifier les sources sonores.
2. qu'un évènement sonore possède les qualités suivantes :
   - Il s'agit d'un son, ou d'un groupe d'éléments sonores (agrafeuse, porte qui se ferme) borné(s) temporellement.
   - Les éléments sonores composant un évènement sont tous issus de la même source. L'identification de la source dépend du niveau d'écoute (acoustique, causal, sémantique) de l'auditeur.

Pour la notion de texture, nous sommes très proches de celle des séquences amorphes proposée par Maffiolo (voir Maffiolo, 1999). De même, on peut voir nos évènements

comme les objets composant les séquences évènementielles.

Cette dimension évènement/texture est "orthogonale" à celle de "bruit de fond" / "événements de premier plan" (*background/foreground*[1]), utilisée dans le langage courant pour discriminer l'environnement urbain.

Concernant les notions de *background* et de *foreground*, nous considérons que l'une et l'autre peuvent être vue comme une somme d'éléments appelés "entités perceptuelles". Une "entité perceptuelle" peut être composée de textures et d'événements regroupés dans le but de faciliter le traitement auditif de la scène. Pour le background, ces textures ou évènements sont peu saillants, c'est à dire qu'ils attirent potentiellement peu l'attention. Pour le foreground, ces textures ou événements sont saillants (cf. section 2.3.2).

4.2 UNE TYPOLOGIE DES SONS URBAINS

Dans le but de créer un corpus de référence pour la synthèse, nous avons réalisé une typologie des sons environnementaux urbains. Dans un premier temps, les éléments présents dans cette typologie sont issus d'une étude bibliographique.

Nous recherchons les sources et ambiances sonores les plus souvent citées dans la littérature. Notre étude porte sur 16 articles ou thèses. Chacun d'eux traite de la manière dont nous discriminons les paysages sonores urbains. Plusieurs approches sont possibles, nous en avons relevés 3 :

- 10 articles abordent le problème par une approche cognitive, identifiant ainsi des catégories de sons : voir Guastavino (2006) Niessen et al. (2010), Brown et al. (2011), Raimbault and Dubois (2005), Raimbault (2002), Guastavino (2003), Devergie (2006) Maffiolo (1999), Maffiolo et al. (1998), (Guastavino et al., 2005).
- 3 articles proposent une classification morpho-typologique, divisant l'environnement sonore urbain en "zones sonores" possédant une identité acoustique forte selon la configuration du site. Polack et al. (2008), Beaumont et al. (2004), Menzel et al. (2011).
- 3 articles répertorient et classifient les sons d'un point de vu expert : Léobon (1997), Leobon (1986), Defréville et al. (2004).

A partir des éléments relevés, nous établissons deux typologies : une pour les textures (cf. figure 9), et une pour les événements (cf. figure 4.1). Afin de structurer ces deux ensembles, nous regroupons leurs éléments en classes hiérarchisées. La nature des classes est établie par rapport aux catégories perceptives les plus souvent citées dans la littérature.

1. Nous conserverons cette appellation de *background/ foreground* dans la suite du rapport.

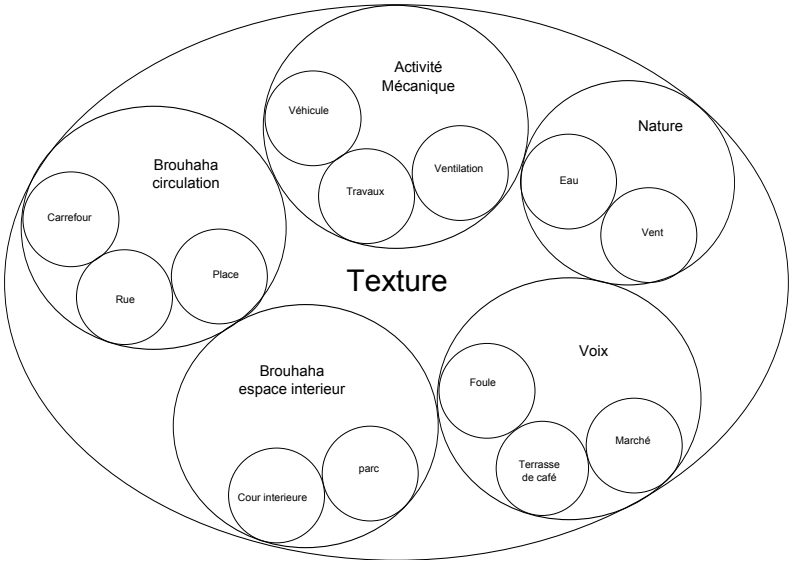

FIGURE 9: Typologie et regroupement des textures sonores

Pour les évènements, les regroupements se font en grande majorité par rapport à la source et sont d'ordre sémantique. Pour les textures nous considérons également la morphologie des lieux hébergeant ces dernières.

Dans un deuxième temps, nos structures typologiques ont été adaptées en fonction des sons disponibles dans les banques de sons, et surtout dans les rues de Paris lors des enregistrements.

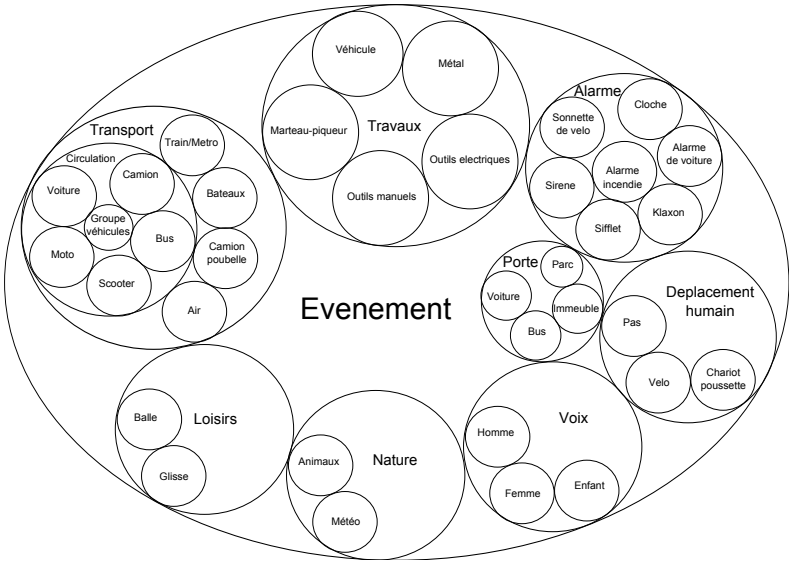

FIGURE 10: Typologie et regroupement des évènements sonores

4.3 CRÉATION DE LA BASE DE DONNÉES

Sur la base des typologies précédemment établies, nous avons collecté 482 sons, dont 381 "sons-sources" d'évènements et 102 "sons-sources" de textures (cf. 3.4 pour plus de détails sur la notion de "son-source").

Parmi les "sons-sources" d'événements :

- 260 sont issus d'enregistrements effectués pendant ce stage
- 89 sont issus de la banque de sons *SoundIdeas* [2]
- 32 sont issus de la banque de sons *Universal SoundBank* [3]

Parmi les "sons-sources" de textures :

- 72 sont issus d'enregistrements effectués pendant ce stage
- 23 sont issus de la banque de sons *SoundIdeas*
- 7 sont issus de la banque de sons *Universal SoundBank*

Tous les enregistrements ont été effectués à l'aide d'un micro canon *AT8035* [4] relié à un enregistreur *ZOOM H4n* [5].

L'utilisation du micro canon nous permet d'isoler les événements sonores du magma urbain. Inversement, pour les textures, il nous permet d'éviter les événements sonores proches du preneur de son, chose impossible avec un micro omnidirectionnel. 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.

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

5

SPEED SOUND FINDING : UNE INTERFACE D'EXPLORATION DU CORPUS DE SONS

Introduction

Dans cette partie nous présentons l'étape de création de l'interface d'exploration *Speed Sound Finding*. Nous expliquons en quoi cette interface est nécessaire à l'expérience de synthèse, détaillons son développement et présentons une expérience visant à tester ses performances.

5.1 LA NÉCESSITE D'UNE INTERFACE D'EXPLORATION

Notre préoccupation est de contrôler au maximum l'influence du processus de synthèse sur le sujet. Parmi les systèmes influents, nous avons identifié l'outil de sélection des sons. Contrairement à des expériences de catégorisation ou de verbalisation classiques, où le sujet, dans sa tâche, n'est limité que par sa propre connaissance du monde, notre expérience ne propose qu'un corpus sonore par essence restreint. Ainsi, le sujet doit être capable :

1. de rechercher rapidement un son en particulier, sans connaissance préalable du corpus sonore disponible
2. d'appréhender rapidement la taille et la diversité du corpus sonore qui lui est proposé

5.1.1 *Le problème d'une recherche par mots clefs*

Il s'agit de penser une interface permettant d'explorer efficacement notre banque de sons. Dans la plupart des cas, une interface d'exploration de banques de sons fonctionne par mots clés. L'utilisateur rentre un mot, caractérisant selon lui le son qu'il a en tête, et l'interface lui présente les sons effectivement décrits par ce mot.

L'efficacité de ce principe repose avant tout sur la structure typologique et la nomenclature de la base de données. Pour notre expérience cela pose trois problèmes majeurs :

1. 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 à la fois précise de chaque son, qui plus est adaptable à la représentation que s'en fait chaque sujet. Ce qui est difficilement envisageable pour nous.

2. 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 visant une diffusion internationale, cette contrainte pose problème.

3. 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 ».

5.1.2 *Explorer une banque de sons à l'aveugle*

Pour limiter l'influence de l'interface sur le sujet, il nous apparaît 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.

Les éléments sonores de notre base de données sont hiérarchisés en classes et sous classes, et disposés sur un plan en deux dimensions. Cette organisation se fonde sur des principes perceptifs et cognitifs. Les classes ont été établies à partir de la littérature traitant de la perception des sons environnementaux urbains (cf. section 4.2). Leur étude nous a permis d'identifier les catégories perceptives les plus fréquemment relevées par les chercheurs. Nous les avons adaptées à notre base de données.

Chaque classe possède un son prototype. Aux derniers niveaux hiérarchiques, nous dévoilons l'ensemble des échantillons relatifs à la sous-classe considérée.

Visuellement, les classes, sous classes et échantillons sonores sont représentés par des cercles. La disposition des cercles dépend de l'organisation hiérarchique de la base de données. Lorsqu'on "clique" sur un cercle, le prototype associé à la classe est joué.

Précisons ici que l'interface donne accès à deux espaces dont l'un est dédié aux textures, l'autre aux évènements (cf. section 4.1).

5.1.3 *La création de l'interface*

L'interface étant destinée à être reliée à l'application web sceneSynth, elle est entièrement programmée en javascript. Nous utilisons les fonctions de *circle packing* de la bibliothèque *D3.js* afin de rendre compte de l'architecture (cercles hiérarchisés) de notre

base de données. Pour que l'interface puisse communiquer avec sceneSynth, nous utilisons la bibliothèque *angular.js* (gestion dynamique et synchronisation des données).

Nous avons développé deux versions de l'interface. La première (interface 1) laisse apparaitre les niveaux hiérarchiques intermédiaires, par exemple *circulation*, *travaux* ou *voiture*, tandis que la deuxième (interface 2) n'affiche que les éléments du plus bas niveau (i.e. les sons composant le corpus) comme *passage-voiture*, *marteau-piqueur* etc. (cf. figure 11).

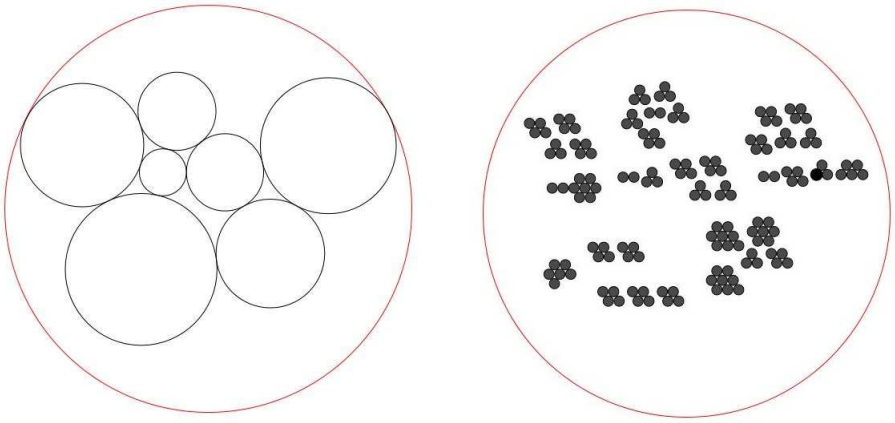

FIGURE 11: Les deux versions de l'interface *Speed Sound Finding* : à gauche l'interface 1 et à droite l'interface 2

Les deux interfaces présentent les mêmes regroupement. Ainsi, si on déplie entièrement les niveaux hiérarchiques de l'interface 1, on retombe sur la configuration spatiale de l'interface 2 (cf. figure 12).

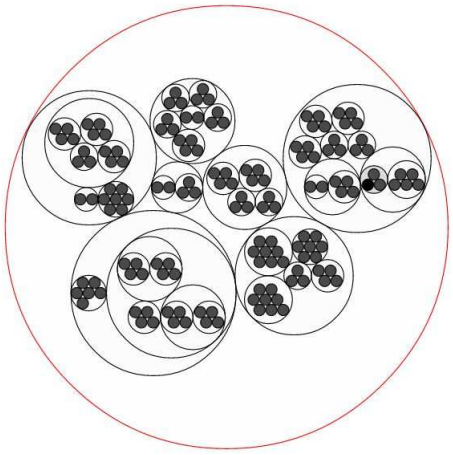

FIGURE 12: L'interface 1 dépliée de *Speed Sound Finding*

5.2 LE TEST DE L'INTERFACE

5.2.1 *Hypothèses de l'expérience*

Afin de valider ces choix nous avons mis au point une expérience au travers de laquelle nnous avons voulu tester et comparer les performances de trois interfaces :

- L'interface 1 qui propose une organisation par classes sémantiques hiérarchisées
- L'interface 2 qui propose une organisation par classes sémantiques non-hiérarchisées
- L'interface 3 qui propose une classification sur la base d'une analyse par descripteurs acoustiques de type MFCC accompagnée d'un positionnement multidimensionnel (*Multidimensional scaling*) (cf. figure 13).

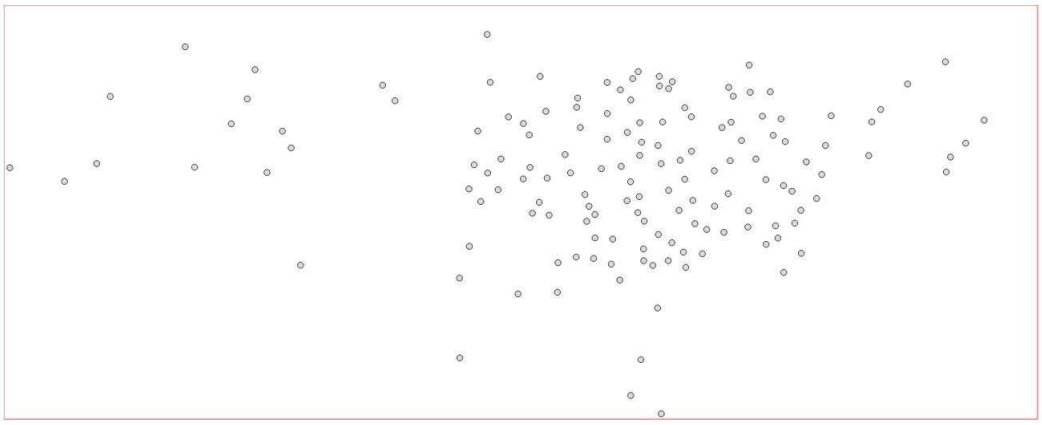

FIGURE 13: L'interface 3 : classification sur la base d'une analyse par descripteurs acoustiques de type MFCC

En confrontant l'interface 1 à l'interface 2, nous avons mis en évidence l'effet d'une hiérarchie imposée à l'utilisateur. En confrontant les interfaces 1 et 2 à l'interface 3, nous avons observé les performances rendues possibles par la classification basée sur les paramètres perceptifs d'une part (interfaces 1 et 2), la classification basée sur l'analyse acoustique d'autre part (interface 3).

5.2.2 *Cadre de l'expérience - Une approche crowdsourcing*

Cette expérience est conçue pour être réalisée entièrement via un navigateur internet. Le but étant d'apprécier la performance d'une interface dans les conditions normales d'utilisation, c'est le choix que nous avons fait.

Approche *crowdsourcing* oblige, les sujets n'ont pas été recrutés suivant des critères particuliers. Nous avons envoyé le lien de l'expérience via les listes de diffusion : *music-ir*, *auditory* et *uuu-IRCAM*.

5.2.3 *Protocole expérimental*

Nous proposons au sujet de retrouver successivement 13 sons cibles à l'intérieur d'une base de données sonores composée de 149 sons environnementaux urbains, tous normalisés au même niveau RMS.

Les sons cibles sont repartis de telle sorte qu'il y en ait au moins deux dans chaque classe du premier niveau hiérarchique de l'interface 1. Afin de limiter les effets d'ordre, nous présentons les sons cibles aléatoirement pour chaque sujet.

Pour écouter le son cible, le sujet clique sur le bouton *Play target sound*. Au moment où, sélectionnant un cercle, il commence sa recherche, un chronomètre démarre. Il peut, s'il le souhaite, réécouter le son cible.

Lorsque le son cible est trouvé, il met un terme à sa recherche en cliquant sur le bouton *Click if found*. Le chronomètre s'arrête. Le son cible suivant est chargé automatiquement.

Deux informations sont communiquées au sujet pendant l'expérience :

- l'état de la recherche (en pause ou en cours)
- le nombre de sons cibles restant à trouver

L'expérience prend fin une fois que tous les sons cibles ont été identifiés.

Le sujet réalise l'expérience sur une seule interface afin d'éviter tout effet d'apprentissage.

5.2.4 *Données observées*

Pendant l'expérience nous relevons pour chaque sujet les paramètres suivant :

- la durée de chaque recherche. La somme de ces durées, pour un sujet donné, constitue la durée effective.
- la durée totale de l'expérience. Appelée aussi durée absolue, elle inclue les pauses entre chaque recherche.
- le nom et le moment de l'écoute pour chaque son entendu lors d'une recherche.

5.3 PRÉSENTATION DES RÉSULTATS DU TEST

Dans cette partie nous présentons les résultats du test de performance de l'interface de sélection. Nous avons collecté 60 résultats lors de l'expérience, soit 20 sur chacune

des interfaces.

5.3.1 *Détection des valeurs aberrantes*

Avant d'analyser les résultats, nous en retirons les valeurs aberrantes(*outlier*). Une valeur aberrante est une valeur significativement éloignée des autres membres du tirage dont elle fait partie. L'identification et la suppression de ces données pose problème, la notion de valeur aberrante n'admettant pas de définition mathématique ou statistique fixe. Une valeur aberrante peut relever d'une erreur de mesure ou d'enregistrement. Elle peut tout aussi bien être une valeur légitime, une valeur "tout simplement - et par hasard - extrême." (voir Howell, 1998).

Supprimer ces données est nécessaire s'agissant d'une expérience de type *crowdsourcing*. L'absence de contrôle sur le déroulement du test conduit inévitablement à des résultats extrêmes, résultats dus le plus souvent à un relâchement, une distraction en cours d'expérience (nous ne pouvons empêcher que certains sujets se prêtent au test tout en ayant l'esprit par ailleurs sollicité, radio, télé. . . , ou encore s'interrompent, par exemple, pour répondre au téléphone (voir Komarov et al., 2013)), plus rarement à la tentation de tricher (voir Buchholz and Latorre, 2011).

Ces résultats sont susceptibles de fausser l'analyse. Pour les détecter, la méthode habituellement utilisée dans les recherches en interactions hommes-machines est de considérer une observation comme aberrante si celle-ci s'éloigne de plus de ± 2 déviations standards de la moyenne. Cette méthode n'est cependant pas robuste à la présence de données extrêmes. Nous lui préférons la méthode présentée dans Komarov et al. (2013), et utilisant l'écart inter-quartile (inter-quartile range :IQR). Cette méthode propose de considérer qu'une donnée est aberrante si elle est supérieure de 3*IQR au troisième quartile, ou inférieure de 3*IQR au premier quartile. Pour des valeurs normalement distribuées, le procédé supprime moins de 0.00023% des données, contre 4.6% avec la méthode précédente.

Pour chaque interface, nous appliquons la méthode à divers paramètres :

- durée moyenne de recherche par sujet
- durée maximale de recherche par sujet
- durée effective de recherche par sujet
- durée absolue de recherche par sujet
- nombre total de sons entendus par sujet
- nombre moyen de sons entendus par sujet
- nombre maximal de sons entendus par sujet

Nous l'appliquons également à certains paramètres variant suivant la position de la recherche, à savoir :

- durée de la recherche suivant les sujets
- nombre de sons entendus lors de la recherche suivant les sujets

Une fois la méthode appliquée, 4 candidats sur les 60 sont détectés comme étant des *outlier*, dont 1 sur l'interface 1, 1 sur l'interface 2, et 2 sur l'interface 3.

Sur ces 4 sujets :

- 2 sont repérés sur la durée absolue (respectivement 5h et 14h de temps d'expérience)
- 1 sur le nombre total de sons entendus (environ 1800 sons, plus de 12 fois la taille du corpus sonore)
- 1 sur le nombre de sons entendus lors de la première recherche (321 sons dont 21 fois le son cible recherché)

Les tableaux 1, 2 et 3 nous permettent d'apprécier l'effet de cette méthode sur l'écart type pour respectivement le nombre de sons entendus, la durée effective et la durée absolue suivant les sujets.

TABLE 1: Variation de l'écart type du nombre de sons entendus par sujet, avec et sans données aberrantes

	Interface 1	Interface 2	Interface 3
Avec données aberrantes	141	155	315
Sans données aberrantes	139	140	146

TABLE 2: Variation de l'écart type de la durée effective par sujet, avec et sans données aberrantes

	Interface 1	Interface 2	Interface 3
Avec données aberrantes	353	277	520
Sans données aberrantes	363	249	273

TABLE 3: Variation de l'écart type de la durée absolue par sujet, avec et sans données aberrantes

	Interface 1	Interface 2	Interface 3
Avec données aberrantes	1401	339	4034
Sans données aberrantes	408	327	273

5.3.2 *Efficacité de la recherche*

Nous réalisons l'analyse sur l'ensemble des résultats moins les *outlier*. Nous dénombrons 19 sujets viables pour l'interface 1, 19 pour l'interface 2 et 18 pour l'interface 3.

Afin d'avoir un premier aperçu des performances, nous observons les résultats pris sur l'ensemble des sujets pour chacune des interfaces. Le tableau 4 nous permet d'observer les moyennes et les écarts types, suivant les sujets, pour différents paramètres. Tant au niveau du temps qu'au niveau du nombre de sons entendus, l'interface 2 semble être la plus performante.

TABLE 4: Moyennes et écarts types suivant les sujets pour le nombre de sons entendus, le nombre de sons entendus sans répétition et les durées de recherche effectives

	Moyenne/écart type du nombre de sons entendus		Moyenne/écart type du nombre de sons entendus sans répétition		Moyenne/écart type des durées de recherche effectives (sec)	
Interface 1	498	139	130	10	847	363
Interface 2	367	140	132	11.5	564	249
Interface 3	526	146	142	7.5	823	273

Il est cependant difficile de comparer directement les moyennes, ces dernières n'étant pas robustes aux données extrêmes, généralement présentes sur des expériences en interaction Homme-Machine. La figure 14 nous permet d'apprécier la distribution des données. Nous observons que les résultats de l'interface 2 sont significativement éloignés de ceux des interfaces 1 et 3, tant en nombres de sons entendus qu'en durées de recherche effectives. L'interface 2 semble être la plus efficace.

Néanmoins, les résultats ne différent pas significativement entre l'interface 1 et 3. Le test de la somme des rangs de Wilcoxon [1] (cf. annexe A) vient confirmer ce fait aussi bien pour le nombre de sons entendus ($p = 0.37$) que pour le temps effectif ($p = 1$). L'interface 1 n'est donc pas plus efficace que l'interface 3.

1. Le test de la somme des rangs de Wilcoxon est un test statistique non paramétrique particulièrement sensible aux différences de tendances centrales entre deux populations. (cf. annexe A)

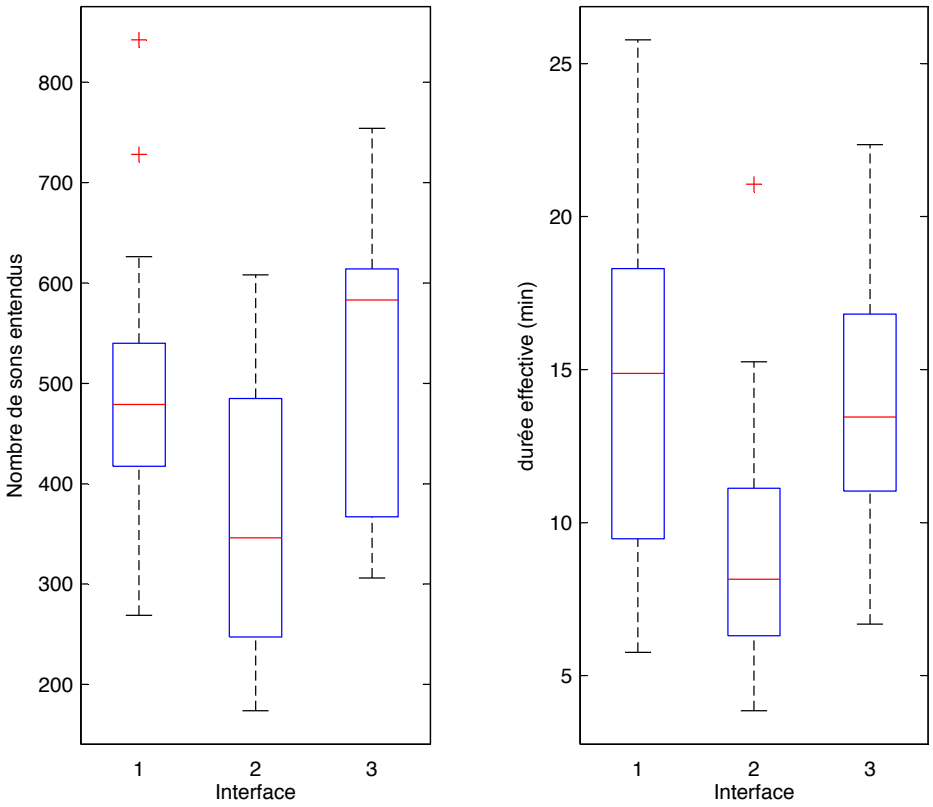

FIGURE 14: Dispersion des données pour la durée effective de recherche : la boite bleue correspond à l'écart inter-quartile, la ligne rouge à la médiane, les croix rouges aux données aberrantes.

Afin de tester la sélectivité de la recherche, nous nous proposons d'observer le nombre de sons entendus sans répétition (cf. figure 15). En d'autres termes, si un sujet écoute 15 fois un même son d'oiseau durant les 13 recherches, ce dernier ne sera comptabilisé qu'une fois. Pour un nombre faible de sons sans répétition, nous supposons que le sujet s'est servi de l'organisation spatiale de l'interface pour améliorer l'efficacité de sa recherche. A l'inverse, pour un nombre élevé de sons sans répétition, nous considérons que le sujet n'a pas compris la manière dont les sons sont repartis dans l'espace. Le maximum de sons entendus sans répétition possible équivaut à la taille du corpus : 149 sons.

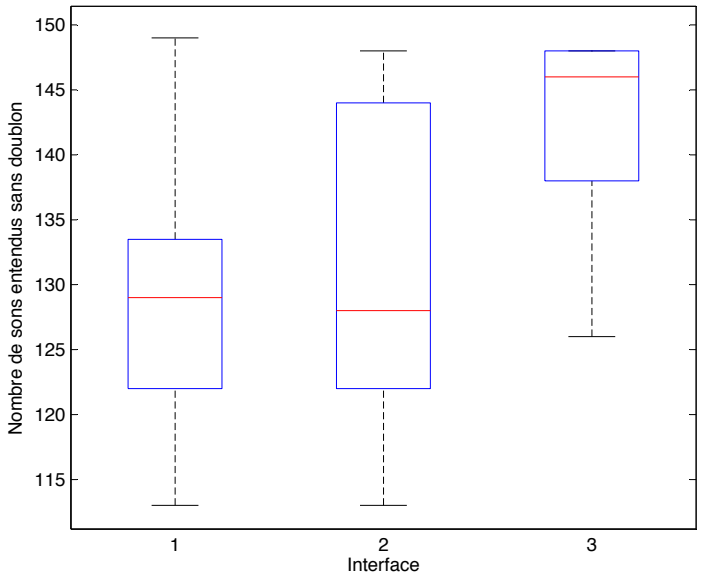

FIGURE 15: Dispersion des données pour le nombre de sons entendus sans répétition : la boite bleue correspond à l'écart inter-quartile, la ligne rouge à la médiane.

Nous observons que les résultats de l'interface 3 sont moins bons que ceux des interfaces 1 et 2. Pour l'interface 3, 75% des sujets ont entendu plus de 138 sons et 25% ont entendu 148 sons soit presque la totalité de la base de données. Pour ce qui est de l'interface 1, 75% des sujets ont entendu moins de 133 sons, contre 144 sons pour l'interface 2. Nous en déduisons que le fait d'imposer visuellement au sujet une hiérarchisation facilite sa connaissance de la base et de son organisation, connaissance par laquelle il optimise rapidement sa recherche. Néanmoins le test de la somme des rangs de Wilcoxon ne permet pas de rejeter l'hypothèse nulle A entre l'interface 1 et 2 (p=0.86).

5.3.3 *Phénomène d'apprentissage*

Nous souhaitons observer la manière dont le sujet s'approprie l'organisation de l'interface. Pour ce faire nous nous proposons d'observer, recherche par recherche, les durées et le nombre des sons entendus. Sur la figure 16 nous affichons sur chaque courbe une régression polynômiale des médianes, afin d'apprécier la tendance globale, ainsi que les premiers et troisièmes quartiles.

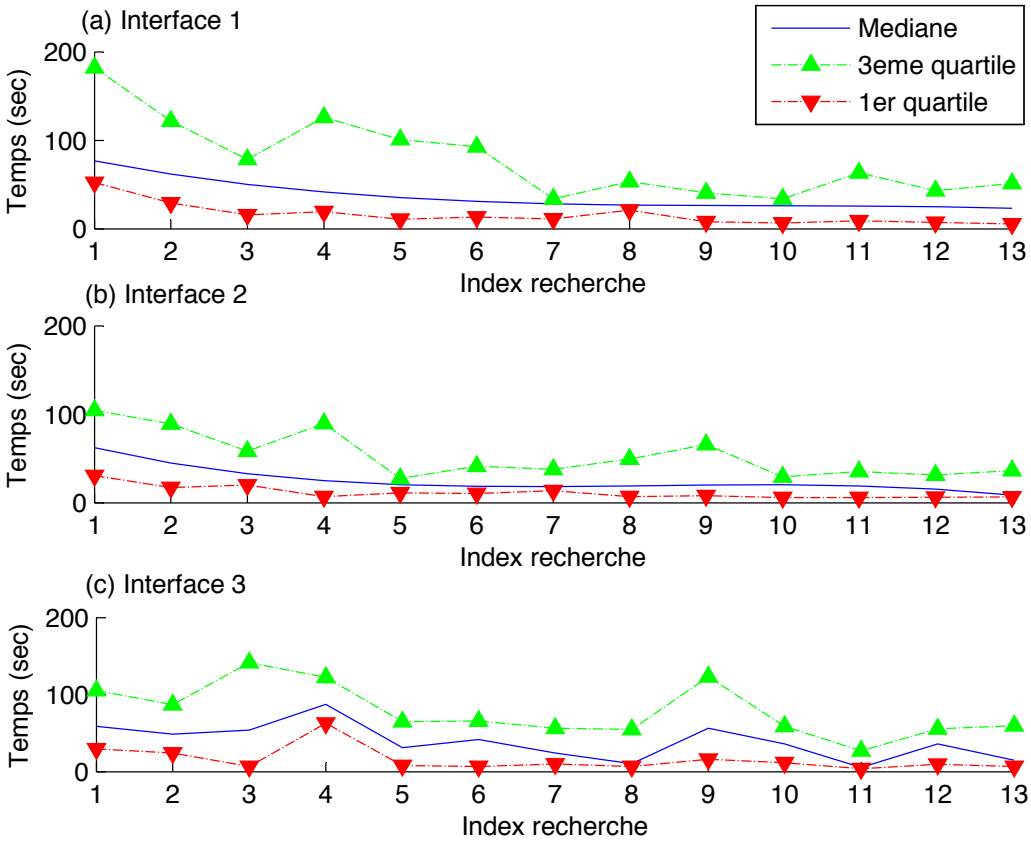

FIGURE 16: Premier quartile, troisième quartile et régression polynomiale des valeurs médianes pour les durées de chaque recherche

Il est intéressant de constater que, pour l'interface 1 comme pour l'interface 2, le temps de recherche maximum est observé lors de la première recherche tandis qu'il est observé lors de la quatrième pour l'interface 3. Cela nous conforte dans l'idée qu'une organisation graphique des sons, basée sur des paramètres perceptifs plutôt qu'acoustiques permet d'acquérir rapidement une connaissance de la base de données, permettant ainsi un meilleur apprentissage.

La figure 17 affiche les valeurs moyennes et médianes sans régression des durées. Au vu de l'allure des courbes, les sujets semblent comprendre plus rapidement l'organisation de l'interface 2 que celle de la l'interface 1. En effet, la courbe de l'interface 2 décroît jusqu'à la sixième recherche avant d'osciller sur des valeurs comprises entre 26 et 55 secondes. La courbe de l'interface 1 ne décroit que jusqu'à la troisième recherche et oscille ensuite entre 26 et 86 secondes. Ces observations plaident en faveur de l'interface 2.

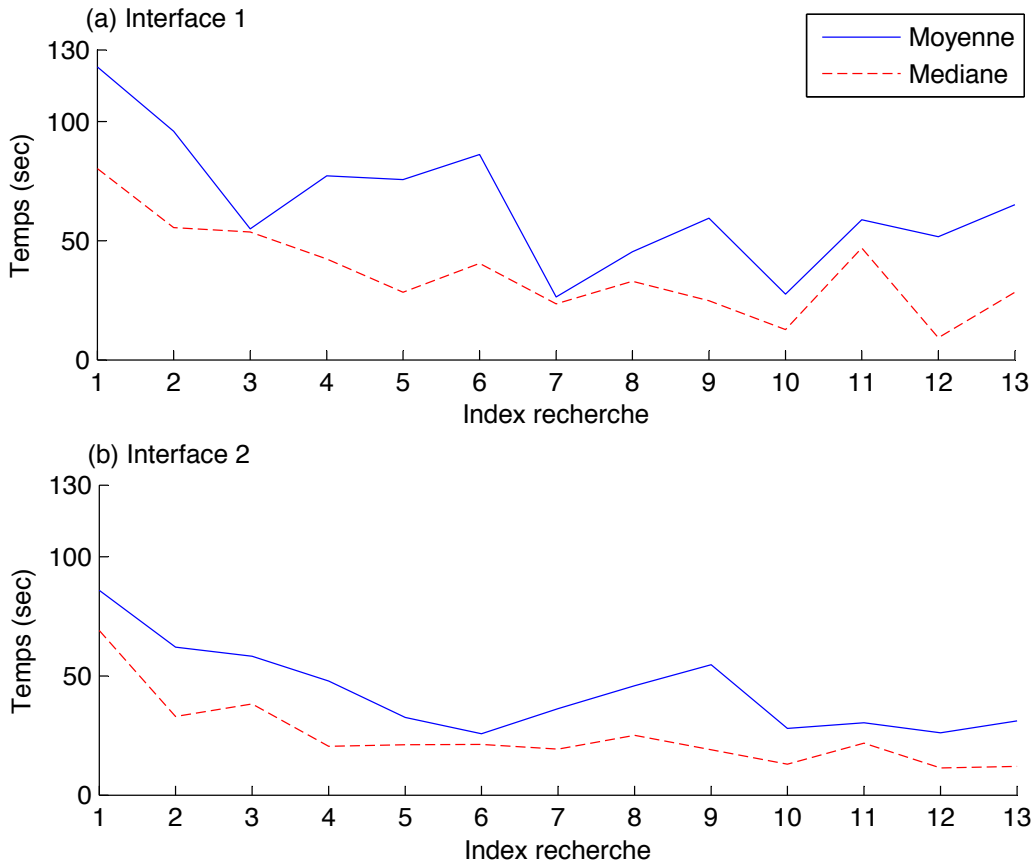

FIGURE 17: Valeurs moyennes et valeurs médianes des durées de chaque recherche

5.4 DISCUSSION

5.4.1 *Performance des interfaces*

Suite à ce test, il apparaît que l'interface 2 est la plus efficace, suggérant ainsi qu'une organisation des données sur la base de paramètres perceptifs et sémantiques est plus performante qu'une organisation à partir de descripteurs acoustiques.

Les interfaces 1 et 3 présentent des résultats d'ordre similaire en temps et en nombre de sons écoutés. Néanmoins l'interface 1 permet une recherche plus sélective. Ce fait est potentiellement dû à la taille réduite de la base de données utilisée. Il serait intéressant de mener une expérience similaire sur une banque de sons comprenant par exemple 1000 éléments, un tel nombre ne permettant plus de parcourir "au hasard" les échantillons sonores.

5.4.2 *L'approche crowdsourcing*

Il nous semble important de tirer ici des conclusions sur différents points spécifiques à l'approche de type *crowdsourcing*. De part la nature de l'outil de synthèse *SceneSynth* (cf. section 3.4), cette approche constitue une composante importante du sujet de ce stage. C'est par ailleurs une méthodologie relativement nouvelle dans le domaine de la psychologie expérimentale qu'il est intéressant d'évaluer.

5.4.2.1 *Le temps de préparation*

Pour un même protocole, le *crowdsourcing* requiert un temps de préparation sensiblement plus important qu'une expérience réalisée en laboratoire.

Premièrement, un temps de programmation est nécessaire, côté serveur d'une part, afin d'héberger l'expérience et de permettre la sauvegarde des données, côté utilisateur d'autre part, afin de prévenir certains effets dommageables inhérents à l'environnement javascript comme le rafraichissement de la page.

Deuxièmement, l'utilisateur n'étant pas en contact direct avec l'expérimentateur, il faut s'assurer que l'interface ainsi que les instructions soient d'une approche aisée, toute incompréhension ou difficulté pouvant rebuter le sujet.

Enfin, les précédents points doivent être abondamment testés.

Nous précisons cependant que ce temps de préparation est en partie compensé par le fait que l'expérience, une fois lancée, ne requiert plus l'attention directe de l'expérimentateur.

5.4.2.2 *La participation*

Nous avons lancé l'expérience sur 3 listes de diffusion. Les volontaires n'ont pas été aussi nombreux qu'escompté, et il a fallu pas moins de 18 jours pour récolter 60 réponses exploitables. Cette faible participation est due en partie au fait que les sujets n'étaient pas rémunérés. Notons qu'il existe plusieurs plateformes comme *Mechanical Turk*[2] qui proposent de diffuser des expériences à un grand nombre de sujets, payés suivant contribution. Le coût reste très inférieur à celui d'une expérience classique, en argent comme en temps. Il y a cependant risque que les données soient bruitées, certains sujets motivés essentiellement par le gain, expédiant l'expérience (voir Buchholz and Latorre, 2011).

La figure 18 nous montre le nombre de données récupérées en fonction du temps (jours). Nous remarquons que 40% des données sont arrivées le premier jour.

2. Pour plus de détails sur *Mechanical Turk* voir https://www.mturk.com/mturk/welcome

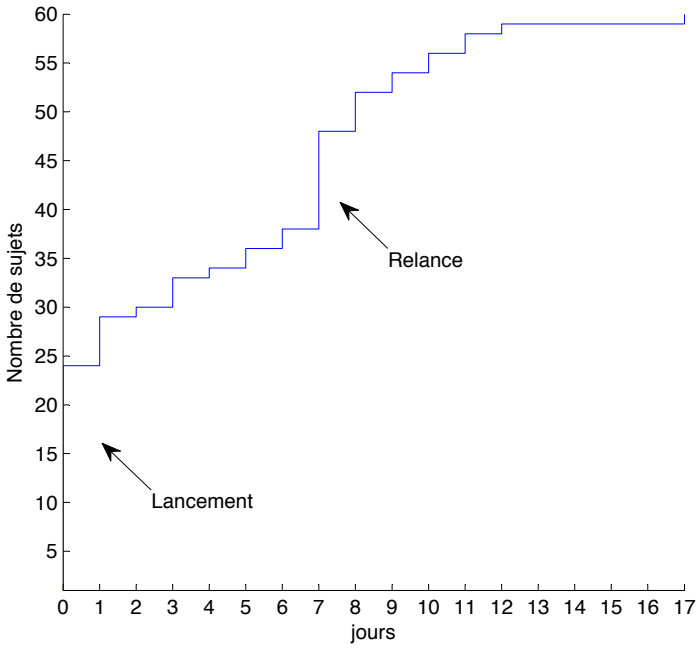

FIGURE 18: Données en fonction du temps (jours)

Nous relevons un total de 13 abandons avérés en cours d'expérience, soit 21% des données utilisables. Par "abandon avéré" nous entendons, un sujet ayant précisé son nom, ayant passé plus de 3 minutes sur l'interface, et n'ayant réalisé qu'une fois l'expérience (avec ou sans succès).

L'expérience est accessible à l'adresse suivante : `http://217.70.189.118/soundthings/speedSoundFinding`. Elle requiert le navigateur chrome pour fonctionner.

6

L'EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS

INTRODUCTION

Dans cette partie, nous présentons l'expérience pilote de synthèse. Cette expérience constitue l'objectif premier du stage. Il s'agit ici de montrer la viabilité de notre approche. L'expérience a par ailleurs nécessité le développement préalable de deux outils, le corpus de sons, et l'interface *Speed Sound Finding* précédemment présentés. Afin de faciliter la lecture de cette partie, nous résumons les points importants touchant à l'aspect pratique de l'expérience, ainsi que les hypothèses testées. Nous terminons par le détail des résultats.

6.1 PRÉSENTATION DE L'EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS

A travers cette expérience, nous souhaitons étudier la manière dont nous nous représentons mentalement les sons de la ville. Pour ce faire nous proposons au sujet de re-synthétiser un paysage sonore urbain, à partir d'un corpus de sons environnementaux. La synthèse est réalisée à l'aide de l'outil SceneSynth

Le corpus a été conçu et structuré afin d'en faciliter l'exploration (cf. chapitre 4). Il est décomposé en deux grandes familles de sons :

- la famille des évènements sonores
- le famille des textures sonores

Ces familles sont organisées en classes hiérarchisées. Lesdites classes sont issues d'études menées sur la perception des sons environnementaux. Chaque classe regroupe des éléments perceptivement liés. Les classes de niveau hiérarchique le plus bas sont dites collections. Elles regroupent des fragments sonores sémantiquement identiques (cf. chapitre 4).

Les sons sont issus :

- d'enregistrements réalisés en cours de stage
- de deux banques de sons existantes : *SoundIdeas* et *Universal SoundBank*

Tous les sons sont monophoniques et présents au format *ogg* dans la base de données.

Nous utilisons l'environnement de synthèse séquentielle par corpus *SceneSynth* (cf. chapitre 3). Afin de permettre à l'utilisateur d'interagir avec la banque de sons sans l'aide d'informations textuelles, nous avons conçu l'interface d'exploration *Speed Sound Finding* et l'avons adaptée à *SceneSynth* (cf. chapitre 5).

Bien que *SceneSynth* soit prévu pour être utilisé dans une approche *crowdsourcing*, nous avons préféré, dans le cadre de l'expérience pilote, faire passer les sujets en laboratoire. Ceci afin d'assurer un contrôle sur l'environnement expérimental et de garantir un nombre donné de sujets à une date fixée.

Nous avons sélectionné 10 sujets sur la base du volontariat. Tous les sujets habitent dans un milieu urbains de type grande ville depuis au moins 1 an.

L'expérience a lieu au sein des locaux de l'équipe "Perception et design sonore" de l'IRCAM. Les sujets sont placés dans des cabines audio-métriques isolées, comportant un ordinateur *Macintosh Mac Book Pro* relié à une carte-son *RME FireFace 800* . L'outil *SceneSynth* est utilisé via le navigateur internet Chrome. Le son est diffusé en mono via des enceintes *Yamaha MSP 5* (enceintes actives).

L'expérience de synthèse est accessible via le lien suivant : `http://217.70.189.118/soundthings/SceneSynth-SSF/`. Elle requiert l'utilisation du navigateur internet chrome pour fonctionner.

6.2 PROTOCOLE EXPÉRIMENTAL

Nous demandons aux sujets de synthétiser successivement deux paysages sonores (cf. section 1.2) :

- un paysage sonore idéal
- un paysage sonore non-idéal

Nous entendons par idéal (respectivement non-idéal), le paysage sonore le plus apprécié du sujet (respectivement le moins apprécié). Nous demandons au sujet de respecter deux points :

- 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 100 millisecondes

Chaque processus de synthèse comprend plusieurs parties :

1. la réalisation de la synthèse

2. l'indexation des éléments sonores introduits (à faire pendant l'étape de synthèse)
3. l'indexation du paysage sonore synthétisé
4. le commentaire libre du paysage sonore synthétisé
5. l'indication des éléments sonores manquants

Une fois les deux scènes sonores réalisées, le sujet est invité à commenter l'ergonomie de l'interface de synthèse *SceneSynth*, ainsi que l'interface de sélection *Speed sound finding*.

Un temps de prise en main de 10 minutes est prévu en début d'expérience. Notons que les sujets sont avertis de la nature des scènes à synthétiser juste avant de commencer.

Le tableau 5 résume les étapes de l'expérience ainsi que leurs durées respectives.

TABLE 5: Déroulement de l'expérience pilote de synthèse séquentielle par corpus

Tache	Durée (minutes)
Présentation de l'expérience de synthèse / Consignes	10
Prise en main de l'outil de synthèse *sceneSynth*	10
Première Synthèse	20
Commentaire de la première Synthèse	10
Deuxième Synthèse	20
Commentaire de la deuxième Synthèse	10
Critique de l'interface	10
Total	1h30

Les documents relatifs à l'expérience peuvent être consultés aux annexes B, C, D et E.

6.3 LES DONNÉES

Les données que nous sauvegardons sont de deux types : numériques et textuelles.

Pour les données textuelles, nous relevons les tags des échantillons sonores sélectionnés par l'utilisateur, les noms donnés par lui à ces échantillons, ainsi qu'à la scène entière. Nous nommons "tag" le vocable sous lequel chaque échantillon est référencé dans le corpus, et "nom" le mot utilisé pour désigner le nom donné par le sujet .

A titre indicatif, nous demandons aussi au sujet de commenter la scène sonore réalisée, à savoir, décrire le résultat final, justifier du choix de tel échantillon, des modifications, et apporter toute autre précision jugée par lui utile, comme en particulier les

objets sonores absents selon lui de la base de données.

Concernant les données numériques, il s'agit des paramètres de contrôle que l'utilisateur applique aux sons choisis. On y trouve le niveau sonore en dB, la position dans la scène ainsi que la pente d'apparition/disparition du son (*fade in/out*). Un dernier paramètre est disponible uniquement pour les sons de type événement. C'est l'*occurence spacing* permettant de régler l'espacement entre deux occurrences d'un événement. Nous invitons le lecteur à se référer à la section 3.4 pour une description plus détaillée de ces paramètres.

Les deux types de données font l'objet d'un traitement séparé.

6.4 RÉSULTATS DE L'EXPÉRIENCE PILOTE

6.4.1 *Les données verbales : titres des scènes synthétisées*

Nous effectuons ici l'analyse des titres donnés par les sujets aux scènes synthétisées. Nous avons 10 titres de scènes idéales dont nous tirons 11 descriptions et 10 titres de scènes non-idéales dont nous tirons 10 descriptions. Précisons qu'un titre peut contenir plusieurs entités sémantiques distinctes. A partir de l'analyse lexicale de ces dernières, nous déduisons des catégories.

Pour les scènes idéales, nous identifions 4 catégories : "parc" (4 descriptions), "espace piéton" (3 descriptions), "cour intérieure" (2 descriptions) et "rue calme" (2 descriptions). Le fait que la catégorie "parc" soit la plus représentée est un résultat assez intuitif. Nous notons que les catégories font majoritairement référence à des lieux privés de circulation (9 descriptions), mais pas de la présence de l'homme.

Pour les scènes non-idéales, toutes les descriptions se réfèrent à la voirie. Là encore nous identifions 4 catégories : "rue" (4 descriptions), "boulevard" (2 descriptions), "avenue" (2 descriptions) et "carrefour" (2 descriptions). 3 descriptions font directement référence à des sons de travaux, et 3 autres à des sons de circulation.

La liste des titres est donnée en annexe C.

6.4.2 *Les données verbales : tags et noms*

Nous analysons ici les données verbales relatives aux tags et aux noms donnés par les sujets aux échantillons sonores sélectionnés. Nous observons d'abord les tags.

Pour les scènes idéales, les sujets ont utilisé 51 événements et 27 textures. Pour les scènes non-idéales nous relevons 96 événements et 32 textures. Ce résultat indique qu'une scène non-idéale est composée d'un nombre plus important de sources sonores qu'une scène idéale. Le tableau 6 affiche les moyennes et les écarts types du nombre

d'évènements et de textures utilisés par les sujets.

TABLE 6: Moyennes et écarts types des évènements et textures utilisés pour la synthèse des scènes idéales et non idéales

	Moyenne et écart type du nombre d'événements utilisés		Moyenne et écart type du nombre de textures utilisées	
Scène Idéale	5.1	2.4	2.7	1.1
Scène non-idéale	9.6	3.1	3.2	1.8

Nous regroupons ensuite les tags suivant leurs classes d'appartenance. Ces classes sont relatives à la structure de notre base de données (cf. figures 10 et 9). Ce regroupement s'effectue à des niveaux hiérarchiques différents selon les tags. Les figures 19 et 20 affichent les résultats pour respectivement les scènes idéales et les scènes non-idéales.

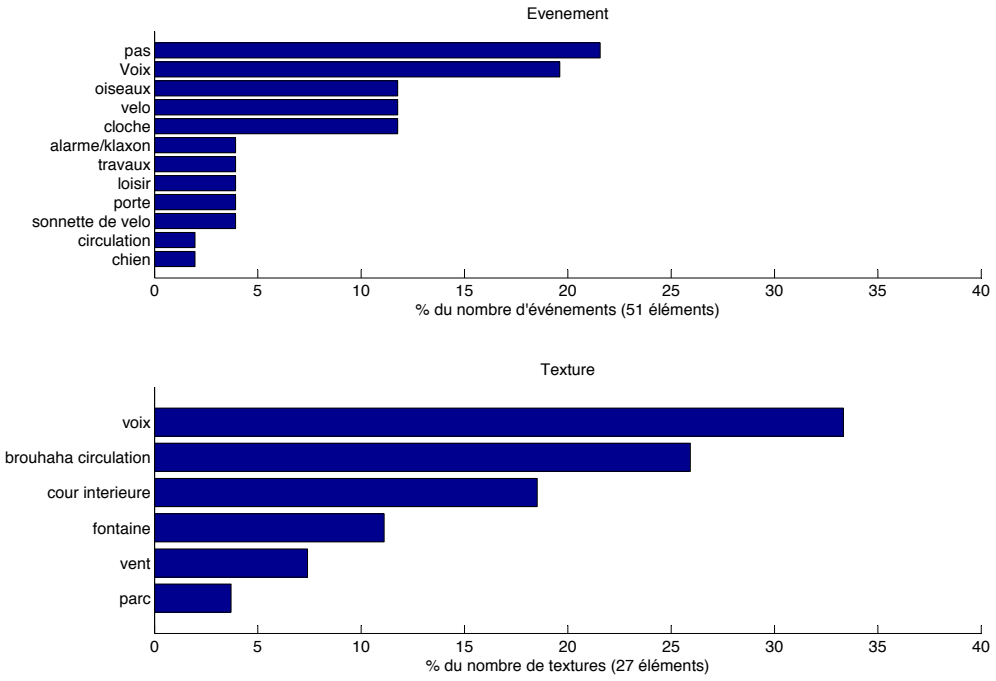

FIGURE 19: Tags relevés pour les scènes idéales

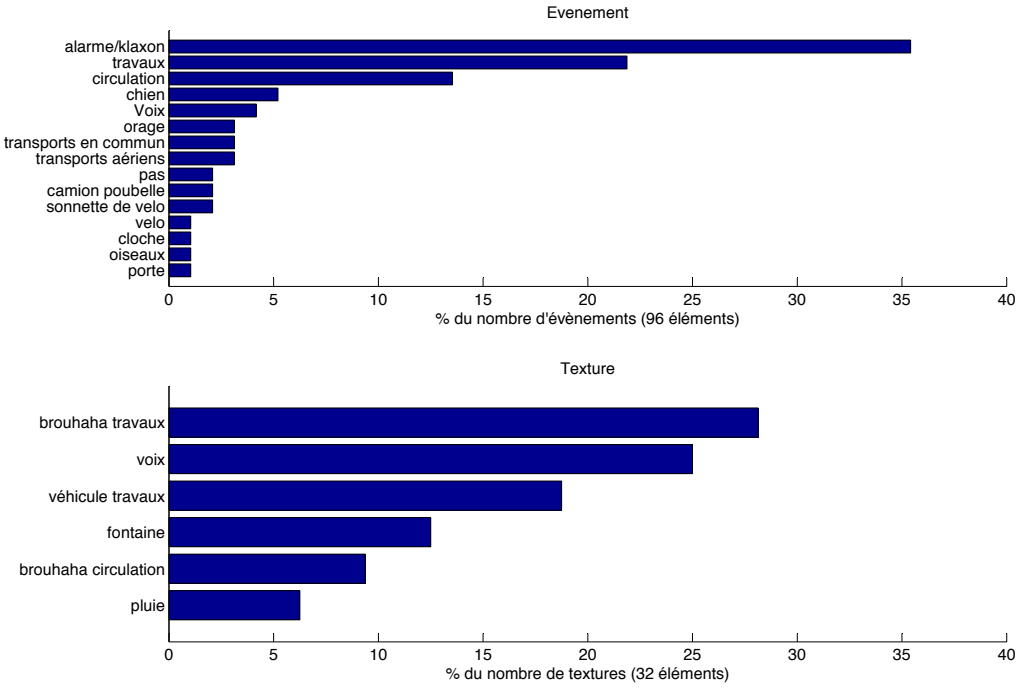

FIGURE 20: Tags relevés pour les scènes non-idéales

Les premiers résultats nous permettent d'observer une plus grande diversité dans le choix des échantillons pour les scènes non-idéales que les scènes idéales. On constate que les événements composant majoritairement les scènes non-idéales sont : des sons d'alarme et de klaxon (35% des occurrences), de travaux (22% des occurrences) et de circulation (13.5% des occurrences). Au niveau des textures, la grande majorité des sons est liée à des sons de travaux (brouhaha : 28% des occurrences, moteur de véhicules : 19% des occurrences) et aux voix (25% des occurrences).

Pour les scènes idéales, les évènements évoquent majoritairement la présence humaine : (pas : 21.6% des occurrences et voix : 19.6 % des occurrences), ainsi que les sons d'oiseaux (11.8 % des occurrences), de vélos (11.8 % des occurrences) et de cloches (11.8% des occurrences). Pour les textures, les sons d'origine humaines sont encore très présents (voix : 33.3% des occurrences).

Plusieurs résultats a priori contre intuitifs apparaissent. Dans les événements idéaux, on trouve des sons d'alarme, de travaux et de circulation. Pour les textures idéales, on observe que les "brouhahas de circulation" sont largement représentés (25.9% des occurrences). De même pour les scènes non-idéales, on observe des textures de "voix humaines" et de "fontaines", sons usuellement associés à une ambiance sonore agréable.

Afin de vérifier si ces résultats ne sont pas dus à une mauvaise identification, nous observons les noms donnés par les sujets aux éléments sonores.

Nous réalisons une analyse linguistique basique, dont l'objectif est d'attribuer à chaque nom, le tag de notre structure typologique le plus proche. Pour ce faire, nous nous appuyons sur plusieurs règles :

- Nous ne relions à un tag que les noms faisant explicitement référence à une source ("pas", "homme interpelle quelqu'un") ou à un fond sonore ("cour intérieure", "ambiance de rue")
- A chaque nom nous faisons correspondre un tag unique. Si le nom fait référence à plusieurs sources, nous lui attribuons le tag en fonction de la première source évoquée.
- Pour les sons ne pouvant être reliés explicitement à un tag, nous vérifions s'ils appartiennent à un même champ lexical. Nous avons alors deux options :
  1. Si nous détectons un champ lexical commun à plusieurs sons, nous regroupons ces derniers sous la même appellation
  2. Si nous ne détectons pas de champ lexical commun, nous éliminons de l'analyse les sons isolés. Ces sons sont alors désignés par le terme "non traité"

L'annexe E présente les correspondances linguistiques réalisées.

Nous affichons les résultats de l'analyse linguistique sur les figures 21 et 22 pour respectivement les scènes idéales et les scènes non-idéales. De manière générale, les évènements ont été bien identifiés avec 90% de correspondance directe pour les scènes idéales, et 92% de pour les scènes non idéales. L'identification s'est moins bien passée pour les textures avec seulement 48% de correspondance directe pour les scènes idéales et 47% pour les scènes non idéales. Nous remarquons de plus que 25% des noms de textures ne font pas référence à un tag, mais à une description globale de la texture.

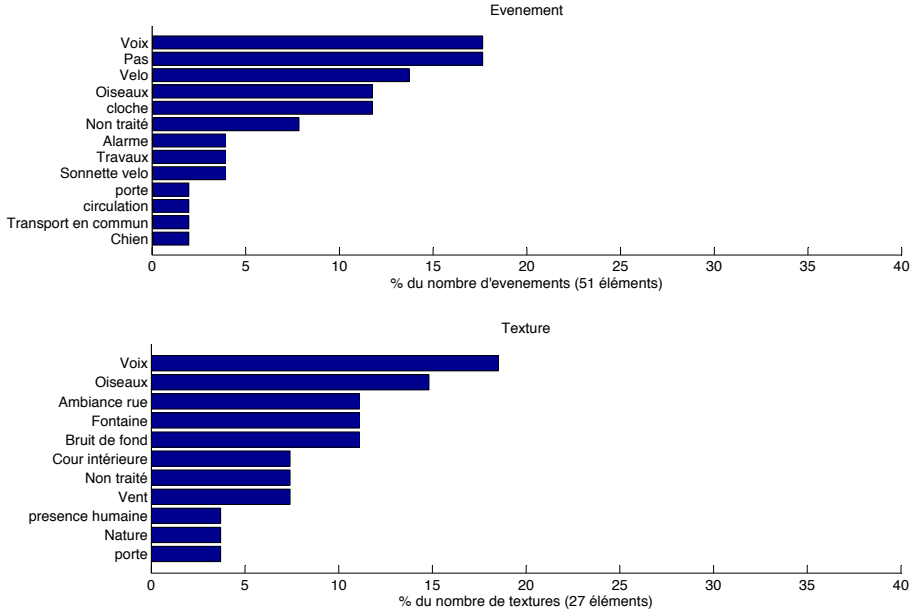

FIGURE 21: Noms relevés pour les scènes idéales

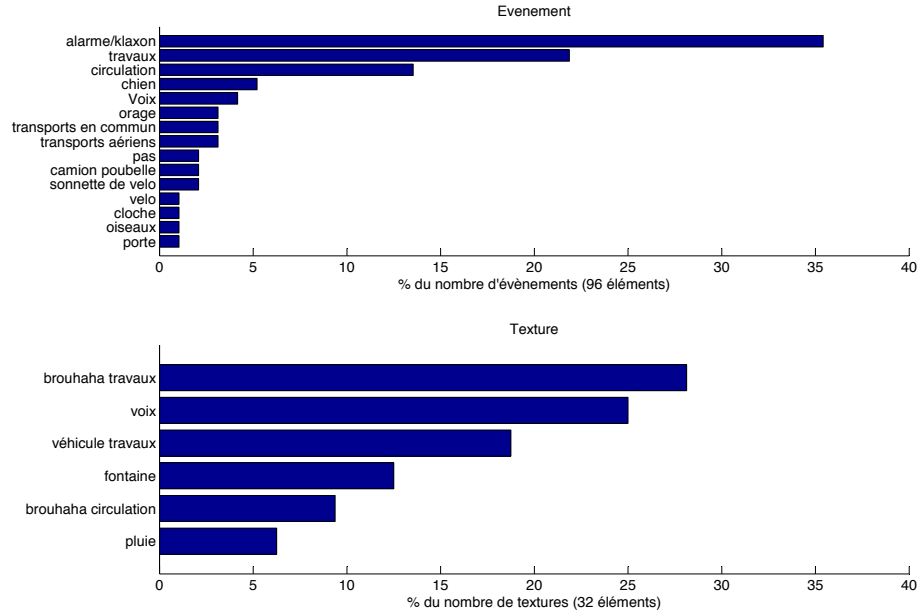

FIGURE 22: Noms relevés pour les scènes non-idéales

Ces observations vont dans le sens de Maffiolo (voir Maffiolo, 1999) qui stipule que les séquences événementielles provoquent une analyse sémantique, alors que les séquences

amorphes, dont la définition est proche de celle nos textures, sont sujettes un traitement holistique. Ce dernier s'effectuant sur la base de considérations acoustiques, il ne favorise pas l'objectivation d'un nom décrivant de manière satisfaisante la séquence.

Nous remarquons également que plusieurs noms données aux textures pour les scènes idéales comme non-idéales renvoient à des sons que nous considérons comme des évènements sonores (oiseaux, transport en commun). Le fait est qu'il est difficile d'obtenir des textures complètement dénuées de tout événement, notamment les sons d'oiseaux, omniprésents dans Paris. Il est cependant intéressant de noter que c'est l'identification de ces évènements sonores, pourtant très faiblement représentés dans les séquences de textures, qui est à l'origine de leur appellation. Cela souligne ainsi l'importance du processus d'identification des sources dans le traitement de l'information perçue.

En ce qui concerne les textures des scènes idéales, nous remarquons une nette diminution du nombre de sons relatifs au "brouhaha de circulation" entre les tags et les noms donnés aux sons. Ces sons se retrouvent sous le nom "ambiance de rue", ou "bruit de fond", indiquant ainsi que les sons liés au trafic urbain sont tolérés s'ils sont en fond sonore. Nous notons aussi un net changement au niveau des textures non-idéales. Les termes relatifs à la "circulation" sont plus présents, au dépend des sons de travaux, de facto mal identifiés.

Nous avons noté la présence d'évènements sonores appartenant à la catégorie alarme/klaxon (sons utilisés par 1 sujet) et travaux (sons utilisés par 2 sujets) dans les scènes idéales. En analysant les descriptions des sujets nous remarquons que les sons d'alarmes ont été consciencieusement choisis. Ceux-ci, d'après le sujet, rendent compte du "murmure" de la ville, "discret et agréable à entendre". Pour les sons de travaux, le premier sujet précise qu'il les a utilisés par un souci de réalisme. Le deuxième sujet indique qu'ils relèvent de l'ambiance caractéristique d'une cour intérieure, les désignant d'ailleurs sous le nom "bricolage" et non "travaux".

6.4.3 *Les données verbales : Les sons manquants*

D'après l'analyse des descriptions nous avons relevé 23 mentions de sons manquants, 14 pour les scènes idéales et 9 pour les scènes non-idéales. Parmi ceux-ci, 3 font référence à des sons musicaux et 10 à des sons présents dans la base de données que les sujets n'ont pas trouvés.Nous avons délibérément exclus les sons musicaux du corpus, notre étude portant sur les sons environnementaux.

Il nous reste donc 10 sons effectivement manquants. Ce faible nombre souligne que notre corpus sonore est bien représentatif de l'ambiance urbaine. Notons que la moitié des sujets a en effet précisé que la diversité du corpus était suffisante pour les scènes idéales (2 sujets), les scènes non idéales (2 sujets), voir les deux (1 sujet).

6.4.4 *Les données verbales : comparaison avec l'étude de Guastavino*

Afin de vérifier la cohérence de nos résultats nous nous proposons ici de les comparer à une étude réalisée par Guastavino (voir Guastavino, 2006) portant également sur les paysages sonores urbains idéaux.

Dans cette étude elle pose par mail à 77 personnes (françaises) la question suivante :

"Quelle serait pour vous l'ambiance sonore idéale d'une ville ?"

Elle collecte 257 descriptions qu'elle regroupe en deux catégories :

- les descriptions se référant à la source du son(76% des occurrences)
- les descriptions relatives au paysage sonore global (24% des occurrences)

A partir de l'analyse lexicale des descriptions relevant de sources sonores, elle identifie des catégories principales (cf. figure 23), et des sous catégories (cf. figures 24 et 25).

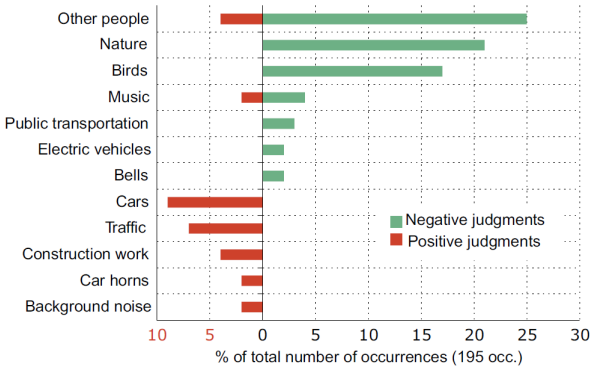

FIGURE 23: Principales catégories de sources émergeant des descriptions spontanées faites par les participants d'un paysage sonore urbain. Figure issue de Guastavino (2006)

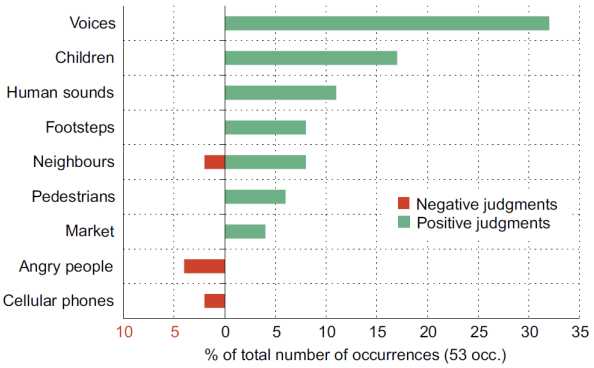

FIGURE 24: Sous-catégories de sources à l'intérieur de la catégorie principale "other people". Figure issue de Guastavino (2006)

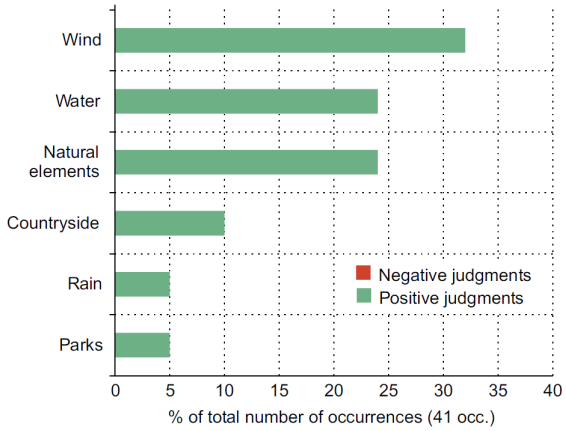

FIGURE 25: Sous-catégories de sources à l'intérieur de la catégorie principale "Nature". Figure issue de Guastavino (2006)

Nous comparons ces catégories à nos résultats obtenus sur les scènes idéales pour les tags et les noms. Les résultats que nous obtenons sur les scènes idéales concordent avec ceux de Guastavino. Dans les deux cas les catégories les plus représentées font référence à :

- des sons humains (voix et pas)
- des sons naturels (animaux,météo)

Dans les deux études, on voit apparaitre des sons de "circulation", "travaux", "alarme/klaxon" pour des villes idéales. Dans le cas de Guastavino, ces termes ont été utilisés dans des tournures de phrases négatives ("sans voitures" ou "avec moins de trafic"), tournures à partir desquelles elle infère un jugement négatif. Dans notre cas, l'analyse des descriptions faites par les sujets de leur scène (cf. annexe D) montre qu'il s'agissait avant tout de reconstruire un paysage urbain certes idéal, mais aussi plausible, l'absence de voiture étant jugé utopique pour une ville. Plusieurs sujets ont par ailleurs clairement jugé les bruits lointains de "circulation" et de "travaux" comme "agréables".

6.4.5 *Les données numériques : intensité et espacement*

Nous nous proposons ici d'analyser les données numériques, issues des paramètres de contrôles de synthèse. Nous commençons par observer les niveaux sonores. Nous affichons les moyennes et les écarts types des niveaux relevés sur les événements et textures sonores, pour les scènes idéales et non idéales (cf. tableau 7). Nous précisons que les sons sont tous fixés à un niveau relatif de 0 dB. Le sujet n'a pas la possibilité d'amplifier un son, juste de l'atténuer.

Les résultats sont plutôt intuitifs. Nous constatons que les niveaux moyens pour les scènes idéales, sont inférieurs à ceux des scènes non-idéales. De même, les niveaux des

TABLE 7: Moyennes et écarts types des niveaux sonores des évènements et textures utilisés pour la synthèse des scènes idéales et non idéales

	Moyenne et écart type du niveau des événements utilisés (dB)		Moyenne et écart type du niveau des textures utilisées (dB)	
Scène Idéale	-4.1	3.2	-9.4	6.6
Scène non-idéale	-0.9	1.1	-4	4.3

textures sont plus bas que ceux des évènements.

Nous analysons maintenant les intervalles de temps séparant les évènements sonores pour les deux types de scènes. Le tableau 8 affiche les moyennes et les écarts types de ces intervalles de temps.

TABLE 8: Moyennes et écarts types des intervalles de temps séparant les évènements utilisés pour la synthèse des scènes idéales et non idéales

	Moyenne (secondes)	écart type
Scène Idéale	10	4.8
Scène non-idéale	9.2	4

Compte tenu de la durée de la scène synthétisée (60 secondes), nous observons que les évènements sonores ont une fréquence d'apparition similaire pour les scènes idéales et non-idéales, de l'ordre de 10 secondes. Ainsi, si la ville idéale possède un niveau sonore global plus bas que la ville non idéale, elle n'en reste pas moins une ville active en terme d'évènements sonores.

6.4.6 *Critique de la fonctionnalité de synthèse de SceneSynth et de l'interface d'exploration Speed Sound Finding*

Concernant les commentaires relatifs à l'ergonomie de *SceneSynth*, tous les sujets nous ont indiqué avoir su rapidement maitriser son mode de fonctionnement. Notons que la moitié d'entre eux n'était pas familière des environnements audio-numériques. Cet apprentissage de l'interface simple et rapide est indispensable dans l'éventualité d'une diffusion de type *crowdsourcing*.

Néanmoins, plusieurs sujets nous ont indiqué avoir eu quelques difficultés dans le maniement des paramètres de contrôle. Enfin, des soucis relatifs à l'ergonomie de l'interface ont été remontés, parmi ceux-ci, l'impossibilité de pouvoir écouter le rendu de la scène sonore à partir d'un moment choisi. Le sujet est en effet obligé de l'écouter depuis le début.

En ce qui concerne l'interface de sélection *Speed Sound Finding*, 6 sujets ont soulignés sa facilité d'utilisation et/ou son caractère intuitif. Au niveau de la répartition spatiale des éléments sonores, 5 sujets ont explicitement indiqués avoir bien compris la nature des regroupements. 1 sujet a cependant précisé n'avoir pas complètement compris comment les sons étaient organisés spatialement. 2 sujets ont indiqués que l'interface permet un "rapide apprentissage de la localisation des sons" et 1 sujet a reconnu que le fait de n'avoir "aucune indication sur le son produit à l'avance, pousse à tous les essayer".

6.4.7 *Discussion*

La comparaison entre nos résultats et ceux de Guastavino montre que notre approche permet d'obtenir des résultats cohérents. Nous notons néanmoins une différence. tandis que l'approche de Guastavino permet d'envisager les représentations mentales d'un environnement sonore urbain dans sa globalité, notre expérience permet d'affiner ces représentations sur la base de considérations écologiques.

De fait, motivé par un souci de réalisme, certains sujets ont volontairement placé dans leurs scènes idéales des événements sonores a priori désagréables. Nous observons également cette tendance au niveaux des textures des scènes idéales dont 25.9% sont des sons relatifs au brouhaha de la circulation.

Par ailleurs, plusieurs sujets nous ont stipulé avoir délibérément supprimé des éléments de leurs paysages afin de n'en pas surcharger le rendu sonore, et ce, particulièrement en ce qui concerne les scènes sonores idéales. Ce fait vient renforcer la validité écologique de notre protocole expérimental. Cette suppression est en effet le fruit d'une prise de conscience, par le sujet, d'un contexte, inhérent à sa scène, le renvoyant à une "réalité du monde". Cette réalité lui est rappelée grâce à l'écoute de la scène, écoute rendue possible par le formalisme de synthèse. La scène synthétisée fait ainsi office de miroir, permettant au sujet d'ajuster ses réponses, pendant le processus de "composition".

L'analyse des données numériques montre une différence notable entre le niveau sonore moyen d'un environnement urbain idéal et celui d'un environnement non-idéal. Néanmoins aucun des environnements ne présente une activité plus faible que l'autre en terme d'apparitions de sources sonores. Ces données montrent que notre protocole expérimental permet d'obtenir des résultats numériques cohérents.

Troisième partie

PERSPECTIVES

7

CONCLUSION ET DÉBOUCHÉS

Le stage a permis de mettre de point un protocole expérimental permettant d'objectiver les représentations mentales des environnements sonores urbains. Nous avons commencé par expliciter les théories sur lesquelles repose notre approche. Pour pouvoir réaliser l'expérience, nous avons constitué un corpus de sons environnementaux de référence, sur la base d'une typologie établie à partir d'une étude bibliographique. Afin de nous affranchir de l'influence d'une nomenclature imposée, ainsi que du manque de termes lexicaux permettant de décrire de manière satisfaisante les phénomènes acoustiques (voir Guastavino, 2006), nous avons développé une interface proposant d'explorer à l'aide des sons en présence, le corpus sonore préalablement structuré suivant des considérations perceptives. Cette interface à été testée via une expérience adoptant une approche *crowdsourcing*. Enfin, nous avons réalisé une expérience pilote, afin de juger de la viabilité de notre protocole.

Cette expérience s'est avérée concluante. En les comparant à une étude précédemment menée (voir Guastavino, 2006), nous avons vérifié la cohérence des résultats obtenus à partir des données verbales. Nous avons par ailleurs montré que notre expérience, de part son formalisme de synthèse, satisfait d'une manière originale aux conditions écologiques inhérentes à une approche cognitive.

Une analyse approfondie reste à mener sur les données numériques afin de vérifier leur pertinence. De plus, le faible nombre de sujets recrutés ne nous a pas permis de caractériser et d'informer les catégories de paysages sonores urbains types mises en évidence par l'analyse des titres des scènes synthétisées.

Concernant la validité écologique, une plus grande attention doit être portée sur le corpus de sons, tant au niveau de sa diversité que sur la manière de l'acquérir et de le diffuser. Dans la même optique, il serait souhaitable que la structure sémantique de laquelle découle l'organisation spatiale de notre interface d'exploration, fasse l'objet d'une étude perceptive à part entière.

Enfin, *SceneSynth* étant prévu pour être utilisé via une approche de type *crowdsourcing*, il reste à tester l'effet potentiel d'un tel système de diffusion sur les résultats. Dans une approche expérimentale cognitive, la formulation de la consigne ainsi que sa juste compréhension par le sujet exercent une influence déterminante sur les données obtenues. L'expérimentateur n'étant pas présent durant une expérience par *crowdsourcing*, un travail en amont sera nécessaire afin de s'assurer de la clarté des instructions.

BIBLIOGRAPHIE

Lawrence W. Barsalou. Perceptual symbol systems. *Behavioral and brain sciences*, 22(04) : 577–660, 1999. URL http://journals.cambridge.org/abstract_S0140525X99002149.

Lawrence W. Barsalou. Grounded cognition : past, present, and future. *Topics in Cognitive Science*, 2(4) :716–724, 2010. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1756-8765.2010.01115.x/full.

J. Beaumont, Stephen Lesaux, J.-D. Polack, Cristina Pronello, C. Arras, and Laurent Droin. Pertinence des descripteurs d'ambiance sonore urbaine. *Acoustique et techniques*, 2004.

D. Botteldooren, B. De Coensel, and T. De Muer. The temporal structure of urban soundscapes. *Journal of Sound and Vibration*, 292(1-2) :105–123, April 2006. ISSN 0022460X. doi : 10.1016/j.jsv.2005.07.026. URL http://linkinghub.elsevier.com/retrieve/pii/S0022460X05004888.

Dick Botteldooren and Bert De Coensel. The role of saliency, attention and source identification in soundscape research. *Proc. Inter· noise,(Ottawa, Canada)*, 2009. URL http://users.ugent.be/~bdcoense/content/data/pdf/conference/24_BotteldoorenIN09.pdf.

A.L. Brown, Jian Kang, and Truls Gjestland. Towards standardization in soundscape preference assessment. *Applied Acoustics*, 72(6) :387–392, May 2011. ISSN 0003682X. doi : 10.1016/j.apacoust.2011.01.001. URL http://linkinghub.elsevier.com/retrieve/pii/S0003682X11000028.

Sabine Buchholz and Javier Latorre. Crowdsourcing preference tests, and how to detect cheating. *Proc. Interspeech2011, Florence*, 2011. URL https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/Speak11To12/IS110306.pdf.

Bert De Coensel and Dick Botteldooren. A model of saliency-based auditory attention to environmental sound. *Proc. ICA,(Sydney, Australia)*, page 1–8, 2010. URL http://archive.ugent.be/input/download?func=downloadFile&fileOId=1180414&recordOId=1180411.

Bert De Coensel, Annelies Bockstael, Luc Dekoninck, Dick Botteldooren, Brigitte Schulte-Fortkamp, Jian Kang, and Mats E. Nilsson. Application of a model for auditory attention to the design of urban soundscapes. 2010. URL https://biblio.ugent.be/publication/1180465/file/1180469.pdf.

Boris Defréville, Catherine Lavandier, and Marc Laniray. Activity of urban sound sources. 2004. URL http://www.icacommission.org/Proceedings/ICA2004Kyoto/pdf/Fr3.F.2.pdf.

Aymeric Devergie. Relations entre perception globale et composition de s´equences sonores. Mémoire de master II, IRCAM ATIAM, 2006.

D. Dubois. *Sémantique et Cognition. catégories, protoypes, typicalité.* CNRS EDITIONS edition, 1993.

Danièle Dubois, Catherine Guastavino, and Manon Raimbault. A cognitive approach to urban soundscapes : Using verbal data to access everyday life auditory categories. *Acta Acustica united with Acustica*, 92(6) :865–874, 2006. URL http://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000006/art00003.

Mounya Elhilali. *Neural basis and computational strategies for auditory processing*. PhD thesis, 2004. URL http://drum.lib.umd.edu/handle/1903/2084.

Mounya Elhilali, Juanjuan Xiang, Shihab A. Shamma, and Jonathan Z. Simon. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology*, 7(6) :e1000129, June 2009. ISSN 1545-7885. doi : 10.1371/journal.pbio.1000129. URL http://dx.plos.org/10.1371/journal.pbio.1000129.

James J. Gibson. *The Senses Considered as Perceptual Systems.* Boston, hougton mifflin company edition, 1966.

James J. Gibson. The ecological approach to the visual perception of pictures. *Leonardo*, 11 :227–235, 1978.

Robert L. Goldstone and Lawrence W. Barsalou. Reuniting perception and conception. *Cognition*, 65(2) :231–262, 1998. URL http://www.sciencedirect.com/science/article/pii/S0010027797000474.

Catherine Guastavino. *Étude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*. PhD thesis, 2003.

Catherine Guastavino. The ideal urban soundscape : Investigatng the sound quality of french cities. *Acta Acustica United with Acustica*, 92 :945–951, 2006.

Catherine Guastavino, Brian FG Katz, Jean-Dominique Polack, Daniel J. Levitin, and Daniele Dubois. Ecological validity of soundscape reproduction. *Acta Acustica united with Acustica*, 91(2) :333–341, 2005. URL http://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000002/art00015.

Olivier Houdé, Daniel Kayser, Olivier Kœnig, Joëlle Proust, and François Rastier. *Vocabulaire des sciences cognitives*. Presses universitaires de france edition, 1998.

Olivier Houix. *Catégorisation auditive des sources sonores.* PhD thesis, 2003.

Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, Patrick Susini, and Isabel Urdapilleta. A lexical analysis of environmental sound categories. *Journal of Experimental Psychology : Applied*, 18(1) :52–80, 2012. ISSN 1939-2192, 1076-898X. doi : 10.1037/a0026240. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0026240.

David C. Howell. *Méthodes Statistiques en Sciences Humaines.* De boeck & larcier s.a. edition, 1998.

Steven Komarov, Katharina Reinecke, and Krzysztof Z. Gajos. Crowdsourcing performance evaluations of user interfaces. 2013. URL http://people.seas.harvard.edu/~reinecke/Publications_files/Komarov_CHI2013.pdf.

A. Leobon. *Analyse psycho-acoustique du paysage sonore urbain*. PhD thesis, 1986.

A. Léobon. Représentation cartographique des ambiances sonores d'une ville. Technical report, 1997.

Valérie Maffiolo. Methodes d'approche de l'environnement sonore urbain : Etude bibliographique. Etude pour la mairie de paris - direction de la protection de l'Environnement réalisée au laboratoire d'Acoustique musicale, 1997.

Valérie Maffiolo. *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. Acoustique appliquée, Université du Mans, 1999.

Valérie Maffiolo, Michèle Castellengo, and Danièle Dubois. Qualité sonore de l'environnement urbain : sémantique et intensité. *Acoustique et techniques*, 16 :14–21, 1998. URL http://www.infobruit.com/revues/78_09499.PDF.

Stephen McAdams and Emmanuel Bigand. *Penser Les Sons : Psychologie cognitive de l'audition*. Psychologie et sciences de la pensée. Presses universitaire de france edition, 1994.

Daniel Menzel, Katsuya Yamauchi, Florian Völk, and Hugo Fastl. Psychoacoustic experiments on feasible sound levels of possible warning signals for quit vehicles, 2011.

Pauline Nadrigny. Paysage sonore et écologie acoustique, 2010. URL http://www.implications-philosophiques.org/implications-de-la-perception/paysage-sonore-et-ecologie-acoustique/.

U. Neisser. *Cognition and Reality*. WH freeman and company edition, 1976.

Maria Niessen, Caroline Cance, and Danièle Dubois. Categories for soundscape : toward a hybrid classification. volume 2010, page 5816–5829, 2010. URL http://www.incas3.eu/wp-content/docs/20100613_Niessen.pdf.

J.-D. Polack, J. Beaumont, C. Arras, M. Zekri, and B. Robin. Perceptive relevance of soundscape desciptors/ a morpho-typological approach. 2008.

Manon Raimbault. *Simulation des ambiances sonores urbines : intégration des aspects qualitifs.* Mécanique, thermique et génie civil, Unversité de Nantes - Ecole polytechnique de Nantes, 2002.

Manon Raimbault. Qualitative judgements of urban soundscapes : Questionning questionnaires and semantic scales. *Acta acustica united with acustica*, 92(6) :929–937, 2006. URL http://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000006/art00011.

Manon Raimbault and Danièle Dubois. Urban soundscapes : Experiences and knowledge. *Cities*, 22(5) :339–350, October 2005. ISSN 02642751. doi : 10.1016/j.cities.2005.05.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0264275105000557.

E. Rosch. Principles of categorization. In *Cognition and Categorization*. New Jersey, lawrence erlbaum, hillsdale edition, 1978.

E. Rosch and B. Lloyd. *Cognition and Categorization*. New Jersey, lawrence erlbaum, hillsdale edition, 1978.

R. M. Schafer. *The new soudscape*. Universale Edition, Vienna, 1969.

R. M. Schafer. *The Soundscape, Our sonic environment and the tuning of the world*. Rochester, Vermont, destiny books edition, 1977.

Brigitte Schulte-Fortkamp. Soundscape-focusing on resources. volume 19, page 040117, 2013. URL http://link.aip.org/link/?PMARCW/19/040117/1.

Corsin Vogel. *Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain*. PhD thesis, Université de Paris VI, 1999.

Quatrième partie

ANNEXES

A

LE TEST DE LA SOMME DES RANGS DE WILCOXON

Les informations présentées ici sont issues du livre (Howell, 1998).

Le test de la somme des rangs de Wilcoxon est un des tests les plus couramment utilisés pour comparer les caractéristiques de deux distributions. Ce test représente une alternative au test non paramétrique de Student, reposant uniquement sur l'ordre des observations relatives aux deux échantillons. Son hypothèse nulle stipule que ces derniers sont "prélevés aléatoirement de populations identiques. Il est ainsi particulièrement sensible aux différences de tendance centrale existant entre les populations" (dans notre cas les médianes).

Considérons deux populations, G_1 et G_2 ayant respectivement n_1 et n_2 valeurs. Admettons que l'hypothèse nulle est fausse entre G_1 et G_2 et que les valeurs de G_1 sont généralement inférieures à G_2 . Si nous devions classer les N valeurs ($N = n_1 + n_2$) par rang, sans tenir compte de leur appartenance à G_1 et G_2, nous devrions observer les valeurs de G_1 parmi les premiers rangs et les valeurs de G_2 parmi les rangs les plus élevés. Ainsi la somme des rangs de G_1 serait inférieure à celle de G_2.

Le test de Wilcoxon est basé sur cette logique. Il calcule la somme des rangs dans l'un des groupes et l'utilise comme statistique de test. Dans le cas d'un test bilatéral, si cette somme est trop petite ou trop grande par rapport à l'autre somme, l'hypothèse nulle est rejetée. Comme statistique générale, nous considérons la somme des rangs attribuée au plus petit groupe. Si nous admettons que $n_1 < n_2$ nous considérons la somme W_1.

$$W_1 = \sum_{i=1}^{n_1} r_i \tag{1}$$

Étant donné cette valeur, nous pouvons utiliser les tables statistiques de Wilcoxon afin de tester l'hypothèse nulle H_0. Si nous nommons ψ_1 et ψ_2 les distributions de probabilité des échantillons G_1 et G_2, l'hypothèse nulle s'écrit :

$$H_0 : \psi_1 = \psi_2 \tag{2}$$

Ces tables nous fournissent une valeur critique ω_s, dépendant d'un seuil α usuellement fixé à 0.05. On accepte H_0 si W_1 appartient à un intervalle I_α.

$$I_\alpha = [\omega_s; 2\bar{\omega} - \omega_s] \tag{3}$$

avec

$$\bar{\omega} = n_1(n_1 + n_2 + 1) \tag{4}$$

B

EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS : CONSIGNE DE L'EXPÉRIENCE

Consignes

1- Description de l'expérience

Cette expérience s'inscrit dans le cadre d'une étude sur la perception des environnements sonores urbains, et des processus cognitifs inhérents. Afin d'étudier les représentations mentales des sons de la ville, nous proposons au sujet de reconstruire deux paysages sonores urbains, à partir d'un corpus de sons environnementaux, conçu et structuré dans une phase préalable de l'étude. La synthèse est réalisée grâce à l'outil sceneSynth, développé dans le cadre du projet HOULE.

2- Déroulement de l'expérience

- A l'aide du logiciel sceneSynth, vous devrez synthétiser deux paysages sonores urbains. La durée de chacune des scènes est fixée à 1 minute. Afin de vous familiariser avec l'outil de synthèse, un temps de prise en main de 10 minutes sera prévu en début d'expérience.

- Durant chaque processus de synthèse, il vous sera demandé de nommer les différents éléments sonores que vous aurez choisis parmi les sons du corpus proposés. Cette demande interviendra à chaque fois que vous sélectionnerez un son.

- Vous devrez également nommer votre paysage sonore. Un espace sera prévu à cet effet.

- A la fin de chaque processus de synthèse, il vous sera demandé, sur un fichier texte séparé, de :

a) Commenter votre scène sonore. (Description du résultat final, choix des échantillons, valeurs des paramètres, réalisme, autres ...)

b) Indiquer les éléments sonores que vous auriez souhaité ajouter à votre paysage, mais que vous n'avez pas trouvé dans le corpus

- A la fin de l'expérience, nous vous proposerons, sur un fichier texte séparé, de critiquer la fonctionnalité de synthèse de sceneSynth, ainsi que son interface de sélection.

Vous trouverez ci-dessous un résumé des étapes de l'expériences ainsi que leurs durées respectives.

Tache	Durée (mins)
Présentation de l'expérience Lecture de la consigne	10
Prise en main de l'outil de synthèse sceneSynth	10
Première Synthèse	20
Commentaire de la première synthèse	10
Deuxième synthèse	20
Commentaire de la deuxième synthèse	10
Critique de l'interface	10
total	1h 30

C

EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS : TITRE DES SCÈNES SYNTHÉTISÉES

Numéro du sujet	Titre des scènes idéales	Titre des scène non-idéales
Sujet 1	calm street	harsh street
Sujet 2	Quartier Passages du Caire, Passage du grand cerf	Avenue en travaux à St Denis
Sujet 3	Vivre dans un immeuble sur cour ou une petite rue pietonne	vivre au dessus d'une grande avenue en travaux
Sujet 4	17h dans un parc	carrefour en travaux
Sujet 5	Cour Intérieure	Grand Boulevard
Sujet 6	Rue isolée	accident au carrefour
Sujet 7	Place Piétonne	Grand Boulevard
Sujet 8	Square des Batignolles	rue bruyante bouchon
Sujet 9	parc au soleil	Rue blindée sous la pluie
Sujet 10	parc calme : sous un arbre à distance d'activité	Petite rue avec embouteillage

D

EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS : DESCRIPTIONS DES SUJETS

Sujet 1

Commentaire du paysage sonore idéal :

This idealized soundscape includes birds chirping throughout and a relatively low background noise level, indicating a respect for the natural environment in the overall organization of the city i.e. there are green spaces integrated into the urban fabric. Bike bells are heard above the overall traffic noise, indicating the ability to navigate by bike. Children's voices indicate the friendliness of the city to all ages. Finally, a continuous church bell is heard to indicate a sense of community.

Commentaire du paysage sonore non-idéal :

This unideal soundscape uses a higher level of background noise as well as harsher sound events, representing greater volumes of traffic as well as use of heavy machinery in ongoing construction work. This constitutes an unfriendly city, in which dogs are barking in response to the noise, no people can be heard talking or walking in the streets except for piercing cries, and ambulances are coming to the rescue.

Sujet 2

Commentaire du paysage sonore idéal :

J'ai pensé au quartier piéton entre la rue St Denis, le Passage du Caire, le Passage du Grand Cerf. Il y a de nombreuses écoles dans ce quartier, j'en ai donc choisie une en bruit de fond, c'est un quartier commerçant, où se trouvent de nombreuses galeries, que j'essaie de représenter par « Passage Clos », « Rue Passante » et « Rue Passante 2 ». C'est, de plus un quartier quasiment exclusivement piétons, d'où l'absence de bruits de circulation, ainsi que les bruits de pas et de vélos (ainsi que le freinage/démarrage). Il s'y trouve en plus quelques squares, représentés par les chants d'oiseaux en arrières fond, ainsi que trois églises, dont on entend les cloches.

Commentaire du paysage sonore non-idéal :

Pour cette deuxième scène "non idéale", je me suis souvenu de la grande avenue traversant St Denis (au niveau de Pleyel) alors en travaux. Les sons les plus divers s'accumulaient au point de provoquer des migraines. Les travaux faisaient le plus de bruit avec des sons réguliers de marteau-piqueur que j'aurai peut-être souhaité mettre plus

fort. Les sons de camion en texture exprime bien l'omniprésence des moteurs. L'école à côté ainsi que la sonnerie ne fait que rajouter au capharnaüm. J'ai réglé le volume en fonction de la proximité des événements, en me situant sur le trottoir longeant les travaux.

Sujet 3

Commentaire du paysage sonore idéal :

Le premier paysage sonore représente un environnement sonore idéal selon moi car il est constitué :

- D'éléments plaisants (rires d'enfants, bruits d'animaux, d'eau et sons de cloches)
- D'un fond sonore urbain présent mais non gênant (sans travaux, ni voitures)
- De légers dialogues de rue
- Et sans que l'ensemble ne soit trop présent ni violent pour l'attention et l'ouïe.

En outres, cela m'évoque l'environnement sonore urbain de petites rues piétonnes avec des commerces mais pas trop, ou encore d'immeubles avec fenêtres sur cour.

Commentaire du paysage sonore non-idéal :

Le deuxième paysage sonore représente un environnement sonore exécrable selon moi car il est constitué :

- D'éléments déplaisants (sons de cloches trop présents, pluie, tonnerre)
- D'un fond sonore urbain gênant (travaux, voitures, klaxons)
- De dialogues de rue pénibles (gens qui parlent forts, rires idiots.)
- L'ensemble créant une atmosphère pénible où l'on a des difficultés de concentration et d'où l'on souhaite s'isoler au plus vite.

En outres, cela m'évoque l'environnement sonore urbain d'une avenue avec beaucoup de trafic (voitures, motos) et de surcroît en travaux.

Sujet 4

Commentaire du paysage sonore idéal :

Il s'agit d'un parc aux alentours de 16h-17h, lorsque les enfants sont sortis de l'école. Le parc est proche de la rue et des voitures dont on peut entendre légèrement le son. Aucun son n'est très fort, et il y a peu "d'événement", car il s'agit plutôt d'une scène tranquille et reposante : il n'y a pas de sursaut à cause d'un klaxon ou d'un cri trop fort.

Commentaire du paysage sonore non-idéal :

Il s'agit d'un carrefour en travaux avec beaucoup de passage : la circulation est embouteillée et l'ambiance est oppressante. Il y a beaucoup trop de monde, il fait chaud, et on respire mal à cause de la poussière des travaux. C'est stressant et fatiguant.

Sujet 5

Commentaire du paysage sonore idéal :

Scène dans une cour intérieure, isolée de la circulation, aux alentours de midi, avec un marché à côté. Peut être un jour de week-end, Samedi par exemple. La cour est fait de graviers et de sol solide, quelqu'un arrive à vélo, pendant qu'une autre personne sort de la cour. Une troisième personne se trouve chez elle en train de bricoler quelque chose, utilisant du bois. Espace calme mais vivant, apaisant, dans un environnement assez "beau" : midi, donc soleil au zénith, petite cour intérieur, de la verdure aux alentours (avec les petits oiseaux !)

Commentaire du paysage sonore non-idéal :

Scène dans la rue, sur un grand Boulevard, avec tous les transports bruyants dans le coin (metro, bus, train, voitures) ainsi que les sirènes de pompiers et de police qui viennent rajouter au vacarme environnant. On entend aussi des travaux au loin. Tout cela se passe sous une pluie battante, il ne fait clairement pas beau. Beaucoup de ces sons sont assez agressifs, et surgissent de nulle part ! Aucun de ces sons ne fait naturel

Sujet 6

Commentaire du paysage sonore idéal :

C'est un paysage urbain avec les bruits habituels des grandes villes. Cependant les sons qui sont désagréables et que l'on souhaite moins entendre sont mis au second plan (travaux/ bruits de voix nombreuses...) On ne peut pas se passer de ces sons mais ils sont en retraits par rapport au reste. Ce n'est pas pour autant le son d'un paysage rural.

Commentaire du paysage sonore non-idéal :

Paysage urbain avec uniquement les sons peut supportables et que l'on entend fréquemment dans certains quartiers. Des sons typiques voir clichés de la ville sont présents. Ils arrivent de façon progressive pour finir le violent bruit de la sirène des pompiers.

Sujet 7

Commentaire du paysage sonore idéal :

J'ai souhaité créer une ambiance d'une place piétonne. Le trafic est audible au loin mais reste assez discret, agréable à entendre, comme un murmure de la ville avec ses klaxons et sirènes de pompier distants. La nature est l'aspect "bucolique" y est omniprésent avec les bruit d'oiseaux, de bicyclette et de clochers au loin. Les cris d'enfant donnent vie à l'ensemble avec le passage des vélo et le bruit des graviers foulés par les passants. Une place paisible et agréable pour profiter de l'air pure et du beau temps.

Commentaire du paysage sonore non-idéal :

Il s'agit d'un grand boulevard parisien. Une ambiance et des textures beaucoup plus bruyantes. Les cris d'enfant se sont transformés en une colonie. Les véhicules et bruits de moteurs sont omniprésents. Le bruit de circulation, des alarmes et des klaxons dénotent du reste et sont à la limite du supportable. Le niveau sonore globale de la scène est bien plus fort. Un boulevard où on ne fait que passer sans écouter ce qui se passe autour, excepter pour éviter les accidents...

Sujet 8

Commentaire du paysage sonore idéal :

Paysage sonore d'un parc, avec gazouillis d'oiseaux, légère brise, des enfants qui jouent et les cloches d'une église non loin. Un vélo passe devant l'auditeur.

Commentaire du paysage sonore non-idéal :

Des éboueurs bloquent une rue pour vider les bennes a ordures, ce qui crée un bouchon. les klaxons sont fréquents et le camion benne redémarre. la voiture arrive dans une place populeuse avec un marché, paysage sonore entrecoupé de sirènes et d'antivol. Le tout se passe sous la pluie battante.

Sujet 9

Commentaire du paysage sonore idéal :

C'est une scène dans un parc, au soleil, les gens sortent, mais le parc n'est pas très connu et donc pas rempli de gens. Il fait beau et les gens sont heureux et calmes. C'est un beau début de printemps.

Commentaire du paysage sonore non-idéal :

C'est horrible, il pleut à verse, les rues sont bondées, les gens énervés, il y a une école à côté où les enfants hurlent, il y a pleins de passage, et ça bouchonne parce qu'il y a eu un accident à côté. On est sur un trottoir à attendre son bus qui n'arrive pas, on se

fait bousculer et il n'y a plus de place sous l'abri bus.

Sujet 10

Commentaire du paysage sonore idéal :

C'est un paysage composé de deux plans. Au premier plan on entend le bruissement des feuilles d'un arbre, nous sommes en train de faire la sieste sous un peuplier. On entend aussi quelques oiseaux. Au loin on entend un son très diffus où on distingue des enfants qui jouent. Entendre de l'activité humaine a quelque chose de rassurant, mais s'ils étaient trop près ils me casseraient les oreilles. Il y a aussi vers le milieu de la scène un homme qui en interpelle un autre. C'est pareil c'est de l'activité humaine, ça nous raccroche un peu à la réalité.

Commentaire du paysage sonore non-idéal :

C'est une scène qui pourrait se passer sur la terrasse d'un café, dans une rue d'habitude relativement calme, mais où il y a aujourd'hui une atmosphère de tension. Les gens sur la terrasse sont bruyants et il y a une certaine agressivité dans l'air. Peut être qu'il fait trop chaud, ou qu'il y a de l'orage. Il y a un véhicule à l'arrêt qui bloque la rue, du coup des gens klaxonnent, s'impatientent, et en plus il y a des scooters qui essayent de passer et manquent de bousculer des passants.

E

EXPÉRIENCE PILOTE DE SYNTHÈSE SÉQUENTIELLE PAR CORPUS : ANALYSE LEXICALE DES NOM DONNÉES PAR LES SUJETS

Analyse lexicale des événements sonores pour les scènes idéales

Index du son	Nom donné par le sujet	Correspondance
1	goingtoschool	non traité
2	bikebell	sonnette vélo
3	bells	cloche
4	Passage Vélo	vélo
5	Démarrage Vélo	vélo
6	Son Talon	pas
7	Freinage Vélo	vélo
8	Son Pas	pas
9	Cloches lointaines	cloche
10	Son Oiseau-Square	oiseaux
11	Rire	voix
12	voix d'enfants	voix
13	oiseaux	oiseaux
14	chiens	chien
15	Rentrer chez soi	non traité
16	bruit de metro	transports en commun
17	église	cloche
18	vie dans le quartier	non traité
19	Dialogue dans le quartier	voix
20	Enfants – récréation	voix
21	chant d'oiseau	oiseaux
22	Pas Gravier	pas
23	Pas Solid	pas
24	Pas Gravier	pas
25	Clocher	cloche
26	Velo Arrive	vélo
27	Bricolage Scie	travaux
28	Portail	porte
29	vehicule_1	circulation
30	velo_1	vélo
31	pas	pas
32	pas_2A	pas
33	pas_2B	pas
34	travaux	travaux
35	Birds	oiseaux
36	Children	voix
37	Laugh	voix
38	Sonnette Vélo	sonnette vélo
39	Clochet	cloche
40	Klaxon	alarme/klaxon
41	Pompier	alarme/klaxon
42	Bruit de Pas	pas
43	cloches église	cloche
44	départ vélo	vélo
45	arrivée vélo	vélo
46	passage	non traité
47	rire	voix
48	rire aussi	voix
49	oiseau	oiseaux
50	oiseaux	oiseaux
51	homme interpelle quelqu'un	voix

Analyse lexicale des textures sonores pour les scènes idéales

Index du son	Nom donné par le sujet	Correspondance
1	urban_background	bruit de fond
2	birds	oiseaux
3	Ecole Lointaine	voix
4	Passage Clos	cour intérieure
5	Rue Passante	Ambiance Rue
6	Rue Passante 2	Ambiance Rue
7	Fontaine	fontaine
8	urbain mais discret	non traité
9	vie dans le quartier 2	non traité
10	fontaine	fontaine
11	Vent dans les arbres	Vent
12	Crissement de porte et voitures dans le fond	porte
13	oiseaux	oiseaux
14	Cour Interieur	cour intérieure
15	Marché	voix
16	BG_1	bruit de fond
17	BG_2	bruit de fond
18	BG_3_BIRDS	oiseaux
19	BG_4_AMBWITHVOICE	voix
20	Ambience Rue	Ambiance Rue
21	gazouillis d'oiseaux	oiseaux
22	cour d'école	voix
23	fontaine wallace	fontaine
24	nature	nature
25	gens	presence humaine
26	enfants lointaints	voix
27	vent dans les feuilles, pluie légère	Vent

Analyse lexicale des événements sonores pour les scènes non-idéales (1)

Index du son	Nom donné par le sujet	Correspondance
1	saw	travaux
2	childyell	voix
3	squeak	non traité
4	dogwhine	chien
5	dogbark	chien
6	motor	circulation
7	birdsflyaway	oiseaux
8	horn	alarme/klaxon
9	shrill	alarme/klaxon
10	sirens	alarme/klaxon
11	siren2	alarme/klaxon
12	Avion/Hélico	transports aérien
13	Sonnerie Ecole	alarme/klaxon
14	Marteau Piqueur	travaux
15	Bulldozer/Machine de chantier	travaux
16	Masse Aspiration	non traité
17	Klaxon 1	alarme/klaxon
18	Klaxon 2	alarme/klaxon
19	Son chantier	travaux
20	rire idiot	voix
21	tronçonneuse	travaux
22	perceuse	travaux
23	outil de chantier	travaux
24	cloches d'église incessantes	cloche
25	Chien qui aboye trop fort	chien
26	Helicoptère	transports aérien
27	démarrage au feu rouge	circulation
28	mobilette	circulation
29	klaxon trop long	alarme/klaxon
30	klaxon bus	alarme/klaxon
31	périférique	circulation
32	tonerre	orage
33	scie	travaux
34	poncer	travaux
35	Camion devant un feu rouge	circulation
36	Déchargement camion	travaux
37	Perceuse	travaux
38	Aboiement chien	chien
39	cri	voix
40	rire fort	voix
41	velo qui fraine	velo
42	sonnerie velo	sonnette velo
43	klaxon long	alarme/klaxon
44	Ambulance	alarme/klaxon
45	Bus qui part	transports en commun
46	Metro	transports en commun
47	Scooter	circulation
48	Train Klaxon	alarme/klaxon
49	Sirene Pompier	alarme/klaxon
50	Sirene Loin	alarme/klaxon
51	Sirene Police	alarme/klaxon

Analyse lexicale des événements sonores pour les scènes non-idéales (2)

Index du son	Nom donné par le sujet	Correspondance
52	Klaxon Voiture 01	alarme/klaxon
53	Klaxon 02	alarme/klaxon
54	Scie Proche	travaux
55	Scie Loin	travaux
56	BUS	transports en commun
57	Moto	circulation
58	ALARM	alarme/klaxon
59	SIRENE	alarme/klaxon
60	klaxon_1	alarme/klaxon
61	klaxon_2	alarme/klaxon
62	VEHICULE_DEBUT	circulation
63	VEHICULE DEBUT 2	circulation
64	Voiture	circulation
65	Bus	transports en commun
66	Pompier	alarme/klaxon
67	Klaxon	alarme/klaxon
68	Alarme	alarme/klaxon
69	Police	alarme/klaxon
70	Helicoptere	transports aérien
71	Moto	circulation
72	Chien	chien
73	Voiture	circulation
74	sirène samu	alarme/klaxon
75	klaxon train	alarme/klaxon
76	klaxon voiture court	alarme/klaxon
77	klaxon voiture long	alarme/klaxon
78	klaxon voiture moyen	alarme/klaxon
79	moteur deblayeuse	travaux
80	antivol voiture	alarme/klaxon
81	moteur scooter	circulation
82	tonnerre	orage
83	éclair	orage
84	scie	travaux
85	scie2	travaux
86	travaux	travaux
87	voitures	circulation
88	moto	circulation
89	klaxon	alarme/klaxon
90	sirénes	alarme/klaxon
91	gens	présence humaine
92	plus de gens	présence humaine
93	marteau	travaux
94	klaxon	alarme/klaxon
95	sonnette velo	sonnette velo
96	scooter qui demarre	circulation

Analyse lexicale des textures sonores pour les scènes non-idéales (1)

Index du son	Nom donné par le sujet	Correspondance
1	loudbackground	bruit de fond
2	heavymachinery	brouhaha travaux
3	Ecole Proche	voix
4	Camions à l'arrêt	brouhaha circulation
5	Circulation	brouhaha circulation
6	Circulation sur terre (travaux)	brouhaha circulation
7	Boucan de rue	rue animée
8	pluie	pluie
9	Gros travaux	brouhaha travaux
10	Pluie Battante	pluie
11	Moteur Bus	transport en commun
12	Train	transport en commun
13	Interieur Gare	Non traité
14	Pluie Marche	pluie
15	BG_1	bruit de fond
16	BG_2	bruit de fond
17	BG_3_RAIN	pluie
18	Circulation	brouhaha circulation
19	Colonie	voix
20	éboueurs	camion poubelle
21	Circulation	brouhaha circulation
22	Groupe électrogène	véhicule travaux
23	moteur voiture	brouhaha circulation
24	place du marché	voix
25	rue animée	rue animée
26	pluie	pluie
27	pluie	pluie
28	travaux	brouhaha travaux
29	enfants enervés	voix
30	véhicules	brouhaha circulation
31	brouhaha avec gens énervés	voix
32	camion point mort	brouhaha circulation