



HAL
open science

**Everyday sound. A journey through perception,
cognition, interaction, and imitation of sounds that
surround us**

Guillaume Lemaitre

► **To cite this version:**

Guillaume Lemaitre. Everyday sound. A journey through perception, cognition, interaction, and imitation of sounds that surround us. Cognitive science. Université Pierre et Marie Curie (Paris 6), 2016. tel-01448929

HAL Id: tel-01448929

<https://hal.science/tel-01448929>

Submitted on 29 Jan 2017

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

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

UPMC

 SORBONNE UNIVERSITÉS

Université Pierre et Marie Curie Paris

Everyday sound

A journey through perception, cognition, interaction, and imitation of sounds that surround us

Le son du quotidien

Un voyage à travers la perception, la cognition, l'interaction, et les imitations des sons qui nous entourent

Habilitation à diriger les recherches

Présentée par Guillaume Lemaitre

Soutenance prévue le 23 novembre 2016 devant le jury composé de:

Monsieur Pascal Belin - rapporteur

Madame Sabine Meunier - rapporteuse

Monsieur Daniel Pressnitzer - rapporteur

Madame Michèle Castellengo - examinatrice

Monsieur Bruno Giordano - examinateur

Madame Sølvi Ystad - examinatrice

Acknowledgement

This document discusses research carried out in different laboratories in France, the United States, and Italy. I am fully indebted to all the persons that made this adventure possible: Stephen McAdams at Ircam, Nicolas Tsingos at INRIA, Laurie Heller at Carnegie Mellon University, Davide Rocchesso at the Università Iuav di Venezia, Patrick Bousard and Christophe Lambourg at Genesis. Above all, I would thank Patrick Susini, who has always supported me since Day One, pulled me out of many quicksands, and kept my canoe afloat across crocodile-infested swamps (these are not metaphors).

I would also like to thank all the colleagues and friends who have made these adventures so fun: Pablo Arias, Jean-Julien Aucouturier, Mondher Ayari, James Ballas, Stephen Barrass, Alexis Baskind, Frédéric Bevilacqua, Anne Caclin, Maxime Carron, Stefano Delle Monache, Murielle Ducas, Jules Françoise, Karmen Franinović, Aurélie Frère, Massimo Grassi, Adam Greenberg, Nicolas Grimault, Andrea Halpern, Hans Hansen, Olivier Houix, Vincent Isnard, Mathieu Lavandier, Enrico Marchetto, Jeremy Marozeau, Antoine Minard, Marie-Hélène Moulet, Nicolas Misdariis, Rozenn Nicol, Pietro Polotti, Emmanuel Ponsot, Geoffroy Peeters, John Pyles, Benneth Smith, Clara Sued, Mike Tarr, Jennifer Tomlinson, Michael Vannier, Sandrine Vieillard, Yon Visell, Frédéric Voisin, and the many more I have forgotten here.

Of course, this work was also made possible by all the students I worked with: Emmanuel Gallo, Karine Aura, Arnaud Dessen, Mathieu Pellerin, Jayant Bhambhani, Clara Charlotte Baron-Hyppolite, Nicolas Zúñiga-Peñaranda, Rentaro Matsukata, Matthew Lehet, Daniele Galante, Sylvain Mysliwiak, Loïc Eymar, Ali Jabbari, Hugo Trad, Hugo Scurto, Lou Seropian. In particular, I would like to give a big hug to Nicole Navolio, who has never failed to amaze me!

Sadly, I would like to remember here Suzanne Winsberg and Vittorio Girotto.

Finally, I would like to thank the committee for having accepted to review this work. I would like in particular to poke Bruno Giordano, whose meticulous work has been an important source of inspiration and a prod to my research.

À Violette et Prunette

Abstract

This document summarizes and discusses the research I carried out since 2004. This research focused on *everyday sounds*: sounds of manufactured products and common events occurring in a listener's daily environment, excluding music, speech, or other communication systems. It was motivated by an overarching question: what does audition tell listeners about their environment? In other words, this research sought to understand what exactly listeners can identify when a sound-producing event occurs in their environment, and how does the auditory system process auditory stimuli to create meaningful representations of the events. Chapter 2 starts with some initial observations made during experiments on timbre perception: even though listeners were required to concentrate on the *sounds*, they could not help but hear the *source* of the sounds. Chapter 3 then addresses the issue of sound identification. In particular, it shows that listeners are better and faster at identifying the sound-producing *actions* than the *objects* on which the actions are executed. Neuroimaging studies showed that the difficulty to identify object properties may be due to additional semantic and visual processing that is not required to identify actions. Chapter 4 then explores *interactions between audition and motor behaviors* and shows that hearing the sound of an action has the potential to facilitate executing that action. Finally, Chapter 5 studies another form of interaction between audition and action: *vocal imitations* of sounds. It shows that vocal imitations are a spontaneous means of communicating sounds and that vocal imitations effectively convey the identity of what they imitate by selecting the most relevant auditory features. The different chapters also illustrate how these results were applied to evaluate the sound quality of industrial products, design sonic interactions in computer interfaces, and develop intuitive sketching tools for sound designers.

Résumé

Ce document synthétise et discute les travaux de recherche que j'ai menés depuis 2004. Cette recherche a pour thème les sons du quotidien, c'est-à-dire les sons de produits manufacturés et d'événements se produisant communément dans l'environnement d'un auditeur, à l'exclusion de la musique, de la parole et de tout autre système de communication. Ces travaux sont traversés par une question générale: qu'est-ce l'auditeur perçoit de son environnement ? En d'autres termes, ces travaux ont cherché à comprendre ce qu'un auditeur perçoit lorsqu'un événement sonore se produit dans son environnement, et comment le système auditif traite les stimuli sonores pour construire une représentation cohérente de cet événement. Le chapitre 2 détaille des observations faites au cours d'études du timbre: bien que les participants des expériences devaient juger les *sons*, ils avaient beaucoup de difficultés à ne pas juger la *source* de ces sons. Le chapitre 3 s'intéresse alors plus particulièrement à l'identification des sources sonores. Il montre notamment que les auditeurs identifient plus facilement et plus rapidement les *actions* qui produisent les sons plutôt que les *objets* sur lesquels ces actions sont effectuées. Des études en neuroimagerie montrent alors que la difficulté à identifier les propriétés des objets est due à des traitements sémantiques et visuels supplémentaires. Le chapitre 4 s'intéresse alors aux *interactions entre audition et comportements moteurs*. Il montre en particulier que l'écoute du son d'une action a le pouvoir de faciliter l'exécution de cette action. Le chapitre 5 discute finalement d'un autre type d'interaction entre audition et comportement moteur: les *imitations vocales* de sons du quotidien. Il montre que les imitations sont utilisées spontanément pour décrire des sons et que ces imitations communiquent efficacement l'identité des sons, en sélectionnant les propriétés les plus pertinentes des sons. A travers ces différents chapitres, ce document illustre également comment ces recherches ont été appliquées à l'évaluation de la qualité sonore de produits industriels, au design d'interactions sonores dans les interfaces informatiques, et au développement d'outils intuitifs d'esquisse sonore.

Contents

Acknowledgement	i
Abstract/Résumé	iii
1 Introduction: everyday sounds?	1
2 Perception: the timbre of everyday sounds	5
2.1 Studying the timbre of everyday sounds	5
2.2 From timbre to sound quality	6
2.3 Discussion	7
3 Cognition: Identifying everyday sounds	9
3.1 How do listeners classify sounds?	9
3.2 Acoustic correlates of sound sources	12
3.3 Objects <i>or</i> actions? Behavioral studies	13
3.4 Objects <i>and</i> actions? fMRI studies	15
3.4.1 Background: cortical networks of sound identification	16
3.4.2 Oddball experiment	17
3.4.3 Discriminating action and material	18
3.5 Discussion	21
4 Interacting with sounds	23
4.1 Priming actions with sounds	23
4.1.1 Nature of the auditory-motor associations	23
4.2 Sonic Interaction Design	27
4.2.1 Sounds to guide gestures: The Spinotron	28
4.2.2 Pleasantness and usability: the Flops	29
4.3 Discussion	31
5 Vocal imitations of everyday sounds	35
5.1 Imitating sounds to communicate them	35
5.2 Can listeners recognize what is imitated?	37

5.3	How do people imitate sounds?	39
5.4	Imitations as a sound sketching tool	40
5.5	Discussion	41
6	Discussion and perspectives	45
	References	50
	Appendices	69
A	Curriculum vitae	71
B	Research projects	75
C	Teaching and mentoring	79
D	List of publications	83

Chapter 1

Introduction: everyday sounds?

This document presents the research I carried out since 2004.¹ The main theme has been the perception of everyday sounds. The starting point was the acknowledgment that listeners hear much more than *sounds*: Hearing provides listeners with a full-fledged representation of the events happening in the environment. I thus set out on a journey that took me from psychoacoustics through the cognitive processes underlying sound identification. The main thesis that emerged from this research is that auditory perception is inherently bound to and interacts with motor behaviors. The idea of a tight connection between *speech* production and perception is probably not unfamiliar to the readers. The document attempts to show that such an intimate connection also exists for *everyday sounds*. It also presents some practical application of this work to the fields of sound design and sonic interaction design. As an introduction, the next paragraphs define everyday sounds, sound quality, and sonic interaction design.

Everyday sounds, sound quality, sonic interactions

The *everyday sounds* studied here are any other sounds than speech, animal communication, or music (Ballas and Mullins, 1991; Gygi and Shafiro, 2007).² More precisely, Vanderveer (1979) defined them as “any possible audible acoustic event which is caused by motions in the ordinary human environment. . . . Besides (1) having real events as their sources . . . , (2) [they] are usually more ‘complex’ than laboratory sinusoids, . . . , (3) [they] are meaningful, in the sense that they specify events in the environment.

¹Although this document *mainly* focuses on research conducted after I completed my PhD, I took the liberty to summarize this work in Chapter 2 because the main questions covered by this document actually originated from observations made during this PhD.

²There is a little confusion in the literature, where the term “environmental sounds” is either used to refer specifically to the sounds natural phenomena, or more generally to every other sounds than speech and music. We thus prefer the term “everyday” sounds, which also insists on the idea that the sounds under study are part of the listener’s daily environment and very often produced by artefacts.

... , (4) the sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation.” This definition pinpoints a very important idea: everyday sounds are caused by physical events (“motions”) occurring in the listeners’ environment and inform them about these events.³ In fact, this document will show that they are often not even perceived as abstract sounds but as the sounds *of something happening*.

Everyday sounds are for the most part produced by manufactured products (e.g. home appliances, vehicles) and contribute to their overall experience. For example, electric or hybrid cars are almost silent at low speeds. Car manufacturers therefore add sounds to these silent vehicles, first and foremost to warn pedestrians about a potential danger. These sounds become an inherent part of the vehicle identity, and thus contribute to the overall aesthetic experience and brand image of the vehicle. *Sounds quality* refers to activities that consist of assessing how well sounds contribute to the functional, emotional, and aesthetic aspects of a product (the sound quality), and specifying acoustic requirements to reach a given sound quality. *Sonic interaction design* also explores ways in which sounds can be used to convey functional, aesthetic and emotional qualities to a product, but in interactive contexts. It particularly focuses on human-computer interfaces that continuously produce sounds in response to users’ gestures (Serafin et al., 2011).

Open questions in everyday sound research

The titles of two articles published in an early issue of *Ecological Psychology* still summarize the major questions addressed by research on everyday sound: “What and how do we hear in the world?” Gaver (1993a,b).

So **what do we hear?** This question, central to the field of **auditory cognition**, is the topic of Chapter 3. A widespread view is that one main function of audition is to direct the eyes to a location of interest, where the visual system takes turn to make sense of what had just happened (Kubovy and Van Valkenburg, 2001). But in fact, sounds may tell a little more than that. Consider for example the following descriptions of sounds (taken from a subject’s descriptions in Lemaitre et al., 2010): “plip-plip-ploop”, “a series of quiet, short, and high-pitched sound”, “water”, “drips”, “a kitchen sink”, “a faucet is leaking”, “plumbing noise”. They all describe the same sound - and all are correct in some respect - but they also describe different views of the same thing. Chapter 3 will show that listeners hear three types of information from a sound: acoustic properties (“high-pitched sounds”), causal phenomena (“a drip”), and

³Note that this definition is in fact close to the notion of “auditory object”: a percept corresponding to a sound (or a group of sounds) perceived as a coherent whole that is assigned (correctly or incorrectly) to a single source or event in the environment (Alain and Arnott, 2000; Bizley and Cohen, 2013).

semantic associations (any kind of information associated with the interpreted sound source: “a faucet is leaking”). This document mainly focuses on the ability of listeners to recover the causal phenomena at the source of the sound and it will argue that their best ability is to tell the action that caused the sounds.

As Chapter 3 will suggest, there is a close connection between hearing and making an action. Chapter 4 will thus reframe Gaver’s question as: **What do we hear for?** How do sounds help us **interact** with our environment? In fact, sounds are *dynamical* stimuli. Sounds unfold in time, cannot be frozen, and as such are similar to actions. Chapter 4 will examine how hearing sounds can help, guide, or prime an action. It does so at an experimental level, but also tries to evaluate how a better understanding of auditory-motor cognition could help designers re-imagine the role of sounds in human-computer interfaces, moving from passive displays of information (often less effective than visual displays) to the substrate of sonic interactions.

The other important question raised by Gaver was: **How do we hear?** How does the auditory system build a meaningful representation of the subject’s surrounding environment? On the one hand, perceiving the high pitch of a signal probably requires little more than processing acoustic features. On the other hand, telling that the sound was produced by a faucet leaking in a kitchen requires some sort of *inference*, informed for example by a priori knowledge of the situation and other sensory modalities. In fact, the semantic interpretation of the cause of a sound has a lot in common with language perception (Ballas, 1993) and neurocognitive studies suggest that the processing of meaningful sounds and language share common neural networks (see for example Lebrun et al., 2001; Orgs et al., 2006; Cummings et al., 2006; Schön et al., 2009; Aramaki et al., 2010). Between these two extremes, the listener’s recovery of the mechanical cause of a sound is usually assumed to be possible without any semantic interpretation: each sound may possess acoustic features that are unequivocally characteristic of its cause.

A fascinating question is thus to understand how the auditory stimuli are represented and memorized. For example, one possibility is that auditory stimuli are represented as abstract percepts and that most details of the sensory stimuli are lost when transferred to different memory stores (Mathias and von Kriegstein, 2014). Conversely, some other results suggest that subtle details of acoustic stimuli may be kept in memory for quite a long time (Agus et al., 2010). Many ingenious experimental methods have been devised to address this question.⁴ These methods seek to query the listeners’ representations (directly or indirectly, implicitly or explicitly) when they hear, memorize, imagine, remember sounds. Chapter 5 will argue that a very natural way for listeners to share their auditory representations is to **imitate the sounds**,

⁴For the sake of concision, I tried to refrain from providing comprehensive reviews of the literature (though I have included a few summaries of important results when I felt it was necessary). I acknowledge the frustration this may create. The interested reader can find them in the articles mentioned in this document.

vocally and with gestures. This last chapter will therefore discuss several studies of vocal imitations of everyday sounds. These studies have mostly tried to understand the ability of people to imitate non-vocal sounds with the voice and the ability of listeners to make sense of these imitations. These studies have now converged toward the result that listeners can recognize what vocal imitations imitate. Because imitations are simplified versions of what they imitate, they offer perspectives to understand what are the pieces of information that are necessary for sound recognition, and thus may open a window into listeners' cognitive representations of sounds. Chapter 5 will also show the results of a project that aimed at using vocalizations and gestures to create intuitive tools for sound design.

Organization of the document

The document is organized around the different themes mentioned in the previous paragraph: Chapter 2 starts with the study of timbre perception to introduce the questions of cognition (Chapter 3), interaction (Chapter 4), and imitations (Chapter 5). The different studies that form the basis of these discussions were in fact conducted within different research projects that are reported in Appendix B. Table 1.1 provides the correspondence between the themes of research and the different projects.⁵

Year	Project	Perception	Cognition	Interaction	Imitation
2000	PhD Klaxon	x			
2004	Opera	x			
2006	CLOSED, SID, MINET		x	x	x
2009	CMU		x	x	
2012	luav		x		x
2013	Genesis	x			
2016	Skat-VG	x	x	x	x

Table 1.1: Correspondence between the themes discussed in this document and research projects the studies took place in.

⁵This document focuses only on the studies that fall into its framework. I have therefore omitted some of the studies carried out within these projects.

Chapter 2

Perception: the timbre of everyday sounds

This chapter first illustrates the psychoacoustical approach to *timbre*. It also details how the results can be used to evaluate the sound quality of industrial products. Importantly, this chapter introduces a key remark: it is very difficult for untrained listeners to judge the timbre of sounds without considering their source. This will pave the way for the next chapters.

2.1 Studying the timbre of everyday sounds

My first foray into the perception of everyday sound was rooted in the psychoacoustical approach to timbre perception. Psychoacoustics aims at establishing quantitative relationships between perceptual attributes (or *auditory features*) and acoustic properties of the sound signals. Some of the attributes are singled out in most Western languages: duration, loudness, and pitch. But a clarinet and a trumpet playing the same note, for the same duration, and at the same loudness still sound different: their *timbres* are different (American Standard Association, 1960; Krumhansl, 1989; Risset and Wessel, 1999). Timbre is in fact multidimensional, bundling up several attributes; the goal of the psychoacoustical approach is precisely to characterize its dimensions.

One method uses dissimilarity ratings and multidimensional scaling (MDS) with a three-step procedure. First, listeners rate the dissimilarities between each pair of sounds of a set. Then, MDS represents the dissimilarity data by distances in a geometrical space (*perceptual space*), whose dimensions represent perceptual attributes. These dimensions are then interpreted by acoustic parameters (called in this case *psychoacoustic descriptors*).

Most studies have investigated the timbre of musical instruments. Following in the footsteps of Susini et al. (1999, 2011), I studied the timbre of car horn sounds during my

PhD (Lemaitre et al., 2007). First a free sorting experiment identified nine families: participants were simply required to create categories of sounds based in whichever criteria they felt relevant. Interestingly, these nine families corresponded precisely to the different mechanisms used in the car horns: type of resonator, number of devices, type of driver, etc. In short, it seemed that the most natural way for the listeners to organize these sounds was to cluster them according to how they were mechanically produced. Then, we submitted a selection of sounds to the MDS approach (McAdams et al., 1995; Winsberg and Soete, 1993), which yielded a spatial model with three dimensions. The first dimension corresponded to the perception of *roughness* (Daniel and Weber, 1997). The second dimension corresponded to the perception of *brightness* (Zwicker and Fastl, 1990). The third dimension was akin to what separates the timbre of a clarinet from a trumpet (Krimphoff et al., 1994).

These results seem remarkably consistent with the other studies on the timbre of musical instruments (see for example McAdams et al., 2010): a low number of dimensions, and dimensions that are often related to similar attributes: brightness, roughness, etc. This raises the question of the generality of the dimensions: are they valid for any sound set? In an attempt to test this idea, Nicolas Misdariis conducted a meta-analysis of several timbre studies (Misdariis et al., 2010). It used sounds from four different products: car engines (Susini et al., 1999), air-conditioning units (Susini et al., 2004), car horns, and car doors (Parizet et al., 2008). Overall, this study confirmed that several dimensions (and brightness in particular) consistently describe the timbre of these different sounds.

2.2 From timbre to sound quality

The psychoacoustical approach to timbre applies to any kind of sounds (musical or industrial). But how do we get from timbre perception to the evaluation of sound quality and to the design of new sounds?

Next paragraphs offer two examples that used timbre to sound quality evaluation and specification. First, I used the results of the study of the car horns to specify the acoustic properties of new sounds (Lemaitre et al., 2009b). Designing the sound of car horns involves a compromise between the need to customize the sounds and the necessity of providing efficient warning signals (the legal function of car horns is to warn road users against potential danger). But what is an efficient warning signal? A signal that is audible and unambiguously recognized as warning (Rogers et al., 2000). Here, we thus defined effective car horn sounds as sounds that are recognized as car warning signals, and sought to identify the acoustic properties that allow sounds to be recognized as such. Existing and new sounds were submitted to a forced-choice experiment: participants indicated whether they thought the sound was produced by a car horn or by something else. Analysis of the consensus between the participants

resulted in a mathematical model that defined the acoustic requirements for a sound to be recognized as a car horn sound, hence mapping the timbre space.

Another study conducted for *Genesis Acoustics* evaluated the unpleasantness of a particular source of car interior noise: wind buffeting noise (aerodynamic turbulences) caused by the vehicle moving at high speed through air. We evaluated the “unpleasantness” of these sounds by having listeners rate the unpleasantness of many recordings in a wind tunnel on a set of scales, with a method that allowed them to compare many different sounds with some positive and negative references (International Telecom Union, 2001-2003). Multilinear regression analysis resulted in a model that predicts the unpleasantness of the sounds based on loudness and fluctuation strength (Zwicker and Fastl, 1990). This model is now part of a tool proposed by the wind tunnel facility to their customers.

2.3 Discussion

The psychoacoustical approach to timbre has proven to be a very valuable tool to study the perception of sounds as diverse as musical instruments and car noises. This approach has identified several perceptual dimensions (and their acoustic correlates) that are remarkably consistent across studies: Sharpness, roughness, fluctuation strength, attack time, for example, are almost systematically found across a wide range of studies. Many algorithms have been devised that calculate descriptors correlated to these dimensions, which are now available in several pieces of commercial software (Head Acoustics’s *Artemis*, *Genesis’s Lea*, Brüel et Kjær’s *Pulse*, etc.).

Describing the timbre of a set of sounds is however only the first part of any sound quality study. It describes the key auditory features that characterize the perception of a particular set of sounds. The issue is then to measure the quality of the product sounds. “Quality” can refer to many different things. The most straightforward method consists of assessing the pleasantness of the sounds, as with the example of wind buffeting noise (pleasantness characterizes the valence of the feelings elicited by listening to a sound, see Section 4.2.2). Specific applications may require more sophisticated assessment, depending on what exactly the sounds are used for in a given product. The function of car horns, for example, is to warn road users of a potential danger; accordingly designers must create new sounds that are still recognized as car horns. In other cases reported in Section 4.2.1, the function of the sounds is to facilitate or guide the manipulation of an interface. In this case, sound quality should focus on manipulation. But in any case, putting quality measurements in relation with timbre descriptors allows the researchers to connect the acoustics of the signals with the aesthetic or functional requirements of the products. The psychoacoustical approach is again extremely valuable for that purpose.

A few qualifications should however be made. The definition of timbre used by the

method is very useful in that it allows a precise methodology based on dissimilarity ratings. But it also hints at another idea: timbre is also what defines an instrument's identity (Hajda et al., 1997). The study of the timbre of car horn sounds, for example clearly showed that listeners could not help but make sense of what they heard, and that it was utterly difficult for them not to ignore the different mechanisms creating the sounds when they judged the dissimilarity between the sounds. In fact, as recommended by Susini et al. (1999, 2011), it is very important that the sounds be homogeneous and not too different. Otherwise listeners become unable to abstract themselves from the source and judge auditory features. Of course, different sound sources are also characterized by different auditory features, but the relationship is often far from a combination of simple features. As a consequence, applying the timbre approach to sounds produced by very different sources provides judgments that are clustered around different sound sources or mechanisms, which violates the assumptions of continuous dimensions used by the psychoacoustical definition of timbre. Next chapter will discuss the question of "what do we hear" (to quote Gaver, 1993b): the *identification of sound sources*.

Another important question is whether the perceptual dimensions highlighted by these studies are genuine psychological percepts or simply reflect the ability of the auditory system to use the most informative cues in a set. Most studies of timbre carefully select the sounds they use. First, the method can only use a few sounds (the number of trials increases as the square of the number of sounds). Second, as noted before, the sound set must be homogeneous. Experimenters thus either create sounds by varying regularly a few synthesis parameters, or select sounds that are closely distributed. So it is very possible that the sound selection procedure creates sets of sounds that distinguish themselves along a few continuous dimensions, and that listeners' judgements simply reflect the selection procedure. In that account, timbre studies can only be interpreted as showing that listeners are sensitive to the statistical variability of these features. It does not prove for example, that these features exist per se (as percepts), are hard coded in the auditory system, or that the auditory system performs some kind of dimensionality reduction, transforming complex auditory inputs into a limited set of auditory features (see for example the discussions in Aucouturier and Bigand, 2012 and Siedenburg et al., 2016). In contrast, the results of Giordano et al. (2012) showed that different auditory features (pitch, loudness, spectral centroid, harmonicity) are actually encoded in different cortical areas, suggesting that auditory system specifically encodes these features (see Chapter 6).

Chapter 3

Cognition: Identifying everyday sounds

As discussed in the previous chapter, it is often very difficult for listeners to focus on *sounds* while ignoring their *source*, to a point that they may even be unable to conceptually distinguish the two. This chapter will confirm that listener's spontaneous organization of a sound set is clustered around the different sources that created the sounds, that listeners spontaneously identify the sources of the sounds but are better able to identify the actions rather than the objects that created the sounds, and that identifying objects recruits more widespread cortical networks than identifying actions.

The fact that listeners spontaneously associate a sound and its source had in fact been acknowledged early on (see for example Vanderveer, 1979; Handel, 1989; Faure, 2000). This idea is probably best summarized by Gaver (1993a,b), who proposed the distinction between *musical listening* (when the listener focuses on the qualities of the acoustic signal)¹ and *everyday listening* (when the listener focuses on the event creating the sound). Accordingly, sounds can be similar in different ways: two sounds can be considered as similar because the acoustic signals are similar (acoustic similarity), or because their sources are also similar (causal similarity). Two similar sources can produce similar signals, but can as well produce very different signals (think of all the different sounds that cars make). Different listeners might pay more attention to one or the other of these aspects (i.e. use different modes of listening), depending on their listening habits, abilities, or skills.

3.1 How do listeners classify sounds?

I investigated this question in a study that compared how naive and expert listeners freely sort sounds into categories (Lemaitre et al., 2010). Categorization and identifi-

¹This idea is close to Schaeffer's "écoute réduite" (Schaeffer, 1966).

cation are very closely related: Identifying the source of a sound consists of connecting auditory perception to concepts (knowledge about a particular category), and concepts to language (McAdams, 1993; Goldstone and Kersten, 2003; Barsalou et al., 2003). Only a few experimental classifications (i.e. the result of a sorting task) of environmental sounds have however been reported (Vanderveer, 1979; Guyot et al., 1996; Marcell et al., 2000; Gérard, 2004; Guastavino, 2007; Merer et al., 2009). The analysis of these data led me to define three types of similarity:

- The similarity of acoustic properties: acoustic similarity;
- The similarity of the identified physical event causing the sound: causal similarity;
- The similarity of some kind of knowledge, or meaning, associated by the listeners to the identified object or event causing the sound: semantic similarity.

In the central experiment of our study, two groups of listeners (expert and non-experts; experts were professional sound artists, audio engineers, and acousticians) freely sorted a selection of sounds of usual objects found in a kitchen, using the experimental procedure described in paragraph 2.1. In addition, they indicated which type of similarity they had created the categories with. The *causal uncertainty* (*Hcu*, Ballas, 1993) of the sounds had been measured in a preliminary experiment where listeners described each sound. *Hcu* is minimal when listeners all provide the same cause, and maximal when each listener provides a different cause. Analyses showed that both expertise and *Hcu* had a significant effect on the strategies used to sort the sounds. Non-expert listeners created categories mostly based on causal similarity whereas expert listeners created categories mostly based on acoustic similarity. Sounds whose cause was difficult to identify (i.e. high causal uncertainty) were more likely to be categorized on the basis of the acoustic similarity, whereas sounds with non ambiguous cause (i.e. low causal uncertainty) were more likely to be categorized according to their source. The result that expert listeners are more likely to group together sounds on the basis of acoustic similarity is directly related to their expertise: judging the content of a sound. What is more striking in our results is that even expert listeners had difficulty overriding the cause of a sound when the source of the sound is obvious.

We continued to explore classification of everyday sounds in a study conducted by Houix et al. (2012). A first experiment studied how listeners sort a large set of environmental sounds. The results highlighted four main categories of sounds: solids, liquids, gases, and machines. The second experiment focused only on sounds made by solid objects. The results suggested a taxonomical organization first distinguishing discrete (e.g. impacts) from continuous interactions (e.g. tearing), and suggested that action is the most important organizational principle (see Figure 3.1). We inferred that listeners organize sounds hierarchically, mainly on the basis of the action or movement generating the sounds. Furthermore, we proposed that the different actions are reflected in the temporal patterns of the sounds.

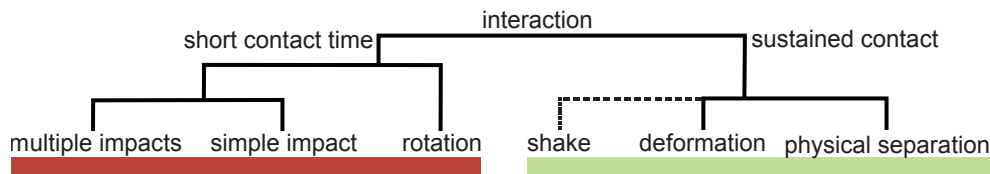


Figure 3.1: The taxonomy of sounds made solid objects proposed by Houix et al. (2012). This taxonomical organization first distinguishes discrete (e.g. impacts) from continuous interactions (e.g. tearing), and suggests that action is the most important organizational principle

The hierarchical (or taxonomical²) organization of sound categories shows that sounds can be described at different levels of specificity. But is there a level that is the most accessible to listeners? One way to approach this question is to find the entry level at which an object is identified, akin to the “basic level” of categories conceptualized by Rosch (1978). In Rosch’s proposal, three levels are distinguished in taxonomy, from the more general to the more specific: superordinate, basic, and subordinate levels. Two properties of the basic level are particularly relevant. First, identification at this level is faster than at other levels, and identification is more accurate at the basic level than at the superordinate level. Second, category labels and sensory representations are the most tightly coupled at the basic level, as evidenced by priming experiments. Laurie Heller and I searched for evidence that the auditory organization of categories of sounds produced by actions includes such a privileged level of description (Lemaitre and Heller, 2013), using these two properties of the basic level: identification time and priming effect. The study used sounds that were created by single objects (or substances) undergoing simple actions that could be described at five levels of specificity in a taxonomy inspired by our previous work (state of matter, type of interaction, specific action, manner of the action, object of the action). The first experiment (identification experiment) compared the accuracies and reaction times of listeners identifying the sounds at five different levels in the taxonomy. The second experiment (semantic priming) compared the priming of a lexical decision task by the prior presentation of a sound, using a paradigm inspired by van Petten and Riefelder (1995). In each trial, a sound was played and a string of letters was then displayed. The listeners answered whether the string of letters formed a word in English. The words and the sounds could be unrelated (incongruent trials) or the word could correspond to one level of description of the sound (congruent trials). Priming is characterized by participants responding faster and more accurately in the congruent trials than the incongruent trials. The results of both experiment showed the labels

²A taxonomy is a system by which categories are related to another by means of class inclusion.

that describe the specific action causing the sound (e.g. tapping, scraping, dripping, etc.) form the basic level in the taxonomy: identification times were minimal and priming effect was maximal for the label describing the specific action. At a theoretical level, semantic priming of a lexical decision is generally accounted for by a spread of activation generated by the prime (Neely, 1991). Our results therefore suggest that listening to simple sound events activate lexical and/or semantic representations that may interact with the lexical or semantic representations activated by the names of the categories of types of interactions and specific actions. Because both the names of these categories describe the actions (and in the majority of the cases the gesture) that produced the sounds, this suggests that hearing a sound activates the semantic representations of the actions (and possibly gestures) producing the sounds. This result is consistent with several studies showing that the sounds of actions uniquely activate motor and pre-motor brain areas (see Section 3.4). It is also consistent with theories of embodied cognition where there are close links between conceptual memory and sensory and motor systems (Barsalou, 2008).

3.2 Acoustic correlates of sound sources

Many studies have reported the listeners' ability to recover auditory properties of events causing sounds (e.g. material, size, shape of objects causing the sounds, see for example Lakatos et al., 1997; Kunkler-Peck and Turvey, 2000; Grassi, 2005). Usually, these studies also seek to identify the acoustic information used by the listeners to recover these properties (Aramaki et al., 2009). For example, McAdams et al. (2004) identified perceptual dimensions correlated with physical parameters of synthesized sounds (physical models) of struck bars with different materials. Massimo Grassi, Massimiliano Pastore and I pursued a similar attempt and investigated the ability of listeners to estimate the size of a ball dropped on a plate (Grassi et al., 2013). Results showed that listeners' estimations were accurate. In addition, they suggested that listeners were likely integrating frequency and amplitude cues in order to produce their estimate.

There are however many other examples of listeners not being able to accurately recover the properties of the sound sources. For example, Lutfi and colleagues showed that listeners do not optimally use the available acoustic information to decide upon the material or the hollowness of the struck bars (Lutfi, 1995; Lutfi and Oh, 1997; Lutfi, 2001; Lutfi et al., 2005). Furthermore, many studies have used synthetic, simple sounds. Using recorded sounds, it has also been sometimes difficult to identify a clear correlation between acoustic properties and the perceived event properties (Freed, 1990). For example, stereotypical (and by consequence not necessarily accurate) relationships between acoustic properties and listeners' responses have been highlighted: slow, loud and low frequency sounds systematically associated with male hand-clappers

(Repp, 1987) or walkers (Li et al., 1991). But the most striking example is probably that of Giordano and McAdams (2006), who used recorded sounds of plates of various dimensions and materials. They reported a perfect discrimination between the “gross categories” of material (metal-glass vs Plexiglas-wood), independent of shape, but an impaired discrimination within the gross categories. A remarkable result was that there was a systematic influence of shape: small objects were systematically identified as glass or wood, irrespective of their actual material.

3.3 Objects *or* actions? Behavioral studies

Sounds are usually produced by an action executed on an object or by several objects in interaction. Thus, in principle sounds could provide information to listeners about both the actions (e.g. hitting) and the objects (e.g. the material) creating the sound. However, there is a strange dissociation in the literature, with two different fields studying each of these aspects. On the one hand, most of the psychoacoustical studies mentioned in the previous paragraphs have studied how people identify object properties from sound (and in particular material). On the other hand, recent neuroimaging studies have focused on human-generated actions, following the discovery of audio-visual mirror neurons in monkeys (see Keysers et al., 2003; Aglioti and Pazzaglia, 2010, for an overview). The following paragraphs describe two different studies conducted at Carnegie Mellon University that directly compared object and action perception.

Overall, the psychoacoustical studies of material perception have shown that performance is not optimal: listeners can only distinguish gross categories, and are prone to confusions due to (inappropriate) stereotypical associations (Avanzini and Rocchesso, 2001; Tucker and Brown, 2003; Giordano, 2003; Giordano and McAdams, 2006). This result may not be so surprising if we consider that the classification experiments previously described show that action, rather than object, is an organizing principle of listeners’ representations of sounds. In fact, it may well be the case that the auditory system is better suited to perceive action than material properties. On the one hand, action properties (type of contact, force, speed, direction) are readily reflected in the temporal evolution of acoustic parameters (loudness, localization, frequency), and thus easily picked up by the auditory system. On the other hand, information about material, size, shape, is very easily graspable from the light reflected by the object, and thus are clearly visual properties.

Laurie Heller and I first investigated this question in a behavioral study (Lemaitre and Heller, 2012). Two experiments used acoustic stimuli created by crossing different actions (bouncing, hitting, rolling, scraping) executing on cylinders made of different materials (glass, plastic, metal, wood). In a first experiment, listeners rated how well each sound event conveyed its materials or actions. The results showed first that listeners did not differentiate materials from within the same gross category (e.g. metal from

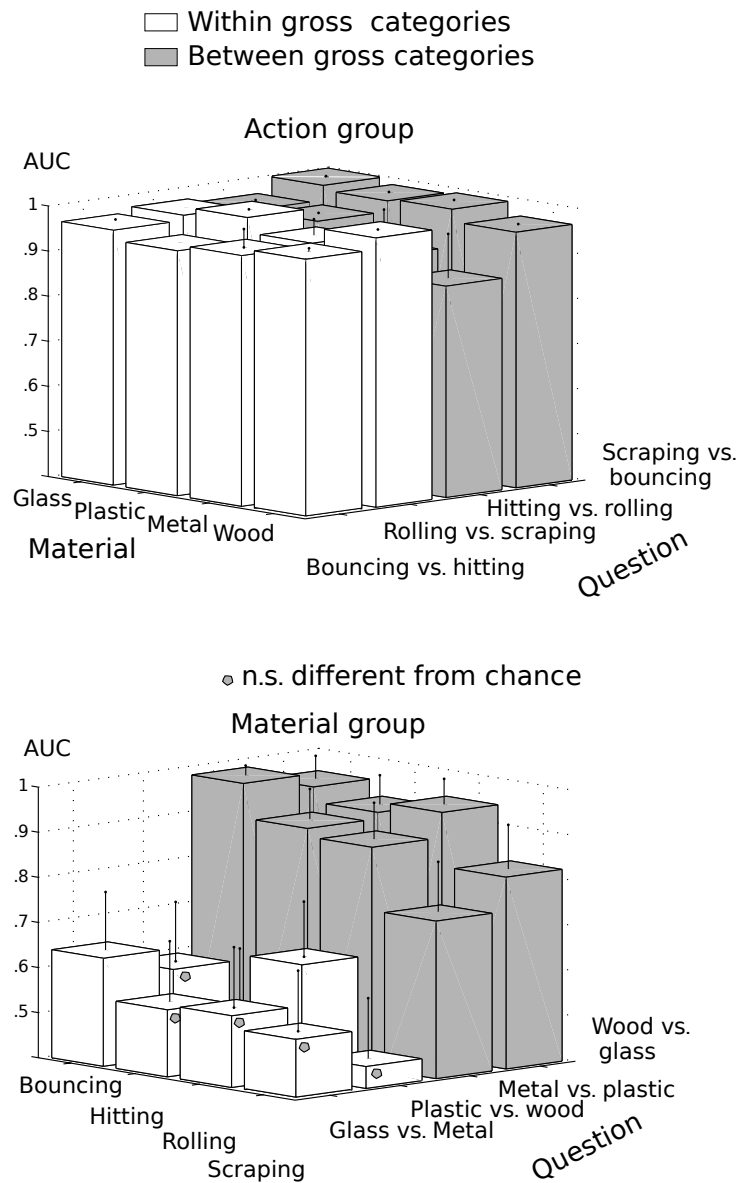


Figure 3.2: Results of the first experiment in Lemaitre and Heller (2012). Listeners discriminated actions across different materials, and material across different actions. AUC is an unbiased measure of accuracy (0.5 corresponds to chance level, 1 corresponds to perfect discrimination, Macmillan and Creelman, 2005). AUC for the different conditions in the action group (upper panel) and material group (lower panel). The vertical bars represent the 95 % confidence interval. The dark bars represent the comparisons between gross categories of materials or actions, the white bars the comparisons within the same gross category. The gray symbols indicate an AUC value not statistically difference from chance (0.5). Listeners could accurately discriminate material only within gross categories (upper panel), whereas they could always discriminate action.

glass). Second, performance for material discrimination varied as a function of action. In particular, when the cylinders were scraped, performance was worse than when the cylinders were hit, bounced, or rolled. This result suggests that material perception was only accurate when the sound event included some form of impact, however small (in theory, material information is conveyed the decay rate across frequencies, thus hearing material requires some sort of impacts, see Wildes and Richards, 1988). Conversely, action information was evident across all of the materials, and discrimination performance was very good for all actions. These results are summarized in Figure 3.2.

In a second experiment, we measured identification performance and identification times for different categories of actions and materials that were easy to discriminate (they were selected from the results of the first experiment): we only used comparisons between the gross categories. The sounds were gated and we tested different gate durations. The most notable result of this experiment concerned the reaction times: Even when we isolated the sounds for which our listeners performed an almost perfect discrimination of material, the listeners were faster to identify the action than to identify the material (see Figure 3.3). Several hypotheses can be advanced. First, it is possible that the acoustic cues used to judge action or material occur at different times. But it is also possible that the psychological processing time was faster for action sounds, because identifying action requires only processing sensory information, whereas identifying material requires deeper cognitive processing.

3.4 Objects *and* actions? fMRI studies

As stated before, many auditory neuroimaging studies have focused on sounds of human-generated actions, but no study has contended with the fact that both types of information are usually simultaneously available in sounds: when someone is knocking on a door, listeners hear both the knocking and the door. To address this issue I looked into the brain processes underlying sound identification. The studies reported in this paragraph were made possible by a Rothberg award grant from the Center for Neural Basis of Cognition (CNBC) at Carnegie Mellon University (“Neural Bases for Sound Identification”). The studies were done with Laurie Heller and in collaboration with John Pyles (CNBC) and Andrea Halpern (Bucknell University, PA). Clara Charlotte Baron-Hyppolite, Jayant Bhambhani, Nicolas Zúñiga-Peñaranda helped us with data collection. Nicole Navolio also helped me with the statistical analysis. Importantly, the grant was also used to collaborate with Matthew Lehet at Carnegie Mellon. The first study reported below was the second year assignment of his PhD.

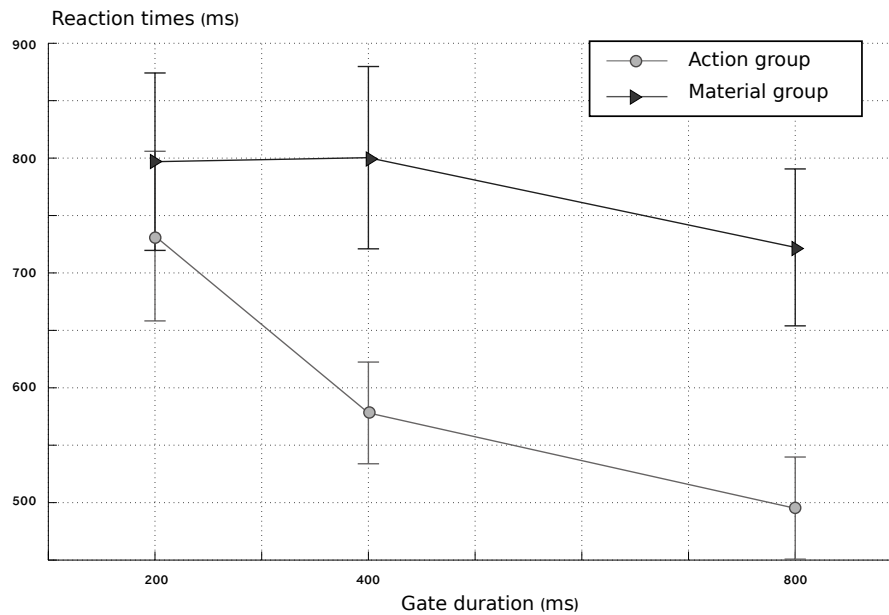


Figure 3.3: Reaction times in the second experiment of Lemaitre and Heller (2012). Listeners were identifying the action or the material that created the sounds. The selected sounds were extremely easy to discriminate (accuracy reached ceiling). The horizontal axis represent the duration of the gates used to shorten the sounds. Note that listeners discriminated action faster than material.

3.4.1 Background: cortical networks of sound identification

Auditory processing is hierarchically organized, with processing becoming more abstract as the distance to primary auditory cortex increases (see Giordano et al., 2010, for a review). One model further characterizes the auditory system as having both a ventral and a dorsal stream, mirroring the dual stream model for vision (Goodale and Milner, 1992). In this model, the ventral stream projects anteriorly along the superior temporal gyrus (STG) toward the inferior frontal gyrus (IFG), and the dorsal stream projects posteriorly through STG and Middle Temporal Gyrus (MTG) toward inferior parietal lobule (IPL) and intraparietal sulcus (IPS) and then to motor and premotor regions in the frontal lobe. The ventral stream (the “what pathway”, “hearing for perception”) is interpreted as processing the sounds’ identities, and the dorsal stream (the “how” and “where” pathway, or “hearing for action”) processes spatial properties necessary to guide actions (see Arnott and Alain, 2011, for a review).

As regards action sounds, a very important series of studies were conducted by James Lewis and colleagues, and served as a basis of the NBSI project (Lewis et al., 2004, 2005; Lewis, 2006, 2010; Lewis et al., 2011, 2012; Engl et al., 2009). They

showed a striking distinction between the cortical networks involved in processing different types of sounds. Sounds made by human actions activate a large network of cortical areas (the “action-sound network”): a left-lateralized fronto-parietal network corresponding to the dorsal stream, and a bilateral complex located in the posterior superior temporal sulcus and middle temporal gyrus (the pSTS/MTG complex, sensitive to biological motion, see for example Beauchamp et al., 2003). The fronto-parietal network includes the inferior frontal gyrus (IFG), inferior frontal sulcus (IFS), and particularly the junction of the IFS and the precentral sulcus (corresponding to the ventral premotor cortex VPMC). It also includes a region centered on the intra-parietal sulcus (IPS) and spreading to the inferior and superior parietal lobules (IPL and SPL). IFG is involved in binding sounds’ spectrotemporal features and conceptual representations (Kaplan and Iacoboni, 2007) and motor programs (in VPMC in particular, potential site of the human mirror neurons, see Keysers et al.; Aglioti and Pazzaglia, 2003; 2010). The functional roles of IPS and IPL are usually interpreted as the integration of auditory information about actions with other sensory modalities (mainly vision), and a form of “action replay” (Arnott et al., 2008). Furthermore, this fronto-parietal network overlaps with regions activated when participants manipulate tools. Therefore, Lewis and colleagues interpreted the fronto-parietal network as “praxis” or “audio-motor association” system, presumably linking one’s own motor repertoire with sound production within the dorsal stream.

In contrast, fewer neuroimaging studies have focused on auditory recovery of object properties. For example, James et al. (2011) shifted the participants’ attention to either the shape or the material of the objects producing the sounds, and interpreted the left posterior MTG as an amodal shape operator. There was no evidence of a material specific area, although a more liberal criterion resulted in activity in the lingual gyri, a visual area found for material processing (Cant and Goodale, 2007). Activation of early visual areas by auditory stimuli has been reported in other studies (Vetter et al., 2014). I used this idea of shifting attention to different properties in the experiment reported next.

3.4.2 Oddball experiment

To study the neural bases of sound identification, our first study used six “human-generated sound events”, each with a unique combination of action and material information (Lehet et al., forthcoming). These six distinct sound events were created by combinations of three actions (hitting, shaking, squeezing) and two materials (liquid, solid). Participants performed an oddball identification task in the scanner: participants had to detect stimuli that were different from the other ones (oddballs). The oddballs included unique “materials” (gasses) so that attending to materials would help listeners pick out oddballs, but it also included some non-unique materials so that attention to actions would also help identify oddball sound events. The oddball task

was chosen to encourage consistent attention, processing, and active listening for the identities of the actions and materials composing each sound event without directing participants to concentrate on one feature.

The data for the oddball task were analyzed with three complementary strategies: (1) a 3 Action \times 2 Material analysis of variance (ANOVA), performed on whole brain, (2) a multivariate pattern analysis (MVPA) using a whole brain searchlight, and (3) an MVPA focusing only on certain regions of interest (bilateral occipital cortices -visual areas- and left motor and premotor cortices). The MVPA searchlight procedure was used to identify spatial patterns of activity that could accurately classify the actions, the materials, or the sound events. The functionally defined ROIs provide a means of identifying properties of sound events that can be decoded from regions used for action and visual object processing (focusing on smaller regions increases statistical power).

The study found three main results. First, activations associated with processing action involved a larger network (large bilateral areas of the dorsal auditory processing pathway extending from primary auditory cortex into superior temporal cortex and inferior frontal regions) than processing material (the main effect of material was confined to a small portion of left primary auditory cortex). Second, areas responsive to action information were spatially congruent with those responsive to individual sound event information. This suggests a functional overlap in neuroanatomical areas that process action and those that process event information in audition. Finally, we showed that action and sound event information could be decoded from activity patterns in motor and visual areas whereas material information could not. Taken together, all three results strongly support the idea that action information is preferentially used over concurrent material information during the processing of these human-generated environmental sounds, and that material processing requires deeper cognitive processing, accessing for example visual information.

3.4.3 Discriminating action and material

One caveat of this study was that the sounds were all different. As noted by Giordano et al. (2012), some of the differences observed could have been at least partially driven by acoustic differences. Therefore, the goal of the second study was to isolate the influence of selective auditory attention toward object or action on the cortical activity within the networks subserving sound identification (Lemaitre et al., in revision). We used sets of sounds for which listeners could successfully identify both the action and the object, and we manipulated the focus of attention toward either action or object properties by using different discrimination tasks. These sounds consisted of recordings of easily discriminable sounds created by tapping and scraping metal and plastic pipes (similar to those used in Lemaitre and Heller, 2012, see section 3.3). Each stimulus consisted of a pair of sounds, and we focused participants' attention by requiring them to indicate whether the actions (A) or the material (M) of the two sounds were the

same or different. We also added a control task in which participants compared scrambled versions of the action sounds (N, meaningless noises sharing the same long-term acoustic properties of other stimuli). As in Lehet's study, we used two different types of analyses: a general linear model analysis (GLM), and a multivoxel pattern analysis (MVPA). The GLM analysis highlighted clusters of voxels whose average cortical activity (percent change of blood oxygenation level) was significantly different between the different tasks. In contrast, the MVPA identified distributed patterns of activity across voxels that were different across the different tasks.

The results are summarized in Figure 3.4. The first and somewhat expected result was that the meaningful tasks overall (A and M) elicited more activation than the Noise task in the left fronto-parietal action-sound network (centered on the left-intraparietal sulcus I-IPS and encompassing a large part of the left Inferior Frontal lobe I-IF, see Lewis et al. 2006). Importantly, the Action and Material tasks activated most of this network equivalently. We therefore suggest that these regions are for the most part not under the control of focused auditory attention and may identify both basic properties of sound objects even when listeners focus on only one property.

Second, contrast analyses revealed significant differences in activation between the Action and Material tasks in four cortical areas: the medial aspect of the Superior Frontal Gyrus (MedSFG), the temporo-parietal junction (I-TPj: posterior part of the left superior and middle temporal gyri, inferior parietal lobule), the left inferior frontal sulcus (I-IFS) and the bilateral inferior occipital lobes (b-IO). Detailed analysis of the time course of the BOLD signals showed that the significant differences found in MedSFG and I-TPj reflected in fact a *greater deactivation of the brain default mode*.

In contrast, the Material task elicited a *greater activation in I-IFS*, positively correlated with the task demand. These results suggest a functional organization of the left inferior frontal region involved in sound source identification, with the inferior frontal sulcus mapping sensory inputs to semantic representations and being sensitive to task demand (activity in VMPC was correlated with reaction times), and VPMC (specifically mapping sensory inputs to motor representations) under the control of the focus of attention (both analyses concluded that focusing attention modulated activity in this area).

Finally, the *Material task elicited significantly more (positive) activation than the Action task in bilateral occipital region*, but activation was not correlated with task demand. This suggests that focusing auditory attention on the material of the objects causing the sounds actually recruited visual areas previously found to process material identification. In short, we showed that discriminating materials elicited increased activity in cortical regions connecting auditory inputs to semantic, motor, and even visual representations, whereas discriminating actions did not increase activity in any regions.

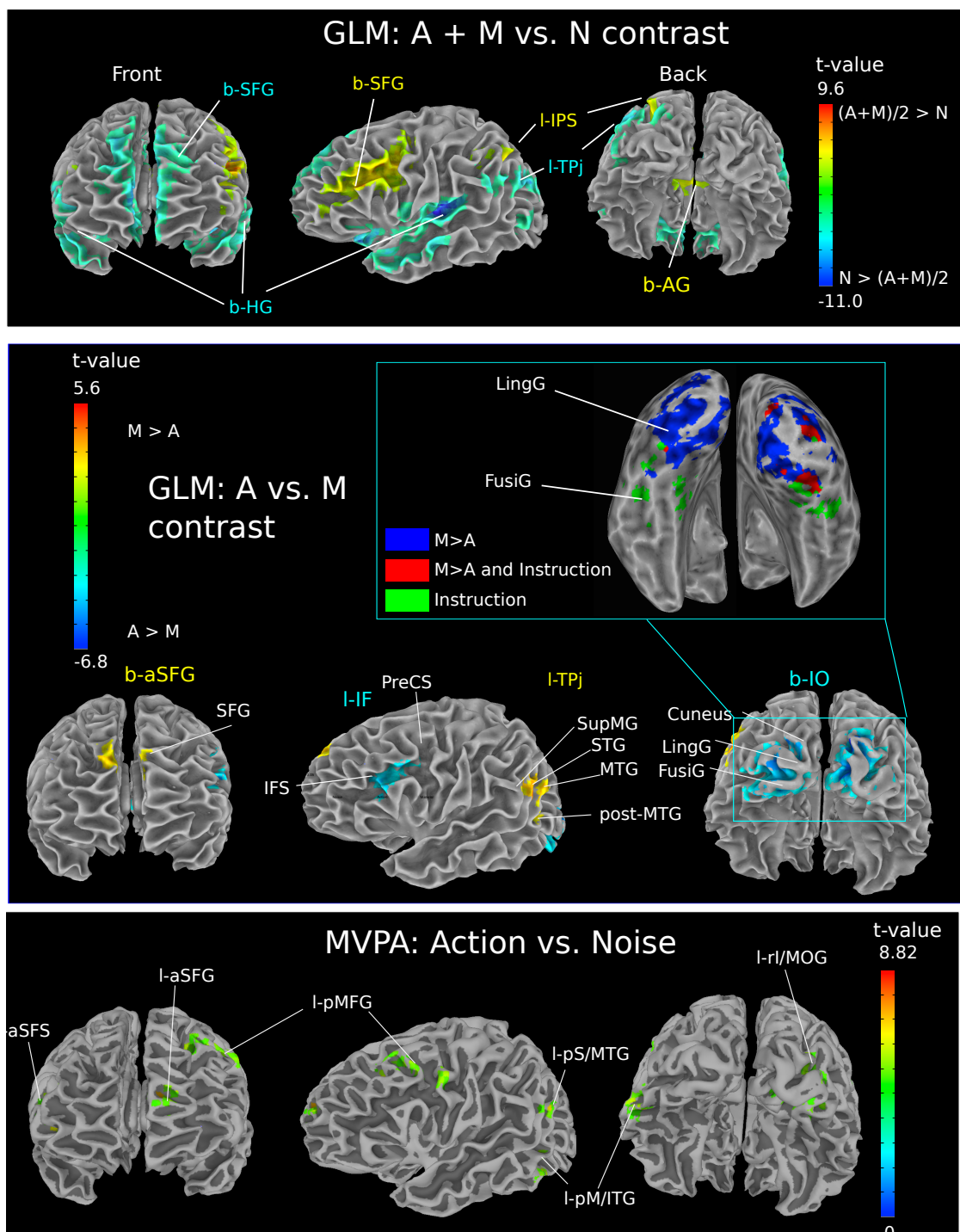


Figure 3.4: Participants in Lemaitre et al. (in revision) were discriminating pairs of sounds, either on the basis of the material (M) or the action (A) causing the sounds. In a control task (N), they discriminated scrambled versions of the sounds. Upper panel: Results of the A+M-2N contrast in the GLM analysis. Regions in yellow correspond to $A+M > 2N$, regions in blue correspond to $2N > A+M$. Middle panel: results of the A - M contrast in the GLM analysis. Regions in yellow correspond to $A > M$. Regions in blue correspond to $M > A$. Lower panel: results of the MVPA (A versus M).

3.5 Discussion

Chapter 2 ended up with the suggestion that listeners spontaneously associate sounds with their source. They do not hear “a series of impulsive sounds with an initial low-frequency impulse followed by a rapidly rising pitch”: they simply hear “water dripping”. The sorting experiments reported in this chapter confirmed this intuition: when listeners sort sounds freely (i.e. with no specific instructions), the most common strategy consists of making groups of sounds based on what we called “causal similarity”. Listeners group together sounds generated by the same source, and not necessarily sounds that sound similar but are produced by different sources. The distinction is important, as using different similarities may result in completely different results: the sound of car horn and a closing door, yet completely different acoustic signals, may well be considered similar, because they are produced by the same product (many participants in Lemaitre et al., 2010 made groups of “food”, “appliances”, “furniture”, etc.). Ballas and Howard (1987) described this phenomenon as “homonym-type sounds”: sounds being discriminated acoustically, but whose sources are confused.

The classification strategies are modulated by two factors. First, the identifiability of the sounds is of course a crucial aspect. When listeners cannot picture what has made the sound, their only available option is to use the acoustics of the sounds. But as soon as they can imagine what has made the sound (and imagination can run wild: even pure tones can be associated with spaceships), it is very difficult for them not to judge the source of a sound instead of the sound itself. Second, expertise plays an important role. Sound engineers for example, whose training consists precisely of analyzing sounds by ear irrespective of their content, are less prone to group sounds based on their sources. But even these persons are strongly influenced by the cause of the sounds, especially for non-musical sounds.

This result has practical consequences, as psychoacoustic experiments (such as those reported in Chapter 2) require participants to concentrate on the auditory features of the sounds and ignore the sources of the sounds. This is not difficult when these experiments use “abstract” sounds (sounds whose source is not identifiable by the listeners: tones, noise bursts, etc.). But this might not be possible in general, and especially for “concrete” sounds (sounds whose cause is unambiguous). This again confirms that psychoacoustical experiments with everyday sounds require care: sound sources must not be too different and participants have to be trained.

So listeners do hear the source of the sounds and not the sounds. But what exactly do they identify? The analysis of the literature and our results suggest that only certain pieces of information may be easily auditorily graspable. Mechanical sounds are generated by actions executed on objects (e.g. hitting a pipe, turbulences caused by vehicle moving through air), and in theory both types of information could be available to listeners. Even if most of the literature has sought to identify what acoustic cues listeners use to identify the properties of the objects (and material above all), our

results suggest that the auditory ability to do so is actually limited: listeners can only tell gross categories apart (e.g. metal from wood, but not metal from glass). Instead, our results show that listeners are much better at identifying and discriminating the actions that make the sounds.

One of the reason might be that the nature of acoustic signals makes them better carriers of information about the actions. Actions are characterized by their dynamics: impacts are characterized by the fact that energy is introduced at one single instant, sustained actions by the fact that energy is continuously introduced to the system, the temporal pattern of events distinguishes broken glass from bouncing glass, etc. Just as music is the “art of time”, everyday sounds evolve in time. Sounds cannot be stopped, cannot be frozen. Stationary sounds (the closest thing to a “frozen” sound) are defined by the fact that their statistical properties do not change *over time*: even stationarity requires time to take place. Thus, the auditory system may have evolved precisely to catch up the information best conveyed in time: actions. Object properties (material, shape, size) are in turn best conveyed by light patterns, and are thus merely visual properties. Note however that our studies did not compare directly vision and audition. We are not claiming that audition is better than vision for picking up actions happening in the environment, but that audition is very good at it.

One of the most remarkable results reported by Lemaitre and Heller (2012) was that identifying material takes always more time than identifying action. The fMRI studies reported in Lehet et al. (forthcoming) and Lemaitre et al. (in revision) offer some insights into the cognitive processes that may correspond to the additional time required to identify action. Identifying action and material both activate the cortical network of sound identification. But identifying material also involves specific brain processes. It deactivates more strongly the brain default mode (suggesting that it requires more concentration) and activates more strongly brain areas traditionally interpreted as providing access to semantic knowledge, and processing visual information. This latter result is particularly striking (though not unique, see for example Vetter et al., 2014). Taking altogether, these results suggest that identifying action may rely only on processing auditory features, whereas identifying material requires other cognitive resources, and in particular, imagining the visual aspects of the objects that have created the sounds. This again suggests that material properties (and maybe other object properties) are essentially visual. Identifying them requires the listeners to connect auditory inputs with visual representations.

Chapter 4

Interacting with sounds

Overall, the previous chapter points toward a very tight interaction between audition and action. This chapter examines more thoroughly how auditory stimuli interact behaviorally with action planning and execution. In a first line of research, I conducted a project at Carnegie Mellon University that showed that hearing an action sound can in fact prime executing that action. The results showed that the auditory-motor associations subserving the priming effect are created on-line, short-lived, and egocentric. The second line of research was developed during the CLOSED and SID projects. It studied how sounds may help users interacting with computer and tangible interfaces, in the context of Sonic Interaction Design (Serafin et al., 2011).

4.1 Priming actions with sounds

This project was carried out at Carnegie Mellon University with Laurie Heller. It involved several undergraduate students (Clara Charlotte Baron-Hyppolite, Jayant Bhambhani, Nicolas Zúñiga-Peñaranda). Notably, Nicole Navolio made a contribution that was extremely valuable, helping with the design and the analysis of the results, and conducted a study on her own (see below). Rentaro Matsukata created the interface.

4.1.1 Nature of the auditory-motor associations

The goal of a first study reported by Lemaitre et al. (2015) was to explore the nature of the associations between auditory perception and action by asking how hearing a sound produced by a manual gesture can prime executing that gesture.

Examples of such behavioral interaction can be found almost exclusively in the vision literature. For example, research on visually guided actions has shown that the affordances of a visual stimulus can prime the participant's actions (see for example Bub and Masson, 2010; Masson et al., 2011; Tucker and Ellis, 2004). To our knowledge

however, this paradigm has not been applied to auditory perception. In fact, there are few ecological situations in which a sound may afford or guide an action (Whyatt and Craig, 2013; Chen, 2011). But for the most part, environmental sounds typically occur *as a consequence* of gesture (i.e. *after* the gesture) and are thus not used to adjust action parameters during gesture planning. As a consequence, the concept of visually guided action does not transfer easily to the auditory domain.

Nonetheless, it is possible that the contingency between a gesture and its acoustic consequence can create a strong association between gestures and sounds, as illustrated by studies of the stimulus-response compatibility effect (see Simon and Rudell, 1967, for a classical example). For example, Hommel (1996) used tones to prime a choice-reaction task. Participants responded to a cue by pressing one of two possible keys. Each key produced either a low-pitched or a high-pitched tone. Prime tones were played before the cue, and participants responded faster when the primes were congruent with the response than when it was incongruent. Our study used this latter paradigm.

Most of the experimental studies of the SR effect use *arbitrary* gesture-sound associations (participants have never encountered these associations before the experiment) that are rapidly created *during* the experiment (i.e. associative learning). As such, the observed effects rely on very recent gesture-sound associations. But other studies in the visual domain have used the *ecological* consequences of an action (Brass et al., 2000). Similarly, gesture priming by sound may also draw upon prior experience and long-term representations of ecological associations (i.e. the natural consequences of an action). The only study to our knowledge that used *ecological sounds* was reported by Castiello et al. (2010).

Accordingly, the goal of our study was to compare priming gestures with sounds that were either arbitrarily or ecologically associated with these gestures, in a S-R paradigm. By the previous line of reasoning, ecological associations should yield more robust priming effects than newly created associations. This assumes that the priming effect draws upon stable long term associations that take time to establish and to update. In addition, ecological associations should persist more durably than newly created associations once the association is no longer reinforced. An alternative possibility is that the priming effect is mediated by short-term, plastic gesture-sound associations, in which case ecological and arbitrary associations should yield the same priming effects, but priming should disappear quickly once the association is no longer reinforced.

The study used five experiments that followed the same S-R paradigm summarized in Figure 4.1: participants responded to a vocal cue (S1 or S2) by executing a response gesture (R1 or R2) that produced a response sound (E1 or E2). Response sounds were used as primes during the test phase. The different experiments varied the response sounds (arbitrary pure tones with different frequencies, tapping and scraping sounds produced by the response gestures), and the response gestures (key lifts on keyboard, tapping and scraping gestures on a custom-made interface). Other variations of the

paradigm used or did not use a learning phase, muffled the sounds of the response interface, and stopped the response sounds halfway through the experiment to observe if the priming effect could persist even when the gesture-sound association was no longer reinforced (see Table 4.1 for a summary of the designs as well as the results of each experiment).¹

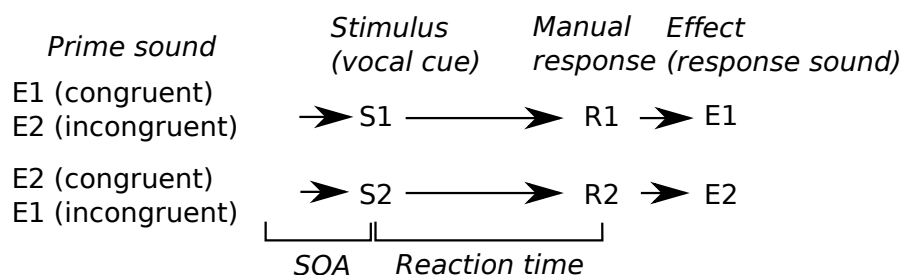


Figure 4.1: The structure of a trial, common to all experiments in Lemaitre et al. (2015) and Navolio et al. (2016). Participants responded to vocal cue S1 or S2 by executing a response gesture R1 or R2. R1 and R2 produced response sounds E1 and E2, respectively, which were also used to prime the gestures. See Table 4.1 for a design summary.

Table 4.1: Outline of the experimental design used in Lemaitre et al. (2015). See Figure 4.1 for the structure of a trial. *: $p < .05$. **: $p < .01$. n.s.: non-significant.

Exp.	S1/S2	R1/R2	E1/E2	Learn.	Effect size
One	Left/right	Left/right key lifts	Low/high tones	Yes	23.0 ms**/4.4 %**
Two	Left/right	Left/right key lifts	Tap/Scrape	Yes	17.9 ms**/2.0 %**
Three	Left/right	Tapping/scraping	Tap/Scrape	No	15.9 ms**/1.7 %**
Four	Left/right	Tapping/scraping	None	No	5.5 ms*/1.0 %**
Five	Left/right	Left/right key lifts	1: L/h tones 2: None	Yes	20.9 ms**/3.2 % 2.4 ms (n.s.)/0.6 %

As a whole, the experimental results confirmed the existence of an auditory motor priming phenomenon in which listening to a sound contingently associated with a gesture primes executing that gesture: participants responded faster and made fewer errors when the response gesture was congruent with the response sound than when it was incongruent. Figure 4.2 illustrates such a result for the experiment that used the tapping and scraping gestures and the tapping and scraping sounds.

¹Data are available at the following URL: <https://zenodo.org/record/20734>, last retrieved on September 19, 2016.

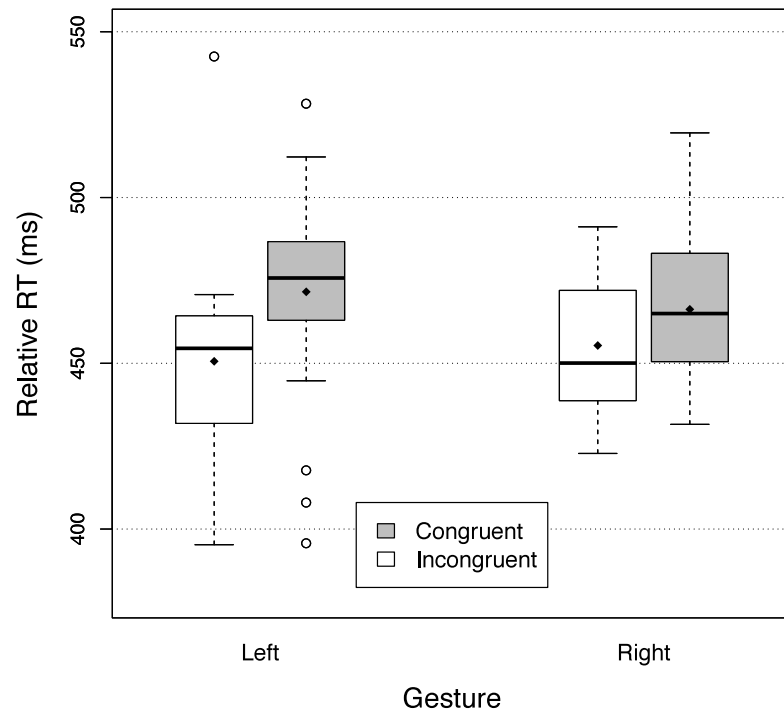


Figure 4.2: Relative reaction times in the third experiment (tapping and scraping a wooden interface) of Lemaitre et al. (2015). The solid box represents the interquartile (IQ) range of the distributions of RTs. The dashed lines represent the interquartile range plus or minus 1.5 IQ. Circles represent data points outside this range (i.e. outliers). The horizontal lines represent the median and the losanges the means of the distributions.

Comparing the different versions of the experiment resulted in two important findings. First, a similar priming effect was observed both for gesture-sound associations newly created by associative learning and for associations that were learned through life-long experience. Second, the priming effect was drastically reduced when the association was no longer being reinforced by the experiment. A first clear conclusion is therefore that long-term prior exposure to the gesture-sound association is not necessary to generate auditory-motor priming. The second notable conclusion is that long-term prior exposure to the gesture-sound association alone is not sufficient to generate auditory-motor priming to the full extent. These results provide evidence that auditory-motor priming is mainly created by rapid learning of the association between sounds and the gestures that produce them. Auditory-motor priming is therefore mediated by short-term associations between gestures and sounds that can be readily reconfigured regardless of prior knowledge.

Using the same priming paradigm, Navolio et al. (2016) investigated the *frame of*

reference of the auditory-motor associations (i.e. whether the association is egocentric or exocentric). This question is especially important when designing human-computer interactions using sounds.² The frame of reference directly influences the learnability, the ease of use, the extent of immersion, and many other factors of the interaction. For example, Salzman et al. (1999) found that an egocentric frame of reference is beneficial for learning local, immersive details, but exocentric perspectives are better for more abstract, global concepts. By using clever variations of the auditory-motor priming paradigm, Nicole showed that the auditory-motor associations subserving the auditory-motor priming effect are egocentric in nature.

4.2 Sonic Interaction Design

The CLOSED and SID projects studied auditory-motor interactions from a more applied perspective. The studies reported here were carried out at Ircam with the PDS group (Patrick Susini, Nicolas Misdariis, and Olivier Houix), and in collaboration with Karmen Franinović and Yon Visell at the Zürich Hochschule für der Künste (Z-HdK). Karmen Franinović and Yon Visell designed the Spinotron and the Flops.

To date, the use of sounds in human computer interfaces has been primarily limited to warning or feedback sounds (Stanton and Edworthy, 1999). Very often, these signals consist of various types of beeps, but Gaver (1986, 1989, 1994) proposed the use of *iconic* sound notifications in computer interfaces. The most famous auditory icon created by William Gaver is probably the sound of a sheet of paper being crumpled and thrown down to a garbage can, used as a feedback to file deletion, developed for Apple Computers. In this example, the meaning of the sound resulted from the causal relationship between the sound and a user's gesture: dragging an object on a surface.

Everyday objects are increasingly embedded with sensors, computation, and communication abilities. Designers are therefore considering sonic augmentations of these everyday objects, either for aesthetic purposes (such as enhancing the roar of a car engine) or for functional reasons (such as improving the usability of a browser on a music player with a small visual display). However, far less is known to guide the designer of continuous auditory feedback coupled to human action in real time. This idea of a continuous coupling of action and sound is exactly what happens when a person learns how to play a musical instrument. For example, to produce a good tone, a violinist bows a string, and (particularly during training) adjusts bowing action continuously by

²Nicole Navolio was particularly interested in transferring the audio-gesture idea to human-computer interactions. She went on to earn an HCI Master's degree at Carnegie Mellon University, and she is now a successful professional HCI designer. She has written a nice blog post explaining why the question of the frame of reference is important for her job: <http://www.summa.com/blog/the-future-of-gesture-sound-interfaces-mobile-technology-human-centered-design>, last retrieved on October 3, 2016.

listening to the sound that is produced. This sonic feedback guides the player's control, modifying bow speed, pressure, angle, and so forth. The idea of transposing this complex aspect of traditional musical instruments to the design of digitally augmented devices was the starting point of my foray into sonic interaction design.

However, only few practical examples of such functional sonic interactions are currently available (Müller-Tomfelde and Münche, 2001; Williamson et al., 2007; Rath and Schleicher, 2008; Cesarini et al., 2016). A very interesting example was developed during the Metason project³: it consisted of sonifying the handwriting of children with dysgraphia (Danna et al., 2013; Thoret et al., 2014). The idea was that the continuous sonic feedback would guide children in improving their handwriting, practically implementing the previous metaphor of the violin. Children were tested before and after training sessions. Training improved writing speed and fluency.

This example is however almost unique in that it developed a precise evaluation of the effect of the sonification. One of the reason of the scarcity of evaluation of sonic interactions is perhaps that it is difficult to define what should be evaluated, let alone how it should be evaluated. The two projects reported below developed evaluation methods for two aspects of sonic interactions: the ability of sounds to guide gestures, and the role of sound's pleasantness and naturalness during sound interaction.

4.2.1 Sounds to guide gestures: The Spinotron

The goal of this study was to assess the functionality of sounds in tangible interfaces, whether continuous auditory feedback could help users to control the interface more finely. It used an interface called the "Spinotron", specifically designed for this study (Lemaitre et al., 2009a). It is a tangible, one degree-of-freedom controller, based on the metaphor of the rotation of a virtual ratcheted wheel, driven in a manner analogous to a child's toy consisting of a spinning top, represented in Figure 4.3. Pumping the Spinotron causes the virtual wheel to spin and the virtual ratchet to clickety-clack.

The study examined whether the sound could guide users in learning how to control the Spinotron so as to drive the speed of the ratcheted wheel. The participants' task was to manipulate the Spinotron so as to maintain a target constant speed of the ratchet. The main experiment compared a group of participants interacting with the Spinotron without auditory feedback, and a group manipulating the Spinotron with feedback from the ratchet sounds. Participants alternated learning phases, in which they were provided with a visual feedback, and test phases in which there was no visual feedback. The performance of the users was measured by the time during which the speed of the ratchet stayed within the target speed area. Crucially, we examined how performance changed over different trials, to observe whether participants could learn how to manipulate the Spinotron more finely. Comparing the two groups showed that

³<http://metason.cnrs-mrs.fr>, last retrieved on 09/16/2016

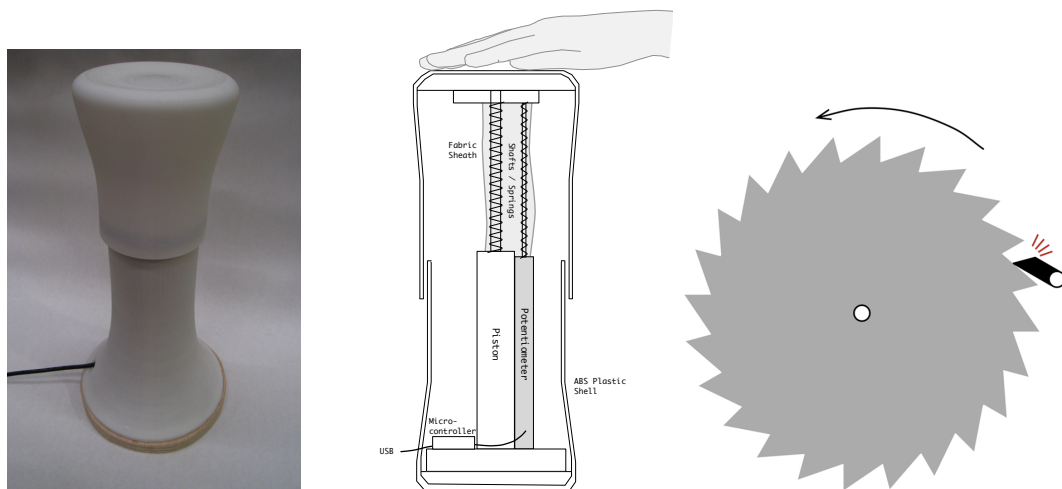


Figure 4.3: The interface used in Lemaitre et al. (2009a). Left: The Spinotron affords vertical pumping and generates digitally synthesized sounds. Middle: The configuration of the spinotron device. Right: A sound synthesis model was designed based on a ratcheted wheel, whose rotation is driven by the pumping motion of the Spinotron.

listening to the sound of the ratchet led to an improvement of performance along trials, indicating that the sounds guided users in learning how to adjust their control gestures. This was not the case when the users did not receive any auditory feedback: without sound, they could not improve their performance across trials, which indicates that they had not succeeded in learning how to control the Spinotron more finely. In this case, the positive effect of adding a continuous feedback to the interface was not that the interface was immediately more useable, but that sound guided the users in learning how to perform the gestures necessary to control the interface. Interestingly, we also replicated this experiment with two other groups of participants. In this case, the model controlling the motion of the virtual wheel did not correspond to a real physical phenomenon (i.e. the spinning top/ratcheted wheel), but was somewhat arbitrary. We observed the same results as with the ratchet model: performance improved across trials with auditory feedback, and the rate of improvement was similar.

4.2.2 Pleasantness and usability: the Flops

The study reported in Lemaitre et al. (2012) addressed the question of sonic interactions from another perspective: the emotions generated by the manipulation of such interfaces. Emotional reactions to the sounds of everyday products have been primarily studied in terms of pleasantness (Zwicker and Fastl, 1990) or annoyance (Guski et al.,



Figure 4.4: The Flops used in Lemaitre et al. (2012). Left panel represents a user pouring virtual balls out of the glass. Right panel represents the metaphor used to generate the sounds.

1999; see also Chapter 2). In particular, several empirical studies have highlighted systematic relationships between sound features and pleasantness (see in particular Västfjäll et al., 2003), when sound is the only potential source of emotion elicitation. However, the generality of these relationships is questionable, in particular in the context of sonic interactions. The interaction, itself, can induce emotions, depending, for example, on how successful is the manipulation with respects to the goals of the user. The goal of this study was to compare these two sources of emotion (sound and interaction) and evaluate whether sounds still influence users' feelings while they are performing tasks with a device instead of passively listening to sounds.

To address this issue, an interactive computational artifact was designed, called the *Flops*. It is a glass embedded with an orientation sensor allowing it to control the generation of virtual impact sounds when tilted. It implements the metaphor of a glass full of virtual items that may be poured out of it (see Figure 4.4).

The study used a dimensional approach to emotions that posits that emotions can be accounted for by two or three *dimensions* (Russell, 1980): *valence*, describing negative to positive feelings, *arousal*, describing the degree of arousal (from calm to excited), and *dominance*, describing how dominated or dominant the subjects feel. Accordingly, participants in our study reported their feelings on these three dimensions, by using the self-assessment manikin developed by Bradley and Lang (1994).

Two experiments were conducted to address these questions. The first experiment (passive listening) observed the feelings reported by listeners passively listening to a set of sounds associated to the Flops interface. Half of the sounds ("natural sounds") were recordings of real impact sounds. The other half ("synthetic sounds") had no clear mechanical counterpart (beeps and noises). The sound selection homogeneously sampled two psychoacoustic descriptors: spectral centroid and tonality index (see Chapter 2).

Participants watched videos of a person using the Flops (left panel of Figure 4.4), with the different sounds. The results showed that participants valued natural sounds with a lower spectral centroid more positively than synthetic sounds with a high spectral centroid (note that the result that higher frequencies make sounds less pleasant has been reported in many studies, see Chapter 2).

A selection of these sounds were used in a second experiment, in which users interacted with the Flops to perform several tasks, with the device being more or less usable, and we observed the contribution of the sounds and the usability to the feelings reported by the users. The results were that the sounds mainly influenced the valence of the emotions (although to a lesser extent than the usability of the interface, and to a lesser extent than in the passive listening experiment), whereas the usability of the device also influenced the arousal and the dominance dimensions of the feelings of the users. Overall, these results show that the aesthetics of the sounds contribute to the valence of the user experience of an interface, even when compared to other, larger manipulations. The results also suggested that pleasant sounds made the task feel slightly easier, and left the users feeling more in control. These results therefore argue that pleasant sounds in computer interfaces are desirable! Naturalness (defined here as whether the sounds were recordings of physical events or synthetic beeps) had little influence in comparison to the other factors. Another study conducted by Susini et al. (2012) suggested that what is in fact important is the perceived congruence between the user's gestures and the resulting sounds.

4.3 Discussion

The studies reported here confirm the intuition gained from previous chapter: audition and action do interact at the behavior level. The most remarkable result is that hearing a sound associated with an action primes that action. Whereas this was demonstrated in a rather experimental context in the priming studies, the Spinotron study shows that the tight coupling between auditory perception and motor behavior can be exploited in sonic interactive interfaces. The Flops study shows in turn that the aesthetics of the sounds should not be overlooked, and that the perceived congruence between the users' gestures and the resulting sounds contributes to the pleasantness of the experience. Here we discuss in more detail the findings of these studies.

Auditory-motor associations and memory representations

The fact that sounds can prime actions is interesting in itself. At a more theoretical level, it raises the question of the nature the auditory-motor associations subserving priming. A very minimal account of this phenomenon is that sensory and motor representations can dialogue. Such an idea is at the core of the *ideomotor* theory (Shin

et al., 2010). The central tenet of this theory is that the motor plans specifying an action are associated in memory with the *consequences* of that action in a distributed and *bidirectional* relationship (consisting of any perceivable consequences of an action, proximal or distal, including proprioceptive, haptic, visual, or auditory consequences). Activating any element of such a distributed representation may activate the whole representation: for example, activating the sensory consequences of an action may activate the motor plans producing that action and thus trigger or prime that action.

Whereas most of the literature has used visual stimuli, our study demonstrates that the ideomotor concepts also apply to sounds. In addition, it offers perspectives on the fate of the auditory-motor associations: how they are created, and how they persist and disappear over time. Overall, the results demonstrate that the gesture-sound associations mediating auditory-motor priming are extremely plastic: they can be formed very quickly in the context of the experiment and reconfigured as quickly when the contingent association is violated, even if they had been established during life-long experience (we found little evidence for any special status for ecological associations). Altogether, these results are strong indications of a type of memory representation that is flexible, in which older items are quickly replaced by new ones, akin to the traditional concept of short-term working memory and consistent with the properties of the dorsal stream of sensory processing.

From auditory-motor associations to sonic interactions

Can auditory-motor associations explain the results of the Spinotron study? The ideomotor theory is completely consistent with these results: manipulating the Spinotron created strong auditory-motor associations in the participants, which in turn helped them control the device. Thinking of the target sound (the regular tick tack of the ratchet spinning at the target pace) would have activated the corresponding motor plans, without the necessity to explicitly learn the dynamics of the model controlling the ratchet sound. For the group that received the audio feedback, that target was indicated auditorily during the training phase, and the audio feedback was also present during the test phase, and this probably created strong auditory-motor associations. For the other group, the target was indicated visually during the training phase, but the visual feedback was absent during the test phase (only the haptic and proprioceptive feedbacks were thus present). This probably led to visuo-motor associations that were too weak to be usable. However, our results do not show that auditory feedback would be more effective than a visual feedback. Interestingly, the Spinotron study showed the same rate of learning when the sound feedback was generated by an ecological model, or by a somewhat more arbitrary model. This is therefore consistent with the idea that auditory-motor associations do not rely on long-term memory but are created on-line and are extremely plastic.

From an applied perspective, these results indicate that sonic interactions can ben-

enefit from auditory-motor associations, but that these associations must be created and reinforced on-line, and cannot rely only on past experience (e.g. ecological associations). This gives freedom to the designers!

On the influence of naturalness

The Flops used sounds with different degrees of “naturalness” (using sounds that range from the natural consequence of the actions to arbitrary abstract beeps). Overall, listeners found natural sounds more pleasant, but this result has to be qualified. First, the usability of the interface influences the perception of naturalness: sounds are perceived as less natural when the interface is malfunctioning. Second, “natural” sounds do not make the manipulation more natural. It is the perceived congruency between a user’s gesture and the resulting sound that makes the whole experience feel natural.

More generally, the *pleasantness* of the sounds appears to be a more important factor. Relationships between auditory features and pleasantness (see Chapter 2) still hold in interactive contexts. In particular, sharpness, measured by the spectral centroid has a great influence on the emotions felt by the users. This feature therefore appears to be strong predictors of the users’ reaction to sounds, independently of whether they are solely passively listening to the sounds, or generating the sounds when interacting with an object: high spectral centroids (and/or synthetic sounds) induce negative emotions, and low spectral centroids (and/or natural sounds) induce positive emotions. Such result has been reported in many other studies (see for example Juslin and Timmers, 2010; Kumar et al., 2008; Takada et al., 2010; Västfjäll et al., 2003; Zwicker and Fastl, 1990).

Chapter 5

Vocal imitations of everyday sounds

Most of the studies reported so far have collected the listeners' descriptions of sounds (for interpreting perceptual dimensions, categories, or documenting interactions with prototypes). One observation was striking: very often, participants ran out of words and started *imitating* the sounds, vocally, but also using a lot of gestures, facial expressions, and pantomimes. In fact, French speakers usually have a very limited vocabulary specific to sounds, and thus rely on other linguistic devices to communicate about sounds (Faure, 2000). I conducted several studies over the years to investigate whether vocal imitations can communicate what they imitate, and how they do so.

I supervised the master's theses of Karine Aura-Rey and Arnaud Dessein, who conducted two initial studies of vocal imitations of everyday sounds. I then spent one year at the University of Venice with Davide Rocchesso to gather the preliminary results confirming the initial intuitions. The results clearly showed that imitations were the most natural way for most people to communicate about their auditory experience, and thus had a great potential for sound designers: vocal imitations could be the equivalent of the architect's sketchpad and pencil for sound designers. We were then lucky enough to be awarded a grant from the European Commission for the SkAT-VG project. During this project, I lead the research group focusing on perception and cognition. We studied what listeners imitate and how they imitate it. In particular, I supervised the master's theses of Ali Jabbari and Hugo Scurto. We also explored how to use vocalizations and gestures to create intuitive tools for sound designers.

5.1 Imitating sounds to communicate them

Imitations are extremely important during infants' development and learning of skills, customs, and behaviors (Chartrand and Bargh, 1999; Heyes, 2001; Meltzoff and Prinz,

2002), and occur in all sorts of situations in adults (matching postures, mannerisms, facial expressions, phonetic features, etc.). But can human vocalizations convincingly reproduce sounds made by *non-human* sources?

The short answer is yes. An important piece of evidence comes from linguistics. Although the most commonly agreed-upon view is that the relationships between signifier and signified (i.e. words and their meaning) is arbitrary (de Saussure, 1916), spoken languages also contain numerous instances of *sound symbolism* wherein the sound of a word is perceptually evocative of its meaning. Onomatopoeias (standardized words that mimic the sound of the object they refer to), ideophones (words that evoke sounds, movements, colors, or shapes by means of a similarity between the sound of the word and the idea it refers to), and phonesthemes (sublexical units referring to higher level attributes of meaning, e.g. “gl”, as in “glitter”, “glow”, “gleam” etc. relates to “vision” and “light”) are classical examples of words evoking some aspects of non-human sounds and other sensory impressions (Sobkowiak, 1990; Assaneo et al., 2011; Schmidtke et al., 2014; Blasi et al., 2016).

Onomatopoeias, ideophones, and phonesthemes are *words* (or parts of words), and as such are constrained by what they refer to, but also by the linguistic system they belong to. Borrowing the words of Rhodes (1994), they are “tame” imitations, as opposed to “wild” imitations, created by speaker on the spot. In comparison, wild vocal imitations have been only rarely (see for example Perlman et al., 2015)

In a first case study of wild vocal imitations of sounds conducted during Karine Aura’s master thesis (whose data were published later in Lemaitre et al., 2014), we observed how French speakers discuss about sounds. We observed couples of participants with limited musical or audio expertise. During the study, one participant listened to different series of sounds (in isolation), and then had to communicate one target sound in the series to the other participant. The task of the second participant was to recover the target sound. We did not specify anything about what to say, do, or not to do, and the participants were free to discuss as long as they liked. The goal was to observe what they would spontaneously use. We used sounds that were very easy or very difficult to identify. Conversations were manually annotated for the presence of vocal imitations. Here is an example of such conversation:

- “It sounds like as if you would take a piece of corrugated cardboard. First you scrape it, then you tear it off and it sounds like Rrrrr off the cardboard. You see? First, Fffff, and then Rrrrr. - Oh I see: Fffff then Rrrr.”

Vocal imitations were commonly used: they were present in 59.3% of the conversations. Conversations (79.6%) included even more often a number of imitative gestures.

5.2 Can listeners recognize what is imitated?

Onomatopoeias are in some cases better identified than the sounds they imitate (Lass et al., 1983), but can listeners also recognize what wild imitations refer to? To answer this question, a first study conducted by Karine Aura-Rey and analyzed by Arnaud Dessein compared how listeners classify a set of unambiguous kitchen sounds (the referent sounds) and vocal imitations of these sounds (Lemaitre et al., 2011). The first step consisted of recording imitations. Participants sat alone in a sound attenuated booth and used a custom interface that allowed them to listen to the different sounds, and record vocal imitations (see Figure 5.1). They were required to imitate sounds “in such a way that another person could recognize them.” They were instructed not to use any words or onomatopoeias. They could listen to their imitations and discard those they did not like¹ (we used variations of this method in many subsequent studies). Then, one set of participants freely sorted the referent sounds and another set sorted the imitations following the method reported in Section 3.1. The comparison of the two categorizations (referent sounds vs. corresponding imitations) showed that the listeners were able to recover the categories of sound events to which most vocal imitations corresponded.



Figure 5.1: Left: A participant in the SkAT-VG project recording vocal imitations of everyday sounds. Right: the interface used to record the sounds.

To investigate whether listeners can reach a finer-grained recognition, Lemaitre and Rocchesso (2014) compared how well listeners could recognize sounds when provided with verbal descriptions or vocal imitations. We used four type of sounds: identifiable complex events, identifiable elementary interactions, artificial sound effects (thus difficult to name), and unidentifiable mechanical sounds. A first set of participants recorded vocal imitations of the sounds as well as verbal portraits (short descriptions

¹The 72 imitations are available at <http://pds.ircam.fr/imitations.html>, date last viewed: 09/13/2016.

of the sounds)². Then, another set of participants were provided with a list of possible referent sounds for each imitation and each verbal portrait, and selected the referent sound that best matched the imitation or the verbal portrait. Figure 5.2 report the results of this experiment. As could be expected, recognition based on verbal portraits was maximally effective when the referent sounds were identifiable, and dropped when identifiability of the sounds decreased. More interestingly, recognition accuracy with vocal imitations did not depend on the identifiability of the referent sounds and was at least as high as with the best verbal portraits.

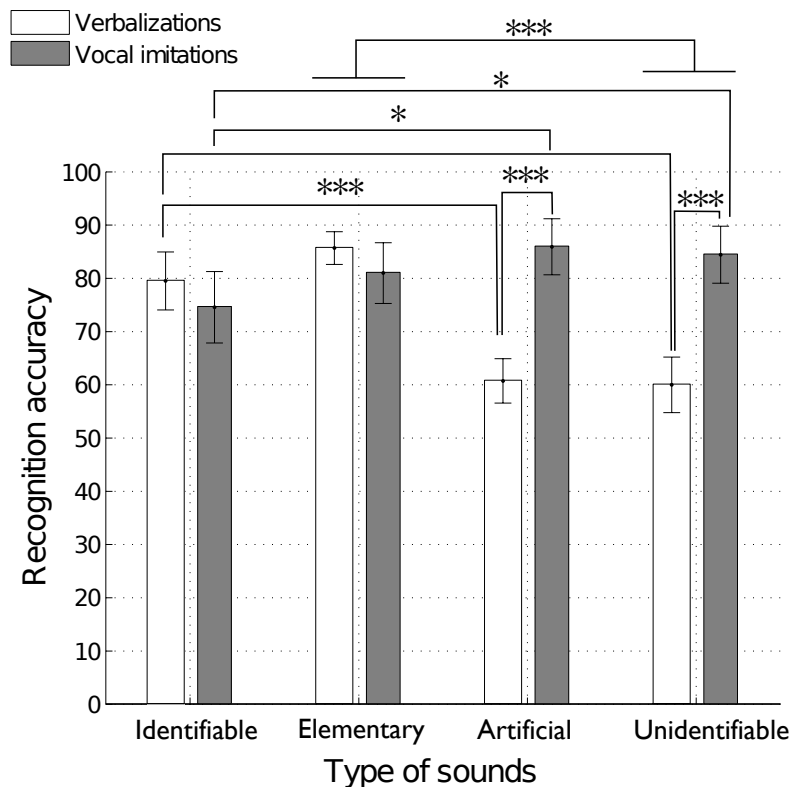


Figure 5.2: Recognition accuracy measured in Lemaitre and Rocchesso (2014). Participants selected the sound that corresponded to either a vocal imitation or a verbal description. Vertical bars represent the 95% confidence interval. Stars represent the significance of t-tests. Accuracy is systematically better with vocal imitations.

Because this experiment provided listeners with both the referent sounds and the imitations, it is possible that they based their judgments simply on the *similarity* between the imitation and the referent sound. In another study, our aim was thus to

²Examples of imitations are available here https://www.researchgate.net/publication/255703236_Sounds_for_Vocal_Imitations (date last viewed 09/13/2016)

assess whether listeners could identify what vocal imitations refer to *without hearing the referent sounds* (Lemaitre et al., In revision). In other words, the goal of this work was to study the semantic representations evoked by simply listening to vocal imitations of sounds. It used a yes-no paradigm: participants were provided with a sound and a verbal description of a target category, and simply indicated whether they thought the sound could correspond to the target category. We also used distractor sounds: sounds that did not correspond to the target categories, but had the same morphological profile as the target sounds (e.g. impulsive target sounds were paired with impulsive distractor sounds, etc.). The study compared vocal imitations produced by human participants to computational *auditory sketches* computed on the basis of sparse mathematical representations of the referent signals, with a method inspired by Suied et al. (2013) and Isnard et al. (2016). These auditory sketches are very interesting because they create degraded versions of the referent sounds by selecting the most energetic components of the sounds and thus minimizing the *overall* distance between the sketch and the referent sound (i.e. taking into account all the sound characteristics). Alternatively, human vocal imitations may *select only a few* important features, while discarding other irrelevant features.

The study used sounds from a large database of vocal and gestural imitations recorded during the SkAT-VG project (Houix et al., unpublished).³ We randomly selected ten imitators imitating referent sounds of manufactured products and basic mechanical interactions, following the method previously described. We also created three auditory sketches for each of the referent sound (with three levels of quality). In the main experiment, a sound was presented at each trial with a description of the target category and the participants indicated whether they felt that the sound corresponded to the description. The results showed that performance with the best vocal imitations was similar to the best auditory sketches for most categories of sounds. More detailed analyses showed that the acoustic distance between a vocal imitation and a referent sound was not sufficient to account for such performance. Instead of trying to reproduce the referent sound as accurately as vocally possible, vocal imitations focus on a few important features, which depended on each particular sound category.

5.3 How do people imitate sounds?

The previous results showed that vocal imitations give listeners access to very detailed representation of what is imitated. But what exactly do imitators do? Do they exactly reproduce all details of the referent sound? Is every kind of sound vocalizable?

The main limitation to what the voice can do probably comes from the glottal signal. The glottal signal is produced by a single vibrational system (the vocal folds), which

³A video documenting this project is available here: <https://vimeo.com/125024731>. Data are available at <https://zenodo.org/record/57468#.V4T1a6uM67A>, last retrieved on 09/13/2016.

implies that vocal signals are most often periodic (even though, chaotic, a-periodic or double-periodic oscillations can also happen), and essentially monophonic (even though some singing techniques can produce the illusion of multiple pitches). Furthermore, the pitch range of the human voice extends overall from about 80 Hz to 1100 Hz, and a single individual's vocal range usually covers less than two octaves. Another kind of limitation comes from speakers's native language. Speakers have a better ability to produce the speech sounds of their native language, and usually encounter utter difficulties when attempting to produce the sounds of a foreign language (Troubetzkoy, 1949; Strange and Shafer, 2008). Finally, some speakers may be better able to invent successful vocal imitations of a sound than other ones.

Ali Jabbari's master project studied how two expert and two lay participants reproduced four basic auditory features: pitch, tempo, sharpness and onset (Lemaitre et al., 2016; a similar attempt was reported by Mehrabi et al., 2015). It used synthetic referent sounds (modulated narrow-band noises and pure tones). The four participants recorded vocal imitations following the method previously described.⁴ We compared the features of the referent sounds and the imitations. Analyses identified three strategies: 1. Vocal imitations of pitch and tempo reproduced faithfully the absolute value of the feature. Tempo was very accurate, and the deviation between the pitch of the referent sounds and the imitations was about a few Hz; 2. Participants "transposed" the sharpness of the referent sounds within the constraints of their vocal apparatus (i.e. higher for female than for male participants) and matched relative differences rather than absolute values of sharpness by compressing or expanding the range of sharpness values; 3. Vocal imitations of onsets categorized the continuum of onset values into two discrete morphological profiles. Overall, these conclusions show that vocally imitating a sound does not amount in simple mimicry. Instead, the participants strive to find an appropriate strategy to convey the variations of this feature within the limits of their vocal capabilities. These strategies are diverse and specific to the different cases: vocal imitation was here not simple mimicry.

5.4 Imitations as a sound sketching tool

Several research groups have tried to harness vocalizations into practical applications. For example, onomatopoeias are ubiquitous in Japanese, and several studies have tried to automatically generate onomatopoeias from an environmental sound (Ishihara et al., 2003, 2004). More generally, many applications that use onomatopoeias as an input have been devised: diagnosis of the sound quality of industrial products (Takada et al., 2001; Bezat, 2007; Sciabica, 2011) or queries in sound database (Nakano et al., 2004; Gillet and Richard, 2005; Sundaram and Narayanan, 2006, 2008), and sound synthesis

⁴Available at <http://recherche.ircam.fr/equipes/pds/skat/LemaitreImitations.htm>, last retrieved on 09/14/2016.

(Nakano and Goto; Cartwright and Pardo; Cartwright and Pardo; Cartwright and Pardo, 2009; 2014; 2015a; 2015b; see also Del Piccolo and Rocchesso, 2016 for a review).

A similar idea motivates the SkAT-VG project: providing intuitive sketching tools for sound designers, using voice and gestures (this idea is developed in Rocchesso et al., 2015). Sketching is a very important aspect of graphical design and architecture but is not really used in sound design. Unlike in other areas of design, sound designers are not much used to teamwork, and often refrain from sharing preliminary sketches: they prefer to offer highly-refined realizations to their clients. Furthermore, tools that enable intuition, expressiveness, and cooperation are still missing in sound design. In particular, there are no tools and methods nearly as effective as hand and pencil, and providing an intuitive control of synthesis is a major research goal. For example, the project *senSons* aimed at developing tools using high-level semantic control and drawing (Ystad, 2010). Similarly, the goal of SkAT-VG was to develop technologies that use a designer's vocal or/and manual gestures to create a synthetic sound that can actually be reproduced, exchanged between collaborators, edited and refined.

In practice, my colleagues have developed several prototypes of applications fulfilling this goal (Houix et al., 2016; Rocchesso et al., submitted). For example, *MiMic* is a microphone-like interface represented in the left panel of Figure 5.3.⁵ It has two modes. First, the users' vocalization automatically selects among different possible of sounds synthesis (the system is based on a on-line machine learning technique). They can then play with the models, controlling the synthesis parameters with their voice and the motion of the microphone. *Mimes* is another approach based the work of Françoise (2013, i.e. mapping by demonstration) and represented in the right panel of Figure 5.3.⁶ Users first associate vocalizations and granular synthesis with gestures manipulating objects. Then can then manipulate the resulting sounds by manipulating the objects. The two prototypes are currently being fused.

These prototypes have been used in a number of workshops. In particular, the "48 hours of sounds design" provided five sound designers with our prototypes and requested them to create sound installations matching a selection of art installations in the vineyard of Château La Cost in Provence.⁷

5.5 Discussion

Taken altogether, these results show that vocal imitations are very effective for communicating everyday sounds. Vocal imitations are classified similarly to the referent sounds they imitate, vocal imitations are always associated with their referent sounds at least as well as the best verbal portraits, and vocal imitations are recognized as well

⁵Demonstration available here: <https://vimeo.com/142351022>, last retrieved on 09/14/2016.

⁶Demonstration available here: <https://vimeo.com/125367426>, last retrieved on 09/14/2016.

⁷Documented here: <https://vimeo.com/169521601>, last retrieved on October 4, 2016.



Figure 5.3: Two prototypes developed during the SkAT-VG project. Left: MiMic. Right: Mimes.

as optimized high-quality auditory sketches. In short, vocal imitations convey enough information to recover to identify the sounds.

Vocal imitations as caricatures

The interesting question is: how? It goes without saying that lay persons are for the most part completely unable to reproduce with their voice *exactly* a sound that they just have heard, or a sound idea that they have. Instead, our results suggest that imitators reproduce only a few characteristic features, and transpose them into their vocal range, or even exaggerate them. The comparison of human vocal imitations and algorithmic sketches by Lemaitre et al. (In revision) exemplifies this idea: whereas the identifiability of the algorithmic sketches could be predicted from the acoustic distance between the sketches and the referent sounds, this was not the case for the vocal imitations. Vocal imitations were in fact more acoustically dissimilar to the referent sounds than the auditory sketches, but they were still recognized equivalently.

On the basis of these observations, we can propose an interpretation of the imitators' strategy. They pick up a few features: those that they could easily render vocally and that they think could maximally distinguish the sounds. But these features are only distinctive of this particular set of sounds: the imitators select information on the basis of the task and the set of sounds. For example, they may use the duration of the sounds to distinguish some of the categories. But if duration is relevant for one particular set of sounds, the duration of the sounds cannot be thought of as characteristic of certain categories of sound events in general. This goes in line with idea mentioned in Chapter 2 that listeners do not have fixed low-dimensional representations of sounds: they can use different aspects of the sounds to adapt their behavior to the task (notice that McAdams et al., 2010 also reported that listeners could use different kinds of

information from a same set of sounds, depending on the task). In that account, vocal imitations are not very different from visual caricatures: the select and reproduce only the most important features for recognition, with some possible exaggeration.

Vocal imitations for exploring auditory cognition

The potential advantages of visual caricatures over more veridical representations have been highlighted for a long time. Early on, Ryan and Schwartz (1956) showed that an object was more quickly identified when represented as cartoons than as photographs or line drawings. Rhodes et al. (1987) asked subjects to name drawings of their colleagues and found faster reaction times for caricatures than for veridical line drawings, which in turn were recognized faster than anticaricatures (deemphasized features): this is the “caricature effect” (Stevenage, 1995; Lee et al., 2000).

The caricature effect has been a useful research paradigm to test different models of face recognition (Byatt and Rhodes, 1998; Rhodes et al., 1998; Lewis and Johnston, 1999; Hurlbert, 2001). So could vocal imitations be used to study sound recognition? To answer this question, our current work focuses on vocalizations of sounds from memory. In fact, in our work so far, there was a *referent sound* that imitators could listen to (as many times as necessary), and imitators were required to “reproduce” them with their voice. This paradigm allowed us to know exactly what the imitators were trying to vocalize. However, the situation may actually be different when the referent sound is not physically present at the time of the imitation (is in *memory*), or because there is no referent sound but the *idea* of a sound. Chapter 6 discusses this idea.

Gestures in imitations

The previous paragraphs have alluded to the fact that people use a lot of gestures when they imitate sounds. In fact, the SkAT-VG project recorded not only a database of vocal imitations, but also of gestural imitations (together with or independently from the voice). Hugo Scurto’s master thesis started to address the question of what the gestures communicate and their interplay with the voice. This study is still ongoing and the results are not published yet, but several conclusions are already available. First, imitators have a strong tendency to micmic the situation in which they imagine the sound was produced. They pantomime the actors, the scene, the objects, the actions, etc. When requested not to pantomime and concentrate on the sounds, we observed that gestures could follow different features of the sounds (pitch direction, tonalness, and granularity). As expected, imitators were using a spatial metaphor to communicate the pitch direction with gestures. More surprisingly, we also observed several other metaphors that were shared across the participants: they rapidly shook their hands whenever the signals were slow modulated (because of random fluctuations

or because the signals were noisy). Furthermore, the regularity of the shaking was indicative of the degree of randomness in the signals.

Chapter 6

Discussion and perspectives

This journey through everyday sounds is coming to a (momentary) halt. It has collected several findings: Listeners tend to perceive the sources of the sounds and not the sounds themselves; They identify more easily the actions than the objects that have caused the sounds; Identifying object properties involves more strongly semantic and visual cortical areas than identifying actions; Hearing the sound of an action primes executing that action; Sounds guide actions in human-computer interactions and influence the aesthetics of the interface; Vocal imitations are spontaneously used to communicate sounds and convey effectively the identity of the sounds; The effectiveness of vocal imitations result from a selection of auditory features that are relevant for sound identification. This last chapter takes a moment to reflect, discuss, and sometimes speculate about what these findings may teach us about auditory perception and cognition, and offers perspectives for new research.

Perception of sound sources

Untrained listeners rate, categorize, describe everyday sounds according to their sources, and not to their acoustic properties. For example, an early version of the vocal imitation protocol did not specify that participants had to imitate the *sounds* (Houix et al., unpublished). In that case, many participants described noise bands as wind gusts or rain textures, pure tones as spaceships or cartoon characters, a repeated tone as a ball bouncing on a glass pane, etc. They were in fact surprised when we suggested to think in terms of features (pitch, loudness). This does not mean that listeners are not able to judge these features: they just do not *spontaneously* focus on them.

Does it mean that listeners can always accurately identify the sound sources? Our results suggest that identification accuracy depends in fact on the type of property investigated. Our data suggest for example that distinguishing actions is easy, based on acoustic properties, whereas identifying material involves deeper cognitive processing

- at least for the sounds we studied. So far however, data are too scarce to make more definitive conclusions. For example, we have shown that gestures form the basic level of action representations. But we do not know if material forms such a stable representation of object properties. As such the comparison of action and material may have been unfair. Furthermore, the perception of the properties of sounding actions have been little studied, except for the velocity of moving objects (Guski, 1992; Carlile and Best, 2002; Neuhoff, 2004; Kaczmarek, 2005).

A systematic effort is thus needed to characterize more thoroughly the “psychomechanics”¹ of sound sources (McAdams et al., 2004). It is first important to not assume that any source property is identifiable in *any circumstances*. For example, research has focused on material, because there is an acoustic cue that unambiguously specifies material for impacts: the partials’ decay rate (the ring of the sound, Wright, 1971). Clean, isolated rings are however not present or masked in many sound events. It is thus not surprising that listeners rely on other, coarser strategies. For example, listeners often rely on spectral centroid to identify material (Lutfi and Oh, 1997), or the frequency of the first partial to identify the hollowness of bars (Lutfi, 2001). These features are not optimal but still provide the listeners with a reasonable guess.

I propose to first study how people rely on sounds to interact with their environment (by video documenting how they use a drill, a coffee machine, etc.) to establish what kind of information is reliably available. My proposal is to first observe normal-hearing persons and then assess how hearing impairment affects these interactions. Only then will it make sense to look for acoustic correlates of the relevant pieces of information. For this latter point I draw inspiration from the work of Robert Lutfi: observing the variability of the acoustic cues in the environment and the different listener strategies to use these cues. As the next paragraphs will argue, it will probably be necessary to move toward more complex representations of the signals than the usual psychoacoustic descriptors.

Dimensions and percepts

Listeners strive to make sense of what they hear and use whatever it takes to do it. When unambiguous information is available (e.g. material ringing), they use it. When relevant information is not fully available, they may rely on strategies that are sub-optimal in general, but will nevertheless provide them with some coarse approximation. This idea that listeners may use any kind of information that they deem relevant in a particular context and for a particular task was in fact proposed by McAdams et al. (2010) but is at odds with the Gibsonian idea that the sensory systems pick up invariant cues that unambiguously specify the environment.

¹I believe Georges Canevet coined this expression.

In other words, these results suggest a versatile auditory system that uses whichever pieces of auditory information fit its purpose. However, this idea also raises questions. Psychoacoustical studies have regularly identified a limited number of dimensions (pitch, duration, loudness, the dimensions of timbre, binaural cues), and brain imagery has identified brain areas that specialize in processing certain types of features (Belin et al., 1998; Poeppel, 2003; Belin et al., 2004; Rauschecker and Scott, 2009; Giordano et al., 2012). Vocal imitation studies also suggest that imitating only a few relevant features is sufficient to convey the identity of a sound source.

Does this mean that auditory information is coded as a limited set of features? Probably not. First, as stated before, the limited number of dimensions found by timbre studies could only reflect the structure of the sound sets used in these studies. Another piece of argument comes from music information retrieval. It turns out that the techniques that best recognize musical instruments, styles, etc. are not based on a few psychoacoustic descriptors: they use as many statistical information as needed (see Aucouturier and Bigand, 2012; Siedenburg et al., 2016, for discussion). Finally, Agus et al. (2010) have shown that repeated exposure to complex, abstract noise textures with little structure improves the performance at a repetition detection task. This suggests that long-term auditory memory for noise encodes some complex statistical properties of sounds, and that these properties may in fact be noise- and listener-dependent (see also Demany and Semal, 2008).

My view is that the auditory system *shapes* the information in a certain way, but does not reduce dramatically the *quantity* of information. Complex spectrotemporal representations in the auditory cortex account very well for instrument classification and timbre similarity judgements (Patil et al., 2012). Importantly, best classification is achieved only with representations that include joint spectro-temporal features, and not purely spectral or purely temporal features (see also Agus et al., 2012). I believe that the auditory system may have evolved to provide an optimized representation of the variability of the acoustic environment. For example, Smith and Lewicki (2006) have shown that the auditory filters could be thought of as a sparse, optimized coding system, implementing the most efficient way of representing the variability of environmental sounds. The consequence of such an idea is that even if there is little information reduction, certain auditory features are more directly accessible, because they are emphasized by the coding scheme. This would explain why certain features (e.g. spectral centroid) are repeatedly found in psychoacoustical studies: they both are good statistical descriptors of the environment and emphasized by auditory coding.

Auditory-motor interactions

The results of the auditory-motor priming studies question another aspect of sound representations. Our interpretation concluded that auditory-motor associations were

plastic and short-lived. But they still do require some sort of short-term memorization. Auditory codes need to be at least temporarily activated in a buffer and associated with motor plans. Such a buffer needs to store representations of different natures (sensory codes from different modalities and motor plans). It could also activate sensory-motor representations retrieved from long-term memory, such as when presenting the name of an object is sufficient to prime an action (Chao and Martin, 2000). This idea implies that motor programs and perceptual information are represented in a common domain. Grounded cognition approaches (Barsalou, 2008), the theory of event coding (Hommel et al., 2001), forward models (Jeannerod, 2006) offer theoretical frameworks that accommodate the combination and interaction of motor and sensory representations, but are not very specific about the nature of the representations.

Further work is thus needed to characterize auditory-motor associations. As with visually guided actions, the question of time and memory is crucial. My proposal is to study how the effect unfolds over time. For example, introducing a time delay between the learning and test phases or using incompatible training will further investigate the life span of the auditory-motor associations. I am in fact still puzzled by the fact that we could not observe a difference between arbitrary and ecological associations.

Another important question is the influence of actions on perception (i.e. the other way round). A few studies have for example shown that visual perception can be influenced by actions (Witt and Sugovic, 2013), but little has been done for auditory perception. For example, a recent piece presented by Teoma Naccarato and John Mac Callum has drawn my attention.² In this piece, the tempo of the music was controlled by the dancer's heartbeat. During rehearsal, the dancer could hear her own heartbeat, either sonified by a metronome or by a continuous signal. In the case of the metronome, she had the impression that the tempo was imposed by a machine whereas she had the impression that her own actions controlled the tempo with the continuous signal. I thus propose to study how different types of auditory-motor association may influence the perception of *agency*.

Sonic interaction design for sport

All these results are of great importance for the *design* of sonic interactions. So far, the practical applications of sonic interaction design have stayed mostly experimental. In my opinion, the most convincing applications have concerned sport (e.g. Bolívar and Bresin, 2012; Cesarini et al., 2016). In fact, sport activities require participants to learn how to execute very precise (and normally silent) motor behaviors. In that sense, practicing sport is very similar to practicing a musical instrument, but with little sound. So let us add continuous interactive sonic feedbacks to prime, guide, or

²<http://forumnet.ircam.fr/fr/event/seminaire-performance-john-maccallum-et-teoma-naccarato/>, last retrieved on September 23, 2016.

facilitate the learning of precise sport gestures! I believe in particular that a better understanding of auditory-motor associations would result in more appropriate design strategies. A project that I would like to develop is a mobile application for runners. Mobile phones and watches are now commonly embedded with motion sensors and localization devices. Training applications already exist, but the way they use sounds is usually very ineffective. I propose the following examples to illustrate how different principles of auditory-motor interactions could result in a better design. A first example is inspired by Marchal et al. (2013) and Turchet et al. (2015). By modifying the sound of the runner's footstep, the system makes the runner adjust the force on the ground, by simulating different ground types (solid, slippery, fragile, etc.). In a second example (inspired by Rath and Rocchesso, 2005 and Eriksson and Bresin, 2010), the vertical acceleration of the runner's center of gravity drives the tilt of a virtual plate that makes a ball roll. By learning how to control the speed of the ball, the runner optimizes his vertical displacement. In an even simpler example, the pace of the runner controls the playback speed of the music. By learning how to achieve a constant playback speed, the runner will control the regularity of his pace.

These examples show how continuous, interactive feedbacks may be used in a sport application. Motor rehabilitation, economic driving are other areas of applications where such continuous feedbacks could be beneficial.

Imitations and auditory cognition

Our work has clearly shown that vocal imitations convey the identity of sounds. It is now time to turn toward the second goal of the project: use vocal imitations as a tool to explore auditory cognition. Vocal imitations tell the researcher what are the important features for identifying a sound category. Thus, vocal imitations could be used to explore the content of auditory representations. It is however first important to acknowledge that they are constrained by the imitators' vocal abilities. As such, we can certainly not hope to study the detailed encoding of auditory representations. A more reasonable expectation is to highlight the types of information stored in *memory*.

In fact, auditory perception cannot be disentangled from memory. As sounds happen in time, memory is required to bind an incoming stream into a coherent whole (Demany and Semal, 2008). Any study that queries a listener's perception requires in fact the listener to reflect on a short-term store of a sound. In our experiments for example, imitations were always produced while or just after the imitators had been listening to the referent sounds. Therefore, they reflect how sounds are encoded in a short-term memory storage.

Little is known about auditory long-term memory (LTM). Most of the work so far has concerned speech and music, and little data is available for everyday sounds. Cohen et al. (2009), for example concluded that auditory LTM is poor compared to visual

LTM. A traditional view of LTM is that stimuli are stored as abstract concepts, and that most of the surface details are lost. However, studies with implicit memory tasks such as those of Agus et al. (2010) show that some details are still retained in memory (similar conclusions have been made for speech). Winkler and Cowan (2005) argue therefore that LTM maintains multiple redundant representations, including abstract regularities and sensory information.

I propose to use vocal imitations to study his question. Our initial plan is to use paradigms in which we separate in time the referent sounds and the imitations. For instance, we will first start by teaching a set of referent sounds to the participants, until they have memorized them with a very good accuracy (tested by an old/new paradigm for example). We will ask them to come back a few days or weeks later, and then imitate the memorized referent sounds. We will then compare imitations made while listening to the sounds and imitations from memory to highlight how the transition to LTM may have modified the auditory representations of the sounds.

Concrete, abstract, natural, artificial, ecological

Throughout this document I have used concepts such as “natural”, “artificial”, “concrete”, “abstract”, or “ecological” sounds. In fact, my research was initially motivated by two ideas: first I wanted to move from pure tones and bursts of noise and study the perception of the sounds of things. Second, I had the idea that sound design should also stop using beeps and tones and embrace ecological sounds. But there is a difficulty, which is not only lexical: what is a “natural sound”? As Richard Kronland-Martinet once commented, sound synthesis can produce sounds that sound exactly like the real thing. How a sound is produced (synthesis vs. mechanical production) is thus not operative. In my work, I often distinguish sounds whose mechanical cause is “unambiguously identified by listeners” (“concrete” sounds) from “abstract sounds”, which listeners cannot imagine a mechanical cause for. But this distinction is also not completely satisfying: in fact, listeners do often associate pure tones with a logical cause: a pure tone is an alarm, the beep of a microwave oven, etc. I think now that the results of the Flops and Spinotron studies point toward a more important distinction: sounds that listeners perceive as resulting from their actions, as opposed to these sounds that they do not perceive as resulting from their actions. In other words, this question has to be considered in terms of *agency*: how listeners feel that they are in control of the sounds that are produced. This distinction has practical consequences: sounds that are perceived as not under the control of the users are perceived as unpleasant and indicating a malfunctioning object. I believe that this is important for sound designers: the quality of a sound design does not lay in the sounds themselves, but in the dynamic relation between the sounds and the users’ actions.

References

- Aglioti, S. M. and Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research*. Published online 04 July 2010. DOI 10.1007/s00221-010-2344-x.
- Agus, T. R., Suied, C., Thorpe, S. J., and Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America*, 131(5):4124–4133.
- Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (2010). Rapid formation of robust auditory memories: insights from noise. *Neuron*, 66:610–618.
- Alain, C. and Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, 5:D202–D212.
- American Standard Association (1960). USA acoustical terminology S1.1-160. American Standard Association.
- Aramaki, M., Brancheriau, L., Kronland-Martinet, R., and Ystad, S. (2009). Perception of impacted materials: sound retrieval and synthesis control perspectives. In *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music: 5th International Symposium, CMMR 2008, Lecture Notes In Computer Science*, pages 134–146, Berlin Heidelberg, Germany. Springer-Verlag.
- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., and Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of cognitive neuroscience*, 22(11):2555–2569.
- Arnott, S. R. and Alain, C. (2011). The auditory dorsal pathway: orienting vision. *Neuroscience & Biobehavioral Reviews*, 35(10):2162–2173.
- Arnott, S. R., Cant, J. S., Dutton, G. N., and Goodale, M. A. (2008). Crinkling and crumpling: An auditory fMRI study of material properties. *Neuroimage*, 8:368–378.
- Assaneo, M. F., Nichols, J. I., and Trevisan, M. A. (2011). The anatomy of onomatopoeia. *PLoS ONE*, 6(12):e28317.

- Aucouturier, J.-J. and Bigand, E. (2012). Mel cepstrum & ann ova: The difficult dialog between mir and music cognition. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 397–402, Porto, Portugal. INESC TEC.
- Avanzini, F. and Rocchesso, D. (2001). Controlling material properties in physical models of sounding objects. In *Proceedings of the International Computer Music Conference (ICMC 2001)*, pages 91–94, La Habana, Cuba.
- Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):250–267.
- Ballas, J. A. and Howard, J. H. (1987). Interpreting the language of environmental sounds. *Environment and Behavior*, 19(1):91–114.
- Ballas, J. A. and Mullins, T. (1991). Effect of context on the identification of everyday sounds. *Human Performance*, 4(3):199–219.
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59:617–45.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *TRENDS in Cognitive Sciences*, 7(2):84–91.
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2003). fmri responses to video and point-light displays of moving humans and manipulable objects. *Journal of cognitive neuroscience*, 15(7):991–1001.
- Belin, P., Fecteau, S., and Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences*, 8(3):129–135.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., and Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of cognitive neuroscience*, 10(4):536–540.
- Bezât, M.-C. (2007). *Perception des bruits d'impact. Application au bruit de fermeture de porte automobile*. PhD thesis, Université de Provence Aix-Marseille 1.
- Bizley, J. K. and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, 14(10):693–707.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*.

- Bolibar, J. and Bresin, R. (2012). Sound feedback for the optimization of performance in running. *TMH-QPSR*, 32(1):39–41.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the semantic differential. *Journal of behavior Therapy and Experimental Psychiatry*, 25:49–59.
- Brass, M., Bekkering, H., Wohlschläger, A., and Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, 44:124–143. DOI: 10.1006/brcg.2000.1225.
- Bub, D. N. and Masson, M. E. J. (2010). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, 36(2):341–358.
- Byatt, G. and Rhodes, G. (1998). Recognition of own-race and other-race caricatures: implications for models of face recognition. *Vision research*, 38:2455–2468.
- Cant, J. S. and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, 17(3):713–731.
- Carlile, S. and Best, V. (2002). Discrimination of sound source velocity in human listeners. *Journal of the Acoustical Society of America*, 111(2):1026–1035.
- Cartwright, M. and Pardo, B. (2014). Synthassist: Querying an audio synthesizer by vocal imitation. In *Proceedings of the Conference on New Interfaces for Musical Expression, London, UK*, Goldsmiths University of London.
- Cartwright, M. and Pardo, B. (2015a). Vocalsketch: vocally imitating audio concepts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), 2015, Seoul, Republic of Korea*, pages 43–46, New York, NY. Association for Computing Machinery.
- Cartwright, M. and Pardo, B. (2015b). Vocalsketch: Vocally imitating audio concepts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 43–46. ACM.
- Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., and Grassi, M. (2010). When ears drive hands: the influence of contact sound to reaching grasp. *PLoS ONE*, 5(8):e12240.
- Cesarini, D., Calvaresi, D., Farnesi, C., Taddei, D., Frediani, S., Ungerechts, B. E., and Hermann, T. (2016). Mediation: An embedded system for auditory feedback of hand-water interaction while swimming. *Procedia Engineering*, 147:324–329.

- Chao, L. L. and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, 12:478–484.
- Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910.
- Chen, C.-C. (2011). Quantitative methodology: Appropriate use in research for blind baseball ergonomics and safety design. *The Journal of Human Resource and Adult Learning*, 9(4):6.
- Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences*, 106(14):6008–6010.
- Cummings, A., Čeponienė, R., Katoma, A., n, A. P. S., Townsend, J., and Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain research*, 115:92–107.
- Daniel, P. and Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acustica united with Acta Acustica*, 83:113–123.
- Danna, J., Velay, J.-L., Paz-Villagrán, V., Capel, A., Pétroz, C., Gondre, C., Thoret, E., Aramaki, M., Ystad, Y., and Kronland-Martinet, R. (2013). Handwriting movement sonification for rehabilitation of dysgraphia. In *Proceedings of the 10th Symposium on Computer Music Multidisciplinary Research*, pages 200–208. Laboratoire de Mécanique et d'Acoustique, Marseille, France.
- de Saussure, F. (1916). *Cours de linguistique générale*. Payot, Paris, France. 510 pages. Edited by Tullio de Mauro, 1995.
- Del Piccolo, A. and Rocchesso, D. (2016). Non-speech voice for sonic interaction: a catalogue. *Journal on Multimodal User Interfaces*, pages 1–17.
- Demany, L. and Semal, C. (2008). The role of memory in auditory perception. In *Auditory perception of sound sources*, pages 77–113. Springer.
- Engl, L. R., Frum, C., Puce, A., Walker, N. A., and Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate different cortical networks. *Neuroimage*, 47:1778–1791.
- Eriksson, M. and Bresin, R. (2010). Improving running mechanisms by use of interactive sonification. In *Proceedings of the 3rd Interactive sonification workshop*, Stockholm, Sweden.

- Faure, A. (2000). *Des sons aux mots: comment parle-t-on du timbre musical?* Unpublished doctoral dissertation, École de Hautes Études en Sciences Sociales, Paris, France.
- Françoise, J. (2013). Gesture–sound mapping by demonstration in interactive music systems. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 1051–1054. ACM.
- Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*, 87(1):311–322.
- Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human Computer Interactions*, 2(2):167–177.
- Gaver, W. W. (1989). The sonic finder: An interface that use auditory icons. *Human Computer Interactions*, 4:67–94.
- Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5(4):285–313.
- Gaver, W. W. (1993b). What do we hear in the world? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29.
- Gaver, W. W. (1994). Using and creating auditory icons. In Kramer, G., editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*. Westview Press.
- Gillet, O. and Richard, G. (2005). Drum loops retrieval from spoken queries. *Journal of Intelligent Information Systems*, 24(2/3):160–177.
- Giordano, B. L. (2003). Everyday listening: an annotated bibliography. In Rocchesso, D. and Fontana, F., editors, *The sounding object*, chapter 1, pages 1–16. Edizioni di Mondo Estremo.
- Giordano, B. L. and McAdams, S. (2006). Material identification of real impact sounds: Effect of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, 119(2):1171–1881.
- Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., and Belin, P. (2012). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral cortex*. First published online July 16, 2012.
- Giordano, B. L., McDonnell, J., and McAdams, S. (2010). Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. *Brain and cognition*, 73:7–19.

- Goldstone, R. L. and Kersten, A. (2003). Concepts and categorization. In Healy, A. F. and Proctor, R. W., editors, *Comprehensive handbook of psychology, Volume 4: experimental psychology*, chapter 22, pages 599–621. Wiley, Hoboken, NJ.
- Goodale, M. A. and Milner, D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25.
- Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception and Psychophysics*, 67(2):274–284.
- Grassi, M., Pastore, M., and Lemaitre, G. (2013). Looking at the world with your ears: how do we get the size of an object from its sound? *Acta Psychologica*, 143:96–104.
- Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology*, 61(1):54–63.
- Guski, R. (1992). Acoustic tau: An easy analogue to visual tau? *Ecological Psychology*, 4(3):189 – 197.
- Guski, R., Felscher-Suhr, I., and Schuemer, R. (1999). The concept of noise annoyance: How international experts see it. *Journal of Sound and Vibration*, 223(4):513–527.
- Guyot, F., Castellengo, M., and Fabre, B. (1996). Étude de la catégorisation d'un corpus de bruits domestiques. In Dubois, D., editor, *Catégorisation, représentation et systèmes symboliques*, chapter 2. Kimé, Paris.
- Gygi, B. and Shafiro, V. (2007). General functions and specific applications of environmental sound research. *Frontiers in bioscience*, 12.
- Gérard, Y. (2004). *Mémoire sémantique et sons de l'environnement (Semantic memory and environmental sounds)*. PhD thesis, Université de Bourgogne, Dijon, France.
- Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harshberger, M. L. (1997). Methodological issues in timbre research. In Deliège, I. and Sloboda, J. A., editors, *Perception and cognition of music*, chapter 12, pages 253–306. Psychology Press, Hove, East Sussex, UK.
- Handel, S. (1989). *Listening: an introduction to the perception of auditory events*, chapter 8: Identification of Speakers, Instruments, and Environmental Events, pages 219–274. The MIT Press, Cambridge, MA.
- Heyes, C. (2001). Causes and consequences of imitations. *TRENDS in cognitive sciences*, 5(6):253–261.

- Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological research*, 59(3):176–186.
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24:849–878.
- Houix, O., Lemaitre, G., Misdariis, N., Susini, P., and Urdapilleta, I. (2012). A lexical analysis of environmental sound categories. *Journal of Experimental Psychology: Applied*, 18(1):52–80.
- Houix, O., Lemaitre, G., Voisin, F., Marchetto, E., Bevilacqua, F., and Helgason, P. (unpublished). Database of vocal and gestural imitations elicited by sounds. Manuscript in preparation.
- Houix, O., Monache, S. D., Lachambre, H., Bevilacqua, F., Rocchesso, D., and Lemaitre, G. (2016). Innovative tools for sound sketching combining vocalizations and gestures. In *Proceedings of Audio Mostly*, Norrköping, Sweden. Visualization Center C.
- Hurlbert, A. (2001). Trading faces. *Nature neuroscience*, 4(1):3–5.
- International Telecom Union (2001-2003). Recommendation ITU-R BS.1534-1. Method for the subjective assessment of intermediate quality level of coding systems. International Telecom Union, Geneva, Switzerland.
- Ishihara, K., Nakatani, T., Ogata, T., and Okuno, H. G. (2004). Automatic sound-imitation word recognition from environmental sounds focusing on ambiguity problem in determining phonemes. In Zhang, C., Guesgen, H. W., and Yeap, W.-K., editors, *PRICAI*, volume 3157 of *Lecture Notes in Computer Science*, pages 909–918, Berlin, Heidelberg, New York. Springer.
- Ishihara, K., Tsubota, Y., and Okuno, H. G. (2003). Automatic transcription of environmental sounds into sound-imitation words based on japanese syllable structure. In *Proceedings of Eurospeech 2003*, pages 3185–3188, Geneva, Switzerland. International Speech Communication Association.
- Isnard, V., Taffou, M., Viaud-Delmon, I., and Sued, C. (2016). Auditory sketches: very sparse representations of signals are still recognizable. *PLOS One*, 11(3):e0150313.
- James, T. W., Stevenson, R. A., Kim, S., VanDerKlok, R. M., and Harman James, K. (2011). Shape from sound: evidence for a shape operator in the lateral occipital cortex. *Neuropsychologia*, 49:1807–1815. DOI: 10.1016/j.neuropsychologia.2011.03.004.

- Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford University Press, Oxford, UK. 220 pages.
- Juslin, P. and Timmers, R. (2010). Expression and communication of emotion in music performance. In Juslin, P. and Sloboda, J., editors, *Handbook of Music and Emotion: Theory, Research, and Applications*, pages 453–489. Oxford University Press, Oxford, UK.
- Kaczmarek, T. (2005). Auditory perception of sound source velocity. *Journal of the Acoustical Society of America*, 117(5):3149–3156.
- Kaplan, J. T. and Iacoboni, M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cognitive processing*, 8:103–113.
- Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual neurons and action recognition. *Experimental Brain Research*, 153:628–636.
- Krimphoff, J., McAdams, S., and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. In *Journal de Physique IV. Colloque C5, supplément au Journal de Physique III, volume 4, mai 1994*, volume 625-628.
- Krumhansl, C. (1989). Why is musical timbre so hard to understand? In Nielzen, S. and Olsson, O., editors, *Structure and perception of electroacoustic sound and music*, pages 43–53. Elsevier, Amsterdam, the Netherlands.
- Kubovy, M. and Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80:97–126.
- Kumar, S., Forster, H. M., Bailey, P., and Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *Journal of the Acoustical Society of America*, 124(6):3810–3817.
- Kunkler-Peck, A. J. and Turvey, M. T. (2000). Hearing shape. *Journal of Experimental psychology: human perception and performance*, 26(1):279–294.
- Lakatos, S., McAdams, S., and Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric forms. *Perception & psychophysics*, 59(8):1180–1190.
- Lass, N. J., Eastham, S. K., Wright, T. L., Hinzman, A. H., Mills, K. J., and Hefferin, A. L. (1983). Listener's identification of human-imitated sounds. *Perceptual and Motor Skills*, 57:995–998.

- Lebrun, N., Clochon, P., Etévenon, P., Lambert, J., Baron, J., and Eustache, F. (2001). An ERD mapping study of the neurocognitive processes involved in the perceptual and semantic analysis of environmental sounds and words. *Cognitive brain research*, 11:235–248.
- Lee, K., Byatt, G., and Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychological science*, 11(5):379–385.
- Lehet, M., Navolio, N., Pyles, J. R., Lemaitre, G., and Heller, L. M. (forthcoming). A neural bias for processing environmental sounds based on action-related information rather than concurrent material information. Manuscript submitted for publication.
- Lemaitre, G., Dessein, A., Susini, P., and Aura, K. (2011). Vocal imitations and the identification of sound events. *Ecological Psychology*, 23:267–307.
- Lemaitre, G. and Heller, L. M. (2012). Auditory perception of material is fragile, while action is strikingly robust. *Journal of the Acoustical Society of America*, 131(2):1337–1348.
- Lemaitre, G. and Heller, L. M. (2013). Evidence for a basic level in a taxonomy of everyday action sounds. *Experimental Brain Research*, 226(2):253–264.
- Lemaitre, G., Heller, L. M., Navolio, N., and Zúñiga-Peñaranda, N. (2015). Priming gestures with sounds. *PloS one*, 10(11):e0141791.
- Lemaitre, G., Houix, O., Misdariis, N., and Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1):16–32.
- Lemaitre, G., Houix, O., Susini, P., Visell, Y., and Franinović, K. (2012). Feelings elicited by the sounds of interactive everyday objects. *IEEE Transactions on Affective Computing*, 3(2):335–348.
- Lemaitre, G., Houix, O., Visell, Y., Franinović, K., Misdariis, N., and Susini, P. (2009a). Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron. *International Journal of Human-Computer Studies, special issue on Sonic Interaction Design*, 67:976–993.
- Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., and Susini, P. (In revision). Vocal imitations as a means to communicate everyday sounds. Manuscript submitted for publication. Plos One.

- Lemaitre, G., Jabbari, A., Misdariis, N., Houix, O., and Susini, P. (2016). Vocal imitations of basic auditory features. *The Journal of the Acoustical Society of America*, 139(1):290–300.
- Lemaitre, G., Pyles, J. R., Halpern, A. R., Navolio, N., Lehet, M., and Heller, L. M. (in revision). Who's that knocking at my door? Neural bases of sound source identification. Manuscript submitted for publication. *Cerebral cortex*.
- Lemaitre, G. and Rocchesso, D. (2014). On the effectiveness of vocal imitation and verbal descriptions of sounds. *Journal of the Acoustical Society of America*, 135(2):862–873.
- Lemaitre, G., Susini, P., Rocchesso, D., Lambourg, C., and Boussard, P. (2014). Non-verbal imitations as a sketching tool for sound design. In Aramaki, M., Derrien, O., Kronland-Martinet, R., and Ystad, S., editors, *Sound, Music, and Motion. Lecture Notes in Computer Sciences*, pages 558–574. Springer, Berlin, Heidelberg, Germany.
- Lemaitre, G., Susini, P., Winsberg, S., Letinturier, B., and McAdams, S. (2007). The sound quality of car horns: a psychoacoustical study of timbre. *Acta Acustica united with Acustica*, 93(3):457–468.
- Lemaitre, G., Susini, P., Winsberg, S., Letinturier, B., and McAdams, S. (2009b). The sound quality of car horns: Designing new representative sounds. *Acta Acustica united with Acustica*, 95(2):356–372.
- Lewis, J. W. (2006). Cortical networks related to human use of tools. *The neuroscientist*, 12(3):211–231.
- Lewis, J. W. (2010). Audio-visual perception of everyday natural objects - Hemodynamic studies in humans. In Naumer, M. J. and Kaiser, J., editors, *Multisensory object perception in the primate brain*, chapter 10, pages 155–190. Springer.
- Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., and DeYoe, E. A. (2005). Distinct cortical pathways for processing tools versus animal sounds. *The journal of neuroscience*, 25(21):5148–5158.
- Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., and Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of cognitive neuroscience*, 23(8):2079–2101.
- Lewis, J. W., Talkington, W. J., Tallaksen, K. C., and Frum, C. A. (2012). Auditory object salience: human cortical processing of non-biological action sounds and their acoustical signal attributes. *Frontiers in systems neuroscience*, 6:Article 27.

- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., and DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral cortex*, 14:1008–1021.
- Lewis, M. B. and Johnston, R. A. (1999). A unified account of the effects of caricaturing faces. *Visual cognition*, 6(1):1–41.
- Li, X., Logan, R. J., and Pastore, R. E. (1991). Perception of acoustic source characteristics: walking sounds. *Journal of the Acoustical Society of America*, 90(6):3036–3049.
- Lutfi, R. A. (1995). Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, 97(2):1333–1334. Letters to the Editor.
- Lutfi, R. A. (2001). Auditory detection of hollowness. *Journal of the Acoustical Society of America*, 110(2):1010–1019.
- Lutfi, R. A., Oh, E., Storm, E., and Alexander, J. M. (2005). Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians and non musicians. *Journal of the Acoustical Society of America*, 118(1):393–404.
- Lutfi, R. A. and Oh, E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America*, 102(6):3647–3656.
- Macmillan, N. A. and Creelman, C. D. (2005). *Detection theory. A user's guide*. Lawrence Erlbaum Associates, Mahwah, NJ, second edition. 292 pages.
- Marcell, M. M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of clinical and experimental neuropsychology*, 22(6):830–864.
- Marchal, M., Cirio, G., Visell, Y., Fontana, F., Serafin, S., Cooperstock, J., and Lécuyer, A. (2013). Multimodal rendering of walking over virtual grounds. In Steinicke, F., Visell, Y., Campos, J., and Lécuyer, A., editors, *Human Walking in Virtual Environments*, pages 263–295. Springer New York.
- Masson, M. E. J., Bub, D. N., and Breuer, A. T. (2011). Priming of reach and grasp actions by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5):1470–1484.
- Mathias, S. R. and von Kriegstein, K. (2014). Percepts, not acoustic properties, are the units of auditory short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2):445.

- McAdams, S. (1993). Recognition of auditory sound sources and events. In McAdams, S. and Bigand, E., editors, *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press, New York, NY.
- McAdams, S., Chaigne, A., and Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, 115(3):1306–1320.
- McAdams, S., Roussarie, V., Chaigne, A., and Giordano, B. L. (2010). The psychomechanics of simulated sound sources: Material properties of impacted thin plates. *Journal of the Acoustical Society of America*, 128:1401–1413.
- McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities and latent subject classes. *Psychological Research*, 58:177–192.
- Mehrabi, A., Dixon, S., and Sandler, M. B. (2015). Vocal imitation of pitch, spectral shape and loudness envelopes. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Malaga, Spain. ATIC Research Group of the University of Malaga.
- Meltzoff, A. N. and Prinz, W. (2002). *The imitative mind. Development, evolution and brain bases*. Cambridge University Press, Cambridge, UK. 364 pages.
- Merer, A., Ystad, S., Kronland-Martinet, R., and Aramaki, M. (2009). Semiotics of sounds evoking motions: categorization and acoustic features. In *Lecture Notes in Computer Science*, volume 4969, pages 139–158. Springer, Berlin/Heidelberg, Germany.
- Misdariis, N., Minard, A., Susini, P., Lemaitre, G., McAdams, S., and Parizet, E. (2010). Environmental sound perception: meta-description and modeling based on independent primary studies. *Eurasip Journal on Speech, Audio and Music Processing*, 2010. Article ID 362013,.
- Müller-Tomfelde, C. and Münche, T. (2001). Modeling and sonifying pen strokes on surfaces. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland.
- Nakano, T. and Goto, M. (2009). Vocalistener: A singing-to-singing synthesis system based on iterative parameter estimation. In *Proceedings of the Sound and Music Computing (SMC) Conference 2009*, pages 343–348, Porto, Portugal. The Sound and Music Computing Network.

- Nakano, T., Ogata, J., Goto, M., and Hiraga, Y. (2004). A drum pattern retrieval method by voice percussion. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, pages 550–553, Barcelona, Spain. The International Society for Music Information Retrieval.
- Navolio, N., Lemaitre, G., Forget, A., and Heller, L. M. (2016). The egocentric nature of action-sound associations. *Frontiers in Psychology*, 7. Perception Science, Article 231.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In Besner, D. and Humphreys, G. W., editors, *Basic processes in reading: visual word recognition*, chapter 9, pages 264–335. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Neuhoff, J. G. (2004). Auditory motion and localization. In Neuhoff, J. G., editor, *Ecological psychoacoustics*, pages 87–111. Brill, Leiden, The Netherlands.
- Orgs, G., Lange, K., Dombrowski, J.-H., and Heil, M. (2006). Conceptual priming for environmental sounds: an ERP study. *Brain and cognition*, 62:267–272.
- Parizet, E., Guyader, E., and Nosulenko, V. (2008). Analysis of car door closing sound quality. *Applied acoustics*, 69:12–22.
- Patil, K., Pressnitzer, D., Shamma, S., and Elhilali, M. (2012). Music in our ears: the biological bases of musical timbre perception. *PLoS Computational Biology*, 8(11):e1002759.
- Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, 2(8):150152.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech perception*, 41:245–255.
- Polotti, P. and Lemaitre, G. (2013). Rhetorical strategies for sound design and auditory display: A case study. *International Journal of Design*, 7(2):67–82.
- Rath, M. and Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE MultiMedia*, 12(2):60–69.
- Rath, M. and Schleicher, D. R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta Acustica united with Acustica*, 94:12–20.

- Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.
- Repp, B. H. (1987). The sound of two hands clapping: an exploratory study. *Journal of the Acoustical Society of America*, 81(4):1100–1109.
- Rhodes, G., Brennan, S., and Carey, S. (1987). Identification and rating of caricatures: implications for mental representations of faces. *Cognitive psychology*, 19:473–497.
- Rhodes, G., Carey, S., Byatt, G., and Proffitt, F. (1998). Coding spatial variations in faces and simple shapes: a test of two models. *Vision Research*, 38:2307–2321.
- Rhodes, R. (1994). Aural images. In Hinton, L., Nichols, J., and Ohala, J., editors, *Sound Symbolism*, pages 276–291. Cambridge University Press, Cambridge, UK.
- Risset, J.-C. and Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In Deutsch, D., editor, *The psychology of music*, Series in cognition and perception, pages 113–169. Academic Press, second edition.
- Rocchesso, D., Cera, A., Baldan, S., Monache, S. D., Bevilacqua, F., and Lemaitre, G. (submitted). Tools for sound sketching.
- Rocchesso, D., Lemaitre, G., and Grassi, M. (2013). Evidence for a spatial bias in the perception of sequences of brief tones. *The Journal of the Acoustical Society of America*, 133(5):EL346–EL350.
- Rocchesso, D., Lemaitre, G., Susini, P., Ternström, S., and Boussard, P. (2015). Sketching sound with voice and gesture. *ACM Interactions*, 22(1):38–41.
- Rogers, W. A., Lamson, N., and Rousseau, G. K. (2000). Warning Research: An Integrative Perspective. *Human Factors*, 42(1):102–139.
- Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B. B., editors, *Cognition and Categorization*, chapter 2, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.
- Ryan, T. A. and Schwartz, C. B. (1956). Speed of perception as a function of the mode of presentation. *American Journal of Psychology*, 69:60–69.

- Salzman, M. C., Dede, C., and Loftin, R. B. (1999). Vr's frames of reference: A visualization technique for mastering abstract multidimensional information. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 489–495. ACM.
- Schaeffer, P. (1966). *Traité des objets musicaux*. Seuil, Paris, France.
- Schmidtke, D. S., Conrad, M., and Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, 5. Article 80.
- Schön, D., Ystad, S., Kronland-Martinet, R., and Besson, M. (2009). The evocative power of sounds: conceptual priming between words and nonverbal sounds. *Journal of cognitive neuroscience*, 22(5):1026–1035.
- Sciabica, J.-F. (2011). *Caractérisation acoustique et perceptive du bruit moteur dans un habitacle automobile*. PhD thesis, Université de Provence Aix-Marseille 1.
- Serafin, S., Franinović, K., Hermann, T., Lemaitre, G., Rinott, M., and Rocchesso, D. (2011). Sonic interaction design. In Hermann, T., Hunt, A., and Neuhoff, J. G., editors, *Sonification handbook*, chapter 5, pages 87–110. Logos Verlag, Berlin, Germany.
- Shin, Y. K., Proctor, R. W., and Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136(6):943–974.
- Siedenburg, K., Fujinaga, I., and McAdams, S. (2016). A comparison of approaches to timbre descriptors in music information retrieval and music psychology. *Journal of New Music Research*, 45(1):27–41.
- Simon, J. R. and Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processings. *Journal of Applied Psychology*, 51:300–304.
- Smith, E. C. and Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439(23):978–982.
- Sobkowiak, W. (1990). On the phonostatistics of English onomatopoeia. *Studia Anglica Posnaniensia*, 23:15–30.
- Stanton, N. A. and Edworthy, J. (1999). Auditory warnings and displays: an overview. In *Human Factors in Auditory Warnings*. Ashgate Publishing Ltd., Farnham, UK. 375 pages.
- Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects? *British Journal of Psychology*, 86:127–146.

- Strange, W. and Shafer, V. (2008). Speech perception in second language learners: the re-education of selective perception. In Hansen Edwards, J. G. and Zampini, M. L., editors, *Phonology and Second language acquisition*, chapter 6, pages 153–192. John Benjamin Publishing Company, Philadelphia, PA.
- Suied, C., Drémeau, A., Pressnitzer, D., and Daudet, L. (2013). Auditory sketches: sparse representations of sounds based on perceptual models. In Aramaki, M., Barthelet, M., Kronland-Martinet, R., and Ivi Ystad, S., editors, *From Sounds to Music and Emotions, 9th International Symposium, CMMR 2012, London, UK, June 19-22, 2012, Revised Selected Papers*, volume 7900 of *Lecture Notes in Computer Science*, pages 154–170. Springer, Berlin/Heidelberg, Germany.
- Sundaram, S. and Narayanan, S. (2006). Vector-based representation and clustering of audio using onomatopoeia words. In *Proceedings of the American Association for Artificial Intelligence (AAAI) symposium series*, pages 55–58, Arlington, VA. American Association for Artificial Intelligence.
- Sundaram, S. and Narayanan, S. (2008). Classification of sound clips by two schemes: using onomatopoeia and semantic labels. In *Proceedings of the IEEE conference on multimedia and expo (ICME)*, pages 1341–1344, Hanover, Germany. Institute of Electrical and Electronics Engineers.
- Susini, P., Lemaitre, G., and McAdams, S. (2011). Psychological measurement for sound description and evaluation. In Berglund, B., Rossi, G. B., Townsend, J. T., and Pendrill, L. R., editors, *Measurement with persons - Theory, Methods and Implementation area*, chapter 11. Psychology Press/Taylor and Francis.
- Susini, P., McAdams, S., and Winsberg, S. (1999). A multidimensional technique for sound quality assessment. *Acustica united with Acta Acustica*, 85:650–656.
- Susini, P., McAdams, S., Winsberg, S., Perry, I., Vieillard, S., and Rodet, X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65(8):763–790.
- Susini, P., Misdariis, N., , Lemaitre, G., and Houix, O. (2012). Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal of Multimodal User Interfaces*, 5(3-4):175–186.
- Takada, M., Takeno, A., and ichiro Iwamiya, S. (2010). Effects of vehicle horn acoustic properties on auditory impressions and interpretation of reason for horn use by other drivers. *Noise Control Engineering Journal*, 58(3):259–272.

- Takada, M., Tanaka, K., Iwamiya, S., Kawahara, K., Takanashi, A., and Mori, A. (2001). Onomatopoeic features of sounds emitted from laser printers and copy machines and their contributions to product image. In *Proceedings of the International Conference on Acoustics ICA 2001, Rome, Italy*, Rome, Italy. Acoustical Society of Italy. CD-ROM available from <http://www.icacommission.org/Proceedings/ICA2001Rome/> (date last viewed 08/09/2013), paper ID: 3C.16.01.
- Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J.-L., and Ystad, S. (2014). From sound to shape: auditory perception of drawing. *Journal of Experimental Psychology: Human Perception and Performance*.
- Troubetzkoy, N. S. (1949). *Principe de phonologie*. Librairie Klincksieck, Paris, France.
- Tsingos, N., Gallo, E., and Drettakis, G. (2004). Perceptual audio rendering of complex virtual environments. *ACM Transactions on Graphics*, 23(3).
- Tucker, M. and Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116(2):185–203.
- Tucker, S. and Brown, G. J. (2003). Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds. In *Audio Engineering Society Convention 114*.
- Turchet, L., Camponogara, I., and Cesari, P. (2015). Interactive footstep sounds modulate the perceptual-motor aftereffect of treadmill walking. *Experimental brain research*, 233(1):205–214.
- van Petten, C. and Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: event related brain potential measures. *Neuropsychologia*, 33(4):485–508.
- Vanderveer, N. J. (1979). *Ecological acoustics: human perception of environmental sounds*. PhD thesis, Cornell University.
- Vetter, P., Smith, F. W., and Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, 24(11):1256–1262.
- Västfjäll, D., Kleiner, M., and Gärling, T. (2003). Affective reactions to interior aircraft sounds. *Acta Acustica united with Acustica*, 89:693–701.
- Whyatt, C. and Craig, C. M. (2013). Interceptive skills in children aged 9–11 years, diagnosed with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 7(5):613–623.

- Wildes, R. P. and Richards, W. A. (1988). Recovering material properties from sound. In Richards, W. A., editor, *Natural computation*, chapter 25, pages 356–363. A Bradford book. The MIT press, Cambridge, MA. London, England.
- Williamson, J., Murray-Smith, R., and Hughes, S. (2007). Shoogle: excitatory multimodal interaction on mobile devices. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 121–124, New York, NY, USA. ACM.
- Winkler, I. and Cowan, N. (2005). From sensory to long-term memory: evidence from auditory memory reactivation studies. *Experimental psychology*, 52(1):3–20.
- Winsberg, S. and Soete, G. D. (1993). A latent class approach to fitting the weighted Euclidian model, CLASCAL. *Psychometrika*, 58(2):315–330.
- Witt, J. K. and Sugovic, M. (2013). Catching ease influences perceived speed: Evidence for action-specific effects from action-based measures. *Psychonomic Bulletin and Review*, 20(6):1364–1370.
- Wright, P. (1971). Linguistic description of auditory signals. *Journal of applied psychology*, 55(3):244–250.
- Ystad, S. I. (2010). Vers le sens des sons: Modélisation sonore et contrôle haut niveau. Mémoire d'habilitation à diriger les recherches. Université de la Méditerranée-Aix-Marseille II.
- Zwicker, E. and Fastl, H. (1990). *Psychoacoustics Facts and Models*. Springer Verlag, Berlin, Heidelberg, Germany. 463 pages.

Appendices

Appendix A

Curriculum vitae

Personal address

32 rue Lantiez
75017 Paris, France.
Phone: +33 6 52 66 39 55
e-mail: GuillaumeJLemaitre@gmail.com
Born on October 5, 1977
Married, one daughter.

Professional address

STMS-IRCAM-CNRS-UPMC
Equipe Perception et Design Sonores
1 Place Stravinsky
75004 Paris, France.
Phone: +33 1 44 78 47 79
e-mail: lemaitre@ircam.fr

A.1 Positions

- 2014-2016 Research associate and work package leader,
STMS-IRCAM-CNRS, équipe Perception & Design Sonores, Paris, France.
European project SkAT-VG.
- 2013 Research engineer,
Genesis Acoustics, Aix-en-Provence, France.
- 2012 Research associate,
Università Iuav di Venezia, Research unit "Interaction", Venice, Italy.
- 2009-2011 Post-doctoral research associate,
Carnegie Mellon University, Department of Psychology, Dietrich College of Humanities and Social Sciences.
Center for Neural Bases of Brain Cognition (CNBC), Pittsburgh, USA.

- 2006-2009 Research associate and project co-coordinator,
STMS-IRCAM-CNRS, équipe Perception & Design Sonores, Paris, France.
European projects CLOSED, SID et MINET.
- 2004-2006 Post-doctoral research associate,
INRIA, projet REVES, Sophia-Antipolis, France. RNTL project OPERA.
- 2000-2004 Doctoral student,
STMS-IRCAM-CNRS, équipe Perception & Cognition Musicales, Paris, France.

A.2 Degrees

- 2004 PhD, acoustics - *Université du Maine*, Le Mans, France.
“Perceptual and acoustical study of new car warning signals”
Defended on 04/08/2004, highest honor (“très honorable avec félicitations”).
Supervisors: Xavier Meynial and Stephen McAdams.
- 2000 Master’s degree, acoustics - *Université du Maine*, Le Mans, France.
- 2000 Master’s degree, electrical engineering - *Institut Supérieur de l’électronique & du Numérique*, Lille, France.

A.3 Grants

- 2012 FP7 FET-Open European project SkAT-VG. Grant awarded in October 2013.
- 2012 Post-doctoral fellowship, *Università Iuav*, Venice, Italy.
- 2011/2012 Rothberg award for brain imagery, *Centre for Neural Bases of Brain Cognition (CNBC)*, Pittsburgh, PA, USA.
- 2008 MINET study visit grant, *Gösta Ekmans Laboratorium, University of Stockholm*, Stockholm, Sweden.
- 2003 Travel grant, Société Française d’Acoustique (SFA), *International Conference on Auditory Displays*, Boston University, Boston, MA, USA.

2000/2003 Doctoral scholarship, Ministère de l'industrie, *SCE Klaxon, Université du Maine, Ircam, Evreux, Le Mans, Paris, France.*

A.4 Organization of scientific events

- 2010 "Sound perception" session, French Congress on Acoustics, Lyon, April 2010.
- 2008 International Congress on Auditory Displays (ICAD 08), IRCAM, July 2008.
- 2008 "Ears Wide Open - Une tête, deux oreilles, trois dimensions - Comment évaluer la qualité des systèmes audio spatialisés ?" Symposium SFA/AES/Orange Labs, Rennes, 2008.
- 2007 Founding days of the Sound Perception Group, SFA, Lyon, January 2007.
- 2004 Second Symposium on Sound Design, Paris.
- 2001 "Journées Jeunes Chercheurs en Acoustique Musicale."

A.5 Committees

- 2016 Edward Collett: "Strategies of auditory categorization in cochlear implant users and normal hearing listeners." PhD thesis, cognitive sciences, Université Paul Sabatier, Toulouse, UMR "Cerveau et Cognition." Pascal Gaillard, Pascal Barone, and Olivier Deguine supervisors.
- 2014 Maureen Fontaine: "Voice recognition in healthy patients," master thesis. Integrative and cognitive neurosciences, Université Aix-Marseille, Institut des Neurosciences de la Timone. Pascal Belin supervisor.

A.6 Scientific expertise

Ad hoc reviewer for the following journal: ACM Transactions on Applied Perception, the Journal of the Acoustical Society of America, Journal of New Musical Research, Linguistics, International Journal of Human Computer Studies, Applied Acoustics, Journal of Automobile Engineering.

A.7 Administrative duties

Work package leader, Perception and cognition, European project SkAT-VG (2014-2016, two groups at Ircam).

Administrative and financial coordinator of the European project CLOSED (2006-2009, four academic partners, 1 million Euros).

Member of the French Society of Acoustics (SFA).

Elected member and treasury of the SFA's Sound Perception Group (2004-2009).

A.8 Outreach

2016 "What sound is this gesture?", Ben P. Stein. Interview et video reportage on the website American Institute of Physics: Inside Science. <https://www.insidescience.org/content/what-sound-gesture/3611>

2015 "Soon we will be able to design custom sounds with voice and gesture", Jennifer Ouellette. Article on the Skat-VG project in Gizmodo: <http://gizmodo.com/soon-we-will-be-able-to-design-custom-sounds-with-voice-1741536277>

2015 Demos at the Ircam Open days.

2012 Demos at the Research Night, University luav of Venice, Italy.

Appendix B

Research projects

2014-2016 - European project SkAT-VG

The European project SkAT-VG (Sketching Audio Technologies using Vocalizations and Gestures)¹ is a FET project (Future and Emerging Technologies) of the Seventh Framework Program. It is coordinated by Davide Rocchesso at the University Iuav of Venice, Italy. In addition, it includes three research groups at Ircam (Sound Perception and design, Interactions Sound/Music/Movement, and Sound Analysis and Synthesis), the KTH University of Stockholm, and the French company Genesis Acoustics. The goal of the project is to develop intuitive sketching tools for sound designers.

The SkAT-VG project aims at enabling designers to use their voice and hands, directly, to sketch the auditory aspects of an object, thereby making it easier to exploit the functional and aesthetic possibilities of sound. The core of this framework is a system able to interpret users' intentions through gestures and vocalizations, to select appropriate sound synthesis modules, and to enable iterative refinement and sharing, as it is commonly done with drawn sketches in the early stages of the design process. To reach its goal, the SkAT-VG project has several work packages: voice production, gesture analysis, cognitive psychology, machine learning, interaction design, and audio application development. The project tasks include case studies of how people naturally use vocalizations and gestures to communicate sounds, evaluation of current practices of sound designers, basic studies of sound identification through vocalizations and gestural production, gesture analysis and machine learning, and development of the sketching tools.

¹<http://skatvg.iuav.it>, last retrieved on September 19, 2016.

2013 - Genesis Acoustics

I conducted research for Genesis Acoustics² on sound quality for industrial customers. In particular, I developed a tool to evaluate the sound quality of aerodynamic car noises, based on psychoacoustical experiments. I also took part to methodological developments of psychoacoustical studies of aircraft fly-by noises (projects Parasoft and Neops). I was also in charge of professional training in psychoacoustics and sound quality.

2012 - Università Iuav di Venezia

The goal of this work with Davide Rocchesso was to draft the SkAT-VG project and gather preliminary results on vocal imitations of sounds. In particular, Davide Rocchesso and I showed that vocal imitations are effective descriptions of sounds, better recognized than verbal portraits (Lemaitre and Rocchesso, 2014).

I also had the occasion to collaborate with Massimo Grassi from the University of Padua, Italy. We carried out a research project on bouncing and colliding balls (Grassi et al., 2013; Rocchesso et al., 2013).

2009-2011 - Auditory lab at Carnegie Mellon University

This work was carried out in Laurie Heller's Auditory Lab (I was the lab manager).³ It consisted of four projects:

Neural Bases of Sound Identification (NBSI) This project was financed by a Rothberg Award for Brain Imaging. It consisted of fMRI studies of listeners identifying different properties of sound sources (action and material). We collaborated with Matt Lehet during his PhD.

Auditory-motor priming The goal of this project was to study how listening to the sound of an action may influence executing that action (Lemaitre et al., 2015; Navolio et al., 2016).

Material perception vs. action perception In this project, we crossed different actions and different materials and we studied how accurate and how fast listeners could identify these different properties (Lemaitre and Heller, 2012).

²<http://genesis-acoustics.com/en/accueil.php>, last retrieved on September 19, 2016.

³<http://www.psy.cmu.edu/~auditorylab/website/index/home.html>, last retrieved on October 4, 2016.

Semantic priming In this project, we studied how hearing a sound could prime a lexical decision (deciding whether a string of letters form a genuine English word), and compared different levels of description of the sounds (Lemaitre and Heller, 2013).

2006-2009 - European projects CLOSED, SID and MINET

The CLOSED project (Closing the loop of sound evaluation and design)⁴ aimed at closing the loop of sound evaluation and design. More specifically, it aimed at providing tools for evaluating the functional, aesthetically, and emotional aspects of sound design.

I carried out three projects:

- Free sorting of everyday sounds (Lemaitre et al., 2010; Houix et al., 2012),
- Vocal imitations of everyday sounds (Lemaitre et al., 2011),
- Sonic interactions with computer interfaces (Lemaitre et al., 2009a; Susini et al., 2012; Lemaitre et al., 2012; Polotti and Lemaitre, 2013).

During the CLOSED project, I also participated to two other “network” projects: MINET (Measuring the Impossible Network)⁵ and SID (Sonic Interaction design).⁶

2004-2006 - Project OPERA

The OPERA project (Optimisation Perceptive du Rendu Audio) was lead by Nicolas Tsingos. Its goal was to develop 3D audio rendering algorithms that can manage a very large number of spatialized sound sources (Tsingos et al., 2004).

During this project, I was in charge of:

- Developing a Matlab library of audio descriptors,
- Conducting perceptual evaluations of the algorithms.

⁴<http://closed.ircam.fr>, last retrieved on September 19, 2016.

⁵<https://minet.wordpress.com>, last retrieved on September 19, 2016

⁶http://www.cost.eu/COST_Actions/ict/IC0601, last retrieved on September 19, 2016

Appendix C

Teaching and mentoring

Table C.1 lists my teaching activities. Table C.2 lists the students I mentored. In particular, I have worked with three doctoral students.

First, I have collaborated with Emmanuel Gallo between 2004 and 2006 during my stay at INRIA. Emmanuel was conducting a PhD in computer science at the University of Nice-Sophia Antipolis on the spatialization of virtual audio scenes (directed by Georges Drettakis and Nicolas Tsingos). I supervised his work on the perceptive evaluation of audio rendering algorithms (chapter 7 of his thesis). We have published together a journal article in EURASIP Journal on Advanced Signal Processing [24] and a conference article at the International Conference on Auditory Displays in Limerick, Ireland, 2005 [44]. Emmanuel has received his PhD in 2006.

I have collaborated in 2009 with Arnaud Dessein during the CLOSED project. During the first year of his PhD in computer science (directed by Gérard Assayag and Arshia Cont), we have continued a study initiated during his second master project on auditory scene analysis (with Arshia Cont). We have published an article the Lecture Notes in Computer Sciences [16] and presented the results at the meeting of the International Society for Music Information Retrieval Conference in 2010 [36]. Arnaud has received his PhD in 2012.

I have finally collaborated with Matthew Lehet between 2013 et 2016. Matthew is pursuing a PhD in psychology at Carnegie Mellon University (directed by Lori Holt). We have collaborated in two fMRI studies. These studies were founded in part by the Rothberg Award for Brain Imagery I received from the CNBC. These studies were his PhD's second year project and are described in detail in Chapter 3. We have recently submitted two journal papers corresponding to these two studies [25, 26].

Degree	Year	University/location	Theme	Duration (hours)
Professional teaching	2013	Genesis Acoustics	Psychoacoustics and sound quality	60
L3	2011-2012	Conservatory C. Pollini, Padua, Italy	Musical acoustics	12
M2P	2009-2010 2008-2009 2007-2008	University Pierre et Marie Curie, Paris, France, Masters Acoustique Industrielle (ACIND) and Acoustique Architecturale (ACAR).	Sound perception and sound quality	99
Engineering school (M2P)	2007-2008	Conservatoire National des Arts et Métiers (CNAM)	Psychoacoustics. Sound quality	26
Engineering school (M2P)	2008-2009 2007-2008 2006-2007 2005-2006 2004 -2005	Ecole Nationale Supérieure d'Ingénieurs du Mans (ENSIM)	Sound quality	18.75
Engineering school (M2P)	2005-2006 2004-2005	Institut Supérieur de Mécanique de Paris	Sound quality	9
Deug 2 (L2)	2003-2004	Université Paris 8. IED. Deug 2 Psychology	Experimental methology in Psychology. Statistics	68.7
Total				293.45

Table C.1: Teaching activities.

Période	Nom	Degree	Subject	Ratio	Pub.	Current situation
2013-2016	Matthew Lehet	PHD , psychology, Carnegie Mellon University	Sound identification, neuroscience	10%	2JS	PhD program
2015	Hugo Scurto	Master 2 ATIAM, UPMC/ENSCachan	Vocal imitations of sounds	70%	2C	PhD program
2014-2015	Nicole Navolio	Master HCI, Carnegie Mellon University	Auditory-motor interactions	50%	2J, 2C	Engineer with Summa Technologie
2011-2013	Nicole Navolio	BSc de Psychology, Carnegie Mellon University	Sound identification, neuroscience	40%	2JS	
2014	Hugo Trad	Licence Science and Philosophy (Biology) Université Paris 4 and UPMC	Music Psychology	30%	-	Cogmaster
2014	Ali Jabbari	Master 2 Arts, Sciences, Technologies, INP Grenoble	Vocal imitations of sounds	100%	1J, 1C	PhD program
2013	Sylvain Mysliwiak	Master 1 acoustics UPMC	Applied psychoacoustics	100%	-	Engineer, Marie de Paris
2013	Loic Eymar	First year of engineering school ENSAM, Aix-en-Provence	Applied psychoacoustics	100%	-	Consultant for Dassault aviation
2012	Daniele Galante	Last year of sound engineering , Conservatory Cesare Pollini, Padoue, Italie	Vocal imitations of sounds	100%	-	Sound designer at Onami Games
2010-2011	Nicolas Zúñiga-Peñaranda	BSc Biology and Pyschology, Carnegie Melllon University	Auditory-motor interactions	50%	1J, 2C	Medical coordnator for Americorps
2010-2011	Clara Baron-Hyppolite	BSc Psychology, Carnegie Mellon University	Sound identification, neuroscience	50%	2C	Education sector
2010-2011	Jayant Bhambhani	BSc Psychology, Carnegie Mellon University	Sound identification, neuroscience	50%	1C	Engineer, Cinch Financial
2009	Arnaud Dessein	PhD computer science, Université Pierre et Marie Curie	Signal processing	5%	1J	Post-doctoral research associate, University of Bordeaux
2009	Arnaud Dessein	Master 2 ATIAM, UPMC	Signal processing	50%	1C	
2008	Arnaud Dessein	Last year of engineering school Ecole Centrale de Lille	Vocal imitations of sounds	100%	1J, 5C	
2007	Karine Aura	Master 2 , Linguistics, Université Toulouse Le-Mirail	Vocal imitations of sounds	100%	1J, 5C	Clinical neuropsycholinguist
2007	Mathieu Pellerin	Last year sound engineering Institut Supérieur des Techniques du son, Ecole Supérieure de Réalisation Audiovisuelle, Paris	Applied psychoacoustics	100%	-	Freelance sound designer
2004-2006	Emmanuel Gallo	PhD computer science, université de Nice Sophia Antipolis	Applied psychoacoustics	10%	1J 1C	Engineer, Autodesk

Table C.2: List of mentored students. Pub.=Publications; J=Journal article; C.=Conference article; JS=submitted journal article.

Appendix D

List of publications

My publications are listed below. The names of the students I mentored are underlined.

International journals		
PLOS One	IF 5 years: 3.70	1
IEEE Transactions on Affective Computing	IF 5 years: 3.19	1
Acta Psychologica	IF 5 years: 2.61	1
Journal of Experimental Psychology: applied	IF 5 years: 2.60	2
Frontiers in Psychology	IF 2014: 2.56	1
Experimental Brain Research	IF 5 years: 2.42	1
International Journal of Human-Computer Studies	IF 5 years: 1.83	1
Journal of the Acoustical Society of America	IF 5 years: 1.74	4
Ecological Psychology	IF 5 years: 1.53	1
Applied Acoustics	IF 5 years: 1.27	2
International Journal of Design	IF 5 years: 1.19	1
EURASIP Journal on Advanced Signal Processing	IF 5 years: 1.07	1
Journal on Multimodal User Interface	IF 5 years: 0.94	1
Acta Acustica united with Acustica	IF 5 years: 0.82	2
EURASIP Journal on Audio, Speech, and Music Processing	IF 5 years: 0.47	1
Lecture Notes in Computer Sciences	IF 2015 : 0.30	2
ACM Interactions	-	1

Table D.1: Publication list by journal. I used the 5-year impact factor when provided by Research Gate. Otherwise I used the last available year.

International journals	24
Chapter in books	2
International conferences (articles)	18
International conferences (abstracts)	20
(invited presentations)	(3)
National conferences	6

Table D.2: Publications by type of publication.

Journal articles (24)

Psychoacoustics and auditory perception

- [1] 2016 Guillaume Lemaitre, Ali Jabbari, Nicolas Misdariis, Olivier Houix, & Patrick Susini, "Vocal imitations of basic auditory features." *The Journal of the Acoustical Society of America*, 139(1), 290-300.
- [2] 2016 Nicole Navolio, Guillaume Lemaitre, Alain Forget & Laurie M. Heller, "The egocentric nature of action-sound associations." *Frontiers in Psychology*, Perception Science, vol. 7, Article 231.
- [3] 2015 Guillaume Lemaitre, Laurie M. Heller, Nicole Navolio & Nicolas Zúñiga-Peñaranda, "Priming gestures with sounds." *PLOS One* 10.11: e0141791.
- [4] 2014 Guillaume Lemaitre & Davide Rocchesso, "On the effectiveness of vocal imitations and verbal descriptions of sounds". *The Journal of the Acoustical Society of America*, vol. 135(2), pp. 862-873.
- [5] 2013 Massimo Grassi, Massimiliano Pastore & Guillaume Lemaitre "Looking at the world with your ears: How do we get the size of an object from its sound?", *Acta Psychologica*, vol. 142(1), pp. 96-104.
- [6] 2013 Guillaume Lemaitre & Laurie M. Heller, "Evidence for a basic level in a taxonomy of everyday action sounds", *Experimental Brain Research*, vol. 226(2), pp. 253-254.
- [7] 2013 Davide Rocchesso, Guillaume Lemaitre & Massimo Grassi, "Evidence for a spatial bias in the perception of sequences of brief tones". *The Journal of the Acoustical Society of America*, *Express Letters*, vol. 133(5), pp. EL346-350.
- [8] 2012 Guillaume Lemaitre & Laurie M. Heller, "Auditory perception of material is fragile, while action is strikingly robust", *Journal of the Acoustical Society of America*, vol. 131(2).

- [9] 2012 Olivier Houix, Guillaume Lemaître, Nicolas Misdariis & Patrick Susini, "A lexical analysis of environmental sound categories", *Journal of Experimental Psychology: applied*, vol. 18(1), pp. 52-80.
- [10] 2011 Guillaume Lemaître, Arnaud Dessein, Karine Aura & Patrick Susini, "Vocal imitations and the identification of sound events", *Ecological Psychology* 23, pp. 267-307.
- [11] 2010 Guillaume Lemaître, Olivier Houix, Nicolas Misdariis & Patrick Susini, "Listener expertise and sound identification influence the categorization of environmental sounds", *Journal of Experimental Psychology: applied*, vol. 16(1), pp. 16-32.

Sound quality and design

- [12] 2015 Davide Rocchesso, Guillaume Lemaître, Patrick Susini, Sten Ternström & Patrick Boussard, "Sketching Sound with Voice and Gesture". *ACM Interactions*, Vol. 22 (1), pages 38-41.
- [13] 2015 Guillaume Lemaître, Christian Vartanian, Christophe Lambourg & Patrick Boussard, "A psychoacoustical study of wind buffeting noise", *Applied Acoustics*, 9: 1-12.
- [14] 2014 Guillaume Lemaître, Patrick Susini, Davide Rocchesso, Christophe Lambourg & Patrick Boussard, "Non-Verbal Imitations as a Sketching Tool for Sound Design". In Mitsuko Aramaki, Olivier Derrien, Richard Kronland-Martinet, Sølvi Ystad, editors, *Sound, Music, Motion*, Lecture Notes in Computer Sciences 8905, pp. 558-574, Springer, Berlin/Heidelberg, Germany.
- [15] 2013 Pietro Polotti & Guillaume Lemaître "Rhetorical strategies for sound design and auditory display: A case study". *International Journal of Design*, vol. 7(2), pp. 67-82.
- [16] 2013 Arnaud Dessein, Arshia Cont & Guillaume Lemaître "Real-Time Detection of Overlapping Sound Events with Non-Negative Matrix Factorization", In Frank Nielsen and Rajendra Bhatia, editors, *Lecture Notes in Computer Sciences: Matrix Information Geometry*, chapter 14, pages 341–371. Springer, Berlin/Heidelberg, Germany, 2013
- [17] 2012 Guillaume Lemaître, Olivier Houix, Patrick Susini, Yon Visell, & Karmen Franinović, "Feelings elicited by the sounds of interactive everyday objects". *IEEE Transactions on Affective Computing*, vol. 3(3), pp. 335-348.
- [18] 2012 Patrick Susini, Nicolas Misdariis, Guillaume Lemaître & Olivier Houix, "Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback", *Journal on Multimodal User Interface*, vol 5 (3-4), pp. 175-186.

- [19] 2010 Nicolas Misdariis, Antoine Minard, Patrick Susini, Guillaume Lemaitre, Stephen McAdams & Etienne Parizet, "Environmental sound perception: meta-description and modeling based on independent primary studies", *EURASIP Journal on Audio, Speech, and Music Processing*, Volume 2010. Article ID 362013.
- [20] 2009 Guillaume Lemaitre, Olivier Houix, Yon Visell, Karmen Franinović, Nicolas Misdariis & Patrick susini, "Toward the Design and Evaluation of Continuous Sound in Tangible Interfaces: The Spinotron", *International Journal of Human-Computer Studies*, vol. 67, pp. 976-993.
- [21] 2009 Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg, Boris Letinturier, & Stephen McAdams, "The sound quality of car horns: Designing new representative sound", *Acta Acustica united with Acustica*, vol. 95(2), pp. 356-372.
- [22] 2008 Guillaume Lemaitre, Boris Letinturier & Bruno Gazengel "Model and estimation method for predicting the sound radiated by a horn loudspeaker – With application to a car horn", *Applied Acoustics*, vol. 69(1), pp. 47-59.
- [23] 2007 Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg, Boris Letinturier, & Stephen McAdams, "The sound quality of car horns: a psychoacoustical study of timbre", *Acta Acustica united with Acustica*, vol. 93(3), pp. 457-468.
- [24] 2007 Emmanuel Gallo, Nicolas Tsingos & Guillaume Lemaitre, "3D-Audio matting, post-editing and re-rendering from field recordings", *EURASIP Journal on Advanced Signal Processing, special issue on Spatial Sound & Virtual Acoustics*, vol. 2007, Article ID 47970.

Submitted articles (3)

- [25] In revision Guillaume Lemaitre, John R. Pyles, Andrea R. Halpern, Nicole Navolio, Matthew Lehet, & Laurie M. Heller "Is that banging or is that made out of wood? An fMRI study of switches of auditory attention to different properties of sound sources". In revision at Cerebral Cortex.
- [26] Submitted Matthew Lehet, Nicole Navolio, John A. Pyles, Guillaume Lemaitre, & Laurie M. Heller "A neural bias for processing environmental sounds based on action-related information rather than concurrent material information".
- [27] In revision Guillaume Lemaitre, Olivier Houix, Nicolas Misdariis, & Patrick Susini, "Identifying vocal imitations". In revision at Plos One.

Chapters in book (2)

- [28] 2011 Patrick Susini, Guillaume Lemaitre & Stephen McAdams, "Psychological measurement for sound description and evaluation", in B. Berglund, Giovanni B. Rossi, James T. Townsend & Leslie R. Pendrill (Eds.) chapter 11, *Measurement with persons - Theory, Methods and Implementation area*, Psychology Press/Taylor and Francis.
- [29] 2011 Stefania Serafin, Karmen Franinović, Thomas Hermann, Guillaume Lemaitre, Michal Rinott & Davide Rocchesso, "Sonic Interaction Design", in T. Hermann, A. Hunt and J. Neuhoff (Eds.) *Sonification Handbook*. Logos Verlag, Berlin, 2011.

International conferences (37)

Articles

- [30] 2016 Olivier Houix, Stefano Delle Monache, Hélène Lachambre, Frédéric Bevilacqua, Davide Rocchesso, and Guillaume Lemaitre. "Innovative tools for sound sketching combining vocalizations and gestures." Proceedings of Audio Mostly, Norrköping, Sweden, 2017.
- [31] 2016 Diemo Schwarz, Guillaume Lemaitre, Mitsuko Aramaki, & Richard Kronland-Martinet. "Effects of Test Duration in Subjective Listening Tests". Proceedings of the International Computer Music Conference (ICMC). Utrecht, the Netherlands, 2016.
- [32] 2013 Guillaume Lemaitre, Davide Rocchesso, Patrick Susini, Christophe Lambourg & Patrick Boussard, "Using Vocal Imitations for Sound Design", in *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, Marseille, France
- [33] 2012 Clara Charlotte Baron-Hyppolite, Guillaume Lemaitre & Laurie M. Heller, "Statistical approaches to mapping perceptual sound categories and acoustic properties", in *Proceedings of the meeting of the Eastern Psychological Association (EPA 2012)*, Pittsburgh, PA.
- [34] 2012 Jayant Bhamhani, Clara Charlotte Baron-Hyppolite, Guillaume Lemaitre, Andrea R. Halpern, & Laurie M. Heller, "Identification of the causes of sounds: Equating actions and materials", in *Proceedings of the meeting of the Eastern Psychological Association (EPA 2012)*, Pittsburgh, PA.
- [35] 2012 Nicolas Zúñiga-Peñaranda, Guillaume Lemaitre & Laurie M. Heller, "Priming of manual actions by their related sounds", in *Proceedings of the meeting of the Eastern Psychological Association (EPA 2012)*, Pittsburgh, PA.

- [36] 2010 Arnaud Dessein, Arshia Cont & Guillaume Lemaitre, "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence", in *Proceedings of the Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, the Netherlands.
- [37] 2009 Patrick Susini, Nicolas Misdariis, Olivier Houix & Guillaume Lemaitre, "Does a natural sonic feedback affect the perceived usability and emotion in a context of use?" in *Proceedings of the Sound & Music Computing Conference (SMC'09)*, Porto, Portugal.
- [38] 2009 Guillaume Lemaitre, Olivier Houix, Karmen Franinović, Yon Visell & Patrick Susini, "The Flops glass: a device to study emotional reactions arising from sonic interactions" in *Proceedings of the Sound & Music Computing Conference (SMC'09)*, Porto, Portugal.
- [39] 2009 Arnaud Dessein & Guillaume Lemaitre, "Free classification of vocal imitations of everyday sounds" in *Proceedings of the Sound & Music Computing Conference (SMC'09)*, Porto, Portugal.
- [40] 2008 Antoine Minard, Patrick Susini, Nicolas Misdariis, Guillaume Lemaitre, Stephen McAdams & étienne Parizet, "Environmental sound description : a meta-analysis of timbre perception", in *Proceedings of the 26th Computer Human Interaction Conference (CHI 2008)*, Florence, Italy.
- [41] 2007 Arnault Nagle, Nicolas Tsingos, Guillaume Lemaitre & Aurélien Sollaud, "On-the-fly auditory masking for scalable VOIP bridges", in *Proceedings of the AES 30th International Conference on Intelligent Audio Environments*, March 15-17, Saariselkä, Finland.
- [42] 2006 Patrick Susini, Nicolas Misdariis, Olivier Houix, Guillaume Lemaitre, Davide Rocchesso, Pietro Polotti, Karmen Franinović, Yon Visell, Klaus Obermayer, Hendrick Purwins & Kamil Adiloglu, "Closing the loop of sound evaluation and design", in *Proceedings of the 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems*, Berlin, Germany.
- [43] 2005 Patrick Susini, Stephen McAdams, Nicolas Misdariis, Guillaume Lemaitre & Suzanne Winsberg, "Timbre des sons environnementaux", in *Actes du Colloque International de Musicologie*, Montréal.
- [44] 2005 Emmanuel Gallo, Guillaume Lemaitre & Nicolas Tsingos, "Prioritizing Signals for Selective Real-time Audio Processing", in *Proceedings of the International Conference on Auditory Displays (ICAD 2005)*, Limerick, Ireland.
- [45] 2004 Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg & Stephen McAdams, "A method to assess the ecological validity of laboratory-recorded car horn sounds", in *Actes du Congrès joint CFA/DAGA'04*, Strasbourg, France.

- [46] 2003 Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg & Stephen McAdams, *Perceptively based design of new car horns sounds*, in *Proceedings of the International Conference on Auditory Displays (ICAD 2003)*, Boston, MA.
- [47] 2001 Guillaume Lemaitre & Christophe Vergez, "Physical modeling of the oboe : Influence of the bore conicity and of the pipe-neck after the reed", in *Proceedings of the International Conference on Systemics, Cybernetics & Informatics (SCI 2001)*, Orlando, FL

Abstracts

- [48] 2016 Guillaume Lemaitre, Olivier Houix, Frédéric Voisin, Nicolas Misdariis & Patrick Susini. "Comparing the identification of vocal imitations and computational sketches of everyday sounds". *Proceedings of the Meeting the ASA Fall Meeting*, Honolulu, HI, 2016.
- [49] 2016 Vincent Isnard & Guillaume Lemaitre. "Auditory bubbles reveal sparse spectro-temporal cues subserving identification of musical voices and instruments". *Proceedings of the Meeting the ASA Fall Meeting*, Honolulu, HI, 2016.
- [50] 2016 Patrick Susini, Olivier Houix, Lou Seropian & G. Lemaitre. "How loudness affects everyday sounds recognition". *Proceedings of the Meeting the ASA Fall Meeting*, Honolulu, HI, 2016.
- [51] 2016 Hugo Scurto, Guillaume Lemaitre, Jules François, Frédéric Bevilacqua, Patrick Susini, & Frédéric Voisin "Embodying sounds: Building and analysis of a database of gestural and vocal imitations" in *ISGS 7 : 7th Conference of the International Society for Gesture Studies*, 18-22 Jul 2016 Paris (France)
- [52] 2015 Hugo Scurto, Guillaume Lemaitre, Jules François, Frédéric Voisin, Frédéric Bevilacqua, & Patrick Susini. "Combining gestures and vocalizations to imitate sounds." *Proceedings of the ASA Fall meeting*, Jacksonville, FL. Also in the *Journal of the Acoustical Society of America* 138, no. 3 (2015): 1780-1780
- [53] 2015 Guillaume Lemaitre, Nicole Navolio, & Laurie M. Heller. "Priming gestures with associated sounds." *Proceedings of the ASA Spring meeting*, Pittsburgh, PA. Also in the *Journal of the Acoustical Society of America* 137, no. 4 (2015): 2205-2205.
- [54] 2015 Guillaume Lemaitre, Ali Jabbari, Olivier Houix, Nicoals Misdariis, & Patrick Susini, (2015). "Vocal imitations of basic auditory features". *Proceedings of the ASA Spring meeting*, Pittsburgh, PA. Also in the *Journal of the Acoustical Society of America*,

137(4), 2268-2268.

- [55] 2014 Guillaume Lemaitre “Auditory perception of the actions causing the sounds: a basis for the design of sonic interactions”, in *Conference on Sonification of Health and Environmental Data*, York, UK. **Invited presentation.**
- [56] 2012 Laurie M. Heller, Nicolas Zuniga-Peñaranda & Guillaume Lemaitre “Auditory-Motor Priming Using Sounds Produced by Simple Actions”, in *Proceedings of the 11th Annual Auditory Perception, Cognition & Action Meeting (APCAM 2012)*, Minneapolis, MN.
- [57] 2010 Guillaume Lemaitre & Laurie M. Heller, “Action is more accessible than material”, in *Proceedings of the 9th Annual Auditory Perception, Cognition & Action Meeting (APCAM 2010)*, Saint Louis, MO.
- [58] 2010 Guillaume Lemaitre & Laurie M. Heller, “Action verbs are the most accessible level of sound event description”, in *Proceedings of the 9th Annual Auditory Perception, Cognition & Action Meeting (APCAM 2010)*, Saint Louis, MO
- [59] 2010 Guillaume Lemaitre, Arnaud Dessein, Karine Aura & Patrick Susini, “Perception of vocal imitations and of the imitated sounds”, in *Proceedings of the ASA meeting*, Baltimore, MD. Also in *Journal of the Acoustical Society of America* vol.127(3), p. 1899.
- [60] 2010 Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis & Patrick Susini, “A perceptive categorization of the sounds produced by the interaction of solid objects”, in *Proceedings of the ASA meeting*, Baltimore, MD. Also in *Journal of the Acoustical Society of America* vol.127(3), p. 1955.
- [61] 2009 Guillaume Lemaitre, Arnaud Dessein, Karine Aura & Patrick Susini “Do vocal imitations enable the identification of the imitated sounds?” in *Proceedings of Auditory Perception, Cognition & Action Meeting (APCAM 2009)*, Boston, MA.
- [62] 2009 Guillaume Lemaitre, Olivier Houix, Nicolas Misdariis & Patrick Susini “Emotional reactions to the manipulation of a sonically augmented interactive device” in *Proceedings of Auditory Perception, Cognition and Action Meeting (APCAM 2009)*, Boston, MA.
- [63] 2009 Guillaume Lemaitre, Arnaud Dessein, Karine Aura & Patrick Susini, “Why and how do we vocally imitate sounds?” in *Proceedings of the SID workshop “Sketching Sonic Interaction Design”*, Holon, Israel. **Invited presentation.**
- [64] 2008 Karine Aura, Guillaume Lemaitre & Patrick Susini, “Vocal imitations of sound events enable recognition of the imitated sound events”, in *Proceedings of the EAA/ASA joint meeting Acoustics’08*, Paris, France. Also in *Journal of the Acoustical Society of America* vol.123 p. 3414.

- [65] 2008 Guillaume Lemaitre, Olivier Houix, Nicolas Misdariis & Patrick Susini “Naïve listeners and listeners with an expertise in sound or music do not use the same strategy to categorize everyday sounds” in *Proceedings of the EAA/ASA joint meeting Acoustics’08*, Paris, France. Also in *Journal of the Acoustical Society of America* p. 3689. **Invited presentation.**
- [66] 2008 Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis & Patrick Susini “Classification of everyday sounds : influence of the degree of sound source identification” in *Proceedings of the EAA/ASA joint meeting Acoustics’08*, Paris, France. Also in *Journal of the Acoustical Society of America* p. 3414. **Invited presentation.**
- [67] 2002 Guillaume Lemaitre, Patrick Susini, Suzanne Winsberg & Stephen McAdams, “Perception of the timbre of car horns”, in *Proceedings of the Forum Acusticum, EEA meeting*, Sevilla, Spain. Also in *Acta Acustica united with Acustica 88, S123 (A)*.

National conferences (6)

- [68] 2016 Guillaume Lemaitre, Olivier Houix, Nicolas Misdariis & Patrick Susini, “Des imitations vocales à l’identification de sons de l’environnement ”, in *Actes du Congrès Français d’Acoustique*, Le Mans, France.
- [69] 2014 Antoine Minard, Guillaume Lemaitre, Christophe Lambourg & Patrick Boussard, “Synthèse de bruits de passage d’avion dans le cadre du projet PARASOFT”, in *Actes du Congrès Français d’Acoustique*, Poitiers, France.
- [70] 2014 Guillaume Lemaitre, Jean-François Sciabica, Stephan Moal, Laurence Vion, Mickael Zeckri & Sylvain Hourcade, “Développement d’un dispositif d’évaluation continue du désagrément acoustique généré par un survol d’avion”, in *Actes du Congrès Français d’Acoustique*, Poitiers, France.
- [71] 2014 Guillaume Lemaitre, Christian Vartanian, Christophe Lambourg & Patrick Boussard, “Etude psychoacoustique des fluctuations et turbulences dans le bruit aérodynamique automobile”, in *Actes du Congrès Français d’Acoustique*, Poitiers, France.
- [72] 2010 Guillaume Lemaitre, Arnaud Dessein, Karine Aura & Patrick Susini, “Do vocal imitations enable the recognition of the imitated sounds?”, in *Actes du 10^{ème} Congrès Français d’Acoustique (CFA)*, Lyon, France.
- [73] 2010 Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, Patrick Susini & Isabel Urdapilleta, “Catégorisation des sons produits par des objets solides”, in *Actes du 10^{ème} Congrès Français d’Acoustique (CFA)*, Lyon, France.