Timbre from Sound Synthesis and High-level Control Perspectives

Sølvi Ystad, Mitsuko Aramaki, Richard Kronland-Martinet

Timbre from Sound Synthesis and High-level Control Perspectives

Sølvi Ystad, Mitsuko Aramaki, and Richard Kronland-Martinet

Aix Marseille Univ, CNRS, PRISM (Perception, Representations, Image, Sound, Music),

31 Chemin J. Aiguier, 13402 Marseille Cedex 20, France

ystad@prism.cnrs.fr (corr.author)

aramaki@prism.cnrs.fr

kronland@prism.cnrs.fr

running title: Timbre, Sound Synthesis and Control

Abstract

Exploring the many surprising facets of timbre through sound manipulations has been a common practice among composers and instrument makers of all times. The digital era radically changed the approach to sounds thanks to the unlimited possibilities offered by computers that made it possible to investigate sounds without physical constraints. In this chapter, we describe investigations on timbre based on the analysis-by-synthesis approach that consists of using digital synthesis algorithms to reproduce sounds and further modify the parameters of the algorithms to investigate their perceptual relevance. In the first part of the chapter, timbre is investigated in a musical context. An examination of the sound quality of different wood species for xylophone making is first presented. Then the influence of physical control on instrumental timbre is described in the case of clarinet and cello performances. In the second part of the chapter, environmental sounds have been investigated in order to identify so-called invariant sound structures that can be considered as the backbone, or the bare minimum, of the information contained in a sound that enables the listener to recognize its source both in terms of structure (e.g. size, material) and action (e.g. hitting, scraping). Such invariants are generally composed of combinations of audio descriptors such as decay, attack, spectral density and pitch. Various investigations on perceived sound properties responsible for the evocations of sound sources are here identified and described through both applied and fundamental studies.

Keywords**:** analysis by synthesis, timbre, intuitive synthesis control, sound semiotics, digital interfaces, musical interpretation, action/object paradigm

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

1 Introduction

1.1 Historical Overview of Timbre and Instrumental Control

Timbre has been one of the main concerns for instrument makers, musicians, and composers throughout history. Certain instruments, such as organs, were particularly well adapted to exploring various timbres due to the numerous pipes that made it possible to combine different harmonics. In the 18th century, Dom Bedos de Celles, a French Benedictine monk and organ maker, published a treatise entitled "The art of the organ-builder" (*L'art du facteur d'orgues*) in which he not only describes the principles behind organ building, but also the many ways of imitating certain instrumental timbres by adding or removing sounds of pipes tuned to multiples of the fundamental frequency (Dom Bedos 1766) (For a detailed description of organ structure and timbre see Angster et al. (2017)). Such techniques, which constitute a practical application of the Fourier (1878) theorem on periodic functions, are claimed to have been used as early as in the XVth century, that is, four centuries before Fourier published his fundamental theorem showing that sounds can be reconstituted by a sum of harmonics and before Helmholtz (1868) related timbre to the proportion of harmonics.

When electricity became viable for use in technology thanks to Faraday in 1831, inventors started to build new musical instruments often based on the additive synthesis technique, which consists in creating sounds by adding elementary signals, typically sine functions with different frequencies and amplitudes. Phonic wheels, thanks to which the harmonics of the sounds could be added and removed to imitate timbres of both known and unknown instruments, were used to develop the Telharmonium in 1897. The Hammond organ developed in the 1930s was based on the same principle but with new control features. With the B-3 model, which offered control of the attack time (Peeters, Chap. 11), the instrument all of a sudden became extremely attractive to jazz musicians due to its new means of adjusting the degree of percussiveness of the sounds (De Wilde 2016). Another invention that focused on

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

the possibilities of controlling timbre variations was the "ondes Martenot," which was based on high-frequency (radio) waves (similarly to the more widespread Theremin). This instrument was equipped with a six-octave keyboard, a sliding metal ring that enabled the performer to produce glissandi, as well as a drawer with timbre controls that made it possible to switch between different waveforms (e.g., sinusoidal, triangle and square waves, pulse waves, and noises) and to route the instrument's output to various loudspeakers providing either reverb effects, sympathetic resonances, or 'halo' effects.

Yet another instrument that offered a huge palette of new timbres was the modular Moog synthesizer developed in the 1960s, which enabled the creation of sounds using four basic modules, namely oscillators, amplifiers, filters, and envelopes. By offering fine envelope control of the attack, release, sustain, and decay parts of the sound, an extremely rich and subtle timbre control was made possible for the musician. Unfortunately, the many control possibilities were not as intuitive and made the first versions of the instrument difficult to use. These examples illustrate musicians' and composers' passionate quest for new timbres, which was nicely expressed by the composer Edgar Varèse (1917, p.1): "I dream of instruments obedient to my thought and which with their contribution to a whole new world of unsuspected sounds, will lend themselves to the exigencies of my inner rhythm." (translated from French by Louise Varese).

1.2 Timbre Studies Induced by the Digital Era

In spite of the many amazing instruments dedicated to analog synthesis (obtained from electric pulses of varying amplitude), the arrival of the digital era in which the computer was introduced revolutionized our conception of musical sounds and perception. In 1957, Max Mathews developed the first sound synthesis computer program (MUSIC I) at the Bell Labs in the USA, which he used to create the first computer generated musical piece in history

(Mathews 1963). The use of sound synthesis enables one generating an infinity of sounds without being constrained by physics. Several pioneers in the field, such as Jean-Claude Risset, David Wessel, and John Chowning, who were both composers and scientists, rapidly took the opportunity to use this new tool as a means to establish links between perception and sound structures by developing an analysis-by-synthesis approach in which the reconstruction of the sound became the criterion for the relevance of the analysis. It was by such an approach that Risset revealed the importance of the temporal evolution of different spectral components in trumpet sounds, for example (Risset 1965). This study pointed out that the increase in spectral bandwidth as a function of amplitude is linked to the brassy effect of the instrument. Similarly, Mathews and colleagues (1965) managed to improve the realism of the attack of bowed string instruments by introducing frequency variations to synthetic sounds.

The analysis-by-synthesis approach was also used in the first studies on perceptual representation of timbre proposed by Grey (1977). This study involved constructing synthetic emulations of musical instruments, in which certain parts of the signal were degraded through simple transformations. Through listening tests, certain acoustic parameters, such as the attack time and the spectral centroid, were identified as relevant from a perceptual point of view. More recent studies based on either resynthesized sounds obtained from the analysis of recorded sounds or from synthesized sounds that are not necessarily perfect imitations of the original sound, have revealed several audio descriptors that are representative of specific sound categories (Peeters et al. 2011, McAdams, Chap. 2; Saitis and Weinzierl, Chap. 5).

At this stage one might think that Varese's dream of instruments that give access to any timbre that a composer imagines would be available due to the many studies that have established links between sound categories and audio descriptors. This is true in theory, but not that easy in practice, since our ability to describe and control sound structures to obtain given timbres is limited. In fact, digital sound synthesis is based on low-level parameters such as amplitudes

and frequencies of spectral components, and their temporal evolution. This signal content is a consequence of the physical behavior of the source and does not necessarily reflect how the sound is perceived (McAdams, Chap. 2). A major challenge in the domain of digital synthesis is therefore to unveil the sound structures that are responsible for the recognition of the sound source (size, shape, …) and the sound producing action (hitting, scraping, …) in order to be able to reproduce and control such evocations in an intuitive manner. This means that various scientific disciplines must be associated in order to link perceptual and cognitive experiments with physical characteristics of sounds.

1.3 Source Recognition and Timbre

Several composers, psychologists, musicians and scientists have worked on human perception of environmental sounds. During the late 1940s, the French scientist, philosopher, and musician Pierre Schaeffer introduced a new musical genre that he called "*musique concrète"* in the "Studio d'essai" of the French public radio (RTF). This new trend consisted in distributing recorded sounds for which the source could not be easily recognized on loudspeakers in order to favor so-called reduced or acousmatic listening, hereby forcing the listeners to focus on the sound itself and not on the source that created the sound. Schaeffer realized that the specification of the physical structure of the sound was not adequate to control the auditory effects, because "music is meant to be heard" and the relation between the physical signal and the perception of musical sounds at the time was from his viewpoint grossly insufficient. In his "Traité des Objets Musicaux" (Schaeffer 1966), he introduced the notion of "reduced listening," which consists of focusing on the morphological aspects of sounds rather than its source and production. By creating abstract or acousmatic sounds (i.e., sounds whose source cannot be easily identified), this kind of listening could be favored.

Schaeffer' ideas can be found in later studies, for instance by Smalley (1994) who introduced the term source bonding as "the natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins" and Gaver (1993) who distinguished what he called ecological or everyday listening (hearing events *per se*) from analytical or musical listening (focusing on intrinsic sound properties as in the case of Schaeffer's reduced listening). Gaver also took his inspiration from Gibson (1979) who introduced the ecological approach to perception in the visual domain. This theory supposes that our perception is direct, without any influence of inference or memory, and is based on the recognition of specific signal morphologies, which can be considered as invariant structures that transmit the perceptual information. In addition to Gaver, several authors have adapted the ecological approach to the auditory domain (Warren and Verbrugge 1984; McAdams 1993). The notion of invariant sound structures is particularly interesting for sound synthesis and control purposes, since the identification of such structures both makes it possible to focus on evocative sound structures to produce sounds and sound metaphors, and enables intuitive or high-level control of sounds from semantic descriptors (e.g., small, big, metal, wood, hollow, plain…).

Accordingly, this chapter will attempt to discuss timbre from the point of view of sound synthesis and control. In the first part, timbre and musical sounds are investigated in various cases. First, wood species are evaluated by a xylophone maker in terms of sound quality. Then the link between instrumental control and timbre is investigated in the case of clarinet and cello performances. In the second part environmental sounds are firstly investigated through studies based on brain imaging techniques aiming at investigating the semiotics of sounds, that is, how meaning is attributed to sounds. Then the implication of audio descriptors in the identification of invariant sound structures responsible for the evocation of sound sources and events is examined. Finally, particular mapping strategies between low-level synthesis parameters, audio

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

descriptors and semantic labels describing the sound sources for intuitive high-level control of sounds are described.

2 Timbre Studies Based on Analysis –Synthesis Approaches in Musical Contexts

This section deals with timbre-related questions of musical instruments, in particular the quality of musical sounds and the role of timbre in instrumental control and musical performances.

2.1 Timbre-Based Wood Selection Made by a Xylophone Maker

The mechanical properties of wood species strongly influence the sound quality. When choosing their wood species, xylophone makers carefully listen to the sounds they produce. Little is known about the criteria they use and the relationship between the sound quality and the physical parameters characterizing wood species. The aim of this study was to identify these perceptual criteria (Aramaki et al. 2007). For this purpose, a professional xylophone maker was asked to evaluate different tropical and subtropical wood species with the same geometry. Sounds were first recorded and classified by the instrument maker through a free classification test. Then the sounds were resynthesized and tuned to the same pitch before the same instrument maker performed a new classification. Statistical analyses of both classifications revealed the influence of pitch on the xylophone maker's judgment and pointed out the importance of two audio descriptors: the frequency-dependent damping and the spectral bandwidth, indicating that the instrument maker searched for highly resonant and crystal-clear sounds. These descriptors can be further related to physical and anatomical characteristics of wood species, thereby providing recommendations for choosing attractive wood species for percussive instruments. Previous studies relating auditory cues to geometry and material properties of vibrating objects have pointed out the importance of internal friction related to the damping factors of the spectral components (Avanzini and Rocchesso 2001; Giordano and McAdams 2006; Lutfi and Oh 1997;

Klatzky *et al.* 2000; McAdams *et al.* 2004, 2010) as theoretically shown by Wildes and Richards (1988). Other studies on sound quality of musical instruments have been performed for instance in the case of violin sounds (Saitis et al. 2012, 2017) based on psycholinguistic analyses of verbal descriptions from experienced musicians. These studies, based on the musician's spontaneous verbalizations describing the playing experience, led to a model linking auditory and haptic sensations to timbre, quality and playability of the instrument.

2.2 Timbre Control in Clarinet Performances

Investigations on musical timbre do not solely focus on the mechanical properties of the instrument itself, but also on the way a musician can control timbre during the instrumental play. In the following section a study on the influence of a clarinet player's pressure and aperture on the resulting timbre using a physical synthesis model is described (Barthet et al. 2010a), followed by an investigation of the influence of timbre on expressiveness in clarinet performance (Barthet et al. 2010b).

2.2.1 Timbre and Instrumental Control

To draw a link between the control parameters and the resulting timbre in clarinet performance, a synthesis model was used to generate perfectly calibrated sounds (Guillemain et al. 2005). Fifteen sounds obtained by different values of reed aperture and blowing pressure were evaluated through dissimilarity ratings. The statistical analyses of the perceptual evaluations resulted in a timbre space with dimensions that correlated well with attack time, spectral centroid, the energy ratio between odd and even harmonics, and the energy of the 2nd to 4th harmonics (2nd tristimulus coefficient). A correlation between the control parameters and the timbre space could also be found, revealing that the pressure control correlated well with the third dimension and had a strong influence on the odd/even ratio. Furthermore, the reed

aperture was well correlated with the first dimension and was correlated with the attack time and the spectral centroid (see Fig. 1). These results allowed for the prediction of the instrumental timbre from the values of the control parameters. Hence, small values of reed opening and blowing pressure result in long attack times and low spectral centroid values, while increasing reed apertures induce increases in the odd/even ratio. (For more information on the acoustics of wind instruments, see for instance Fletcher and Rossing (1998), Joe Wolfe (2018) and Thomas R. Moore (2016)).

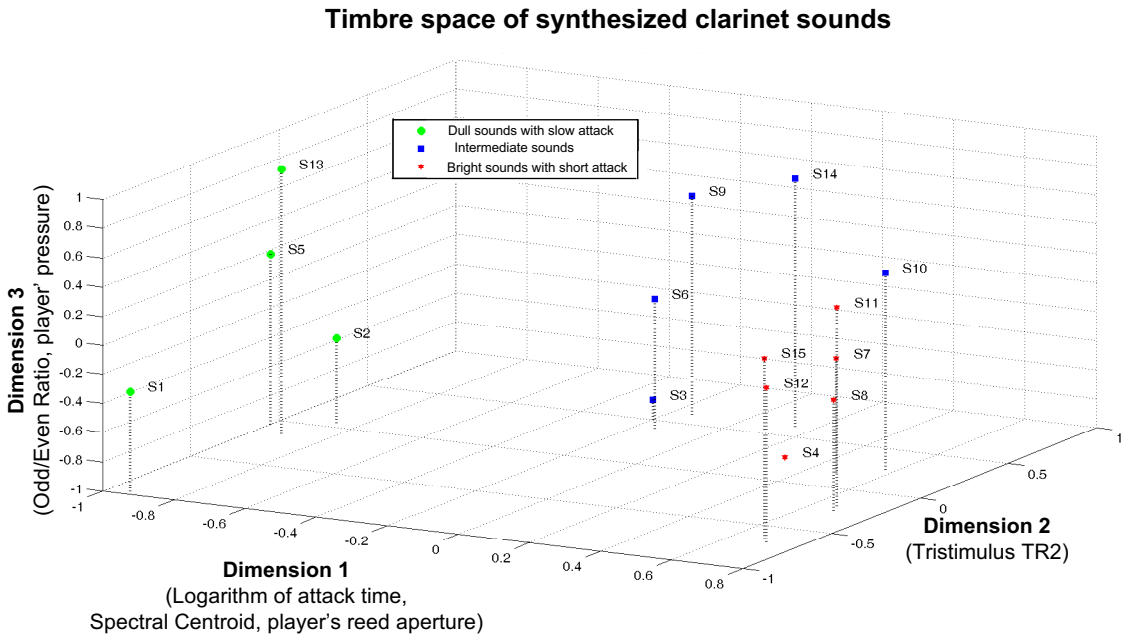

Fig. 1 *Timbre space of clarinet sounds for different reed apertures and blowing pressures. The reed aperture is correlated with Dimension 1, whereas the pressure is correlated with Dimension 3.*

Studies on musical performance have revealed rhythmic and intensity deviations with respect to the musical score, leading to proposals of various musical rules (Sundberg 2000). Although timbre variations are likely to be used by musicians as a means to add expressivity to the performance, they have been more or less ignored, probably for two reasons: 1) scores do not contain timbre specifications, and 2) timbre variations strongly depend on the specificities of

each instrument and might therefore be difficult to integrate with general performance rules. In this study timbre variations were analyzed in order to investigate the influence of timbre on expressiveness in clarinet performance. Mechanical and expressive clarinet performances of excerpts from Bach and Mozart were recorded. An objective performance analysis was then conducted, focusing on the acoustical correlates of timbre. A strong interaction between the expressive intentions and the audio descriptors (attack time, spectral centroid, odd/even ratio) was observed in the case of both musical excerpts. The timbre-related changes across expressive levels did not occur at every note, but were specific to some notes or groups of notes in the musical phrases (such as the first note in a phrase or specific passages). The most salient changes were in the mean spectral centroid and odd/even ratio values and in the range of variation in the durations of the tones. These changes seemed to be made more frequently in the case of long notes (such as half and quarter notes), possibly because a performer needs a certain time to control the timbre while playing.

In a companion study (Barthet et al. 2011), the perceptual influence of certain acoustical timbre correlates (spectral centroid, SC), timing (Intertone onset interval, IOI) and intensity (root mean square envelope) on listeners' preferences between various renderings was studied. An analysis-by-synthesis approach was used to transform previously recorded clarinet performances by reducing the expressive deviations from the SC, the IOI, and the dynamics (root mean square envelope). Twenty skilled musicians were asked to select which version (recorded vs. transformed) they preferred in a paired-comparison task. Results showed that the removal of the SC variations most significantly decreased the musical preference of the performances. This finding seems to be due to the fact that this transformation altered the original timbre of the clarinet tones (the identity of the instrument), and drastically affected the time-evolving spectral shapes, causing the tones to be static and unlively. This result suggests that acoustical morphology, which strongly depends on the context (i.e., the previous and

following notes, a fast or slow or musical tempo, etc.), is important to convey expressiveness in music. More recent studies have analyzed the combination of timbre from different musical instruments and how timbres of certain instruments can blend together and whether musicians consciously control such blend during performance (Lembke et al. 2017). Results revealed that musicians adjusted their timbre, in particular the frequencies of the main formant or spectral centroid depending on whether they had a role as leader or follower during the performance. Other studies have explored the more general role of timbre in orchestration and musical tension and proposed a typology of orchestral gestures based on large-scale timbral and textural changes (Goodchild et al. 2017, McAdams, Chap. 8).

2.3 Timbre and the Notion of Harshness in Cello Performances

Another aspect that appears to be important for musical expressiveness is the musician's movements during the performance. Are the sound-producing gestures solely responsible for the sound quality and expressiveness, or do ancillary gestures that are not directly involved in the sound production also play a role? Several studies on ancillary gestures have been performed in the case of the clarinet (Wanderley et al. 2005; Desmet et al. 2012), the piano (Thompson 2012, Jensenius 2007), the harp (Chadefaux et al. 2013), and the violin (Van Zijl and Luck 2013). In the case of clarinet performance, it was shown that the body movements of the musician generate amplitude modulations of partials of the sounds, often perceived as beating effects. Such modulations are essentially due to changes in directivity of the instrument that follows the ancillary gestures of the musician. In the case of piano performance, a circular movement of the elbow enables a larger displacement of the hand (Jensenius 2007). This gesture depends on parts of the body that are not directly implied in the instrumental gesture (Thompson and Luck 2012). In a study on ancillary gestures in cello performances (Rozé et al.

2016; 2017), professional cellists were asked to play a score as expressively as possible in four postural conditions. The four conditions were a normal condition (N), a "mentally constrained" condition in which the cellists were asked to move as little as possible ("Static Mental"), and two physically constrained conditions in which the torso was attached to the back of the chair with a race harness ("Static Chest") and, for the most constrained condition (Fig.2), also the head (in addition to the torso) was immobilized by a collar neck ("Static Chest Head").

Fig. 2 *Constrained postural condition in which the cellist is attached to the chair by a race harness and his head is immobilized by a collar neck. The reflective markers on the picture enable body movement recordings by motion capture cameras.*

A musical score divided in six parts based on cello exercises with specific technical difficulties was constructed. The tempo (45 beats per minute) was given by a metronome before the beginning of each session and two bowing modes (detached and legato) were compared. Sounds and body movements were recorded. The analyses of the performances revealed that for certain notes the timbre was modified in the fully constrained condition in addition to certain rhythmic deviations. In particular, in a specific passage of the score, a degradation of the timbre inducing

perceived harshness could be perceived. An analysis-by-synthesis approach associated with listening tests revealed that this phenomenon could be characterized by an energy transfer or a formant shift towards higher-order harmonics, a decrease in attack time, and an increase in fluctuation of harmonic amplitudes. Based on these results, a predictive model of perceived harshness depending on three audio descriptors, that is, the attack time, the ratio between the first and second mel-frequency cepstral coefficients characterizing slow fluctuations of the spectral envelope and the harmonic spectral variation reflecting the evolution of the energy of the harmonic components over time was proposed. The three-dimensional space resulting from this analysis presented tight analogies with the acoustic correlates of classical timbre spaces (Grey 1997; McAdams et al. 1995, Hajda et al. 1997). Its first dimension indicated that participants were particularly sensitive to spectral fluctuation properties (Harmonic spectral variation), while the second and third dimensions were respectively well explained by spectral attributes (Harmonic spectral centroid, MFCC ratio) and a temporal attribute (Attack slope). This indicates that a greater brightness combined with a softer attack would contribute to increase the perceived harshness of a cello sound.

3 Semiotics of Environmental Sounds

In timbre research, musical sounds have been given a lot of attention since the first multidimensional representations proposed by Grey (McAdams, Chap. 2; Saitis and Weinzierl, Chap. 5). Fewer studies are available on environmental sounds, possibly because such sounds are hard to control and often complex to model both from physical and signal points of view. Although environmental sounds have been an important source of inspiration for composers of all times, we will not focus on the musical context in this section, but rather present a pragmatic approach that focuses on the way we perceive and attribute meaning to environmental sounds.

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

In daily life, people are confronted with environmental sounds that are more or less consciously processed. Sounds tell us about the weather, living creatures in our surroundings, potential dangers, etc., which means that our environment constantly communicates infomation to us. How do people interpret and attribute meaning to such sounds? Can these sounds provide new ways to communicate if we manage to extract their perceptual essence and further implement it in sound synthesis processes? Can a common sense be attributed to such environmental languages that can be compared to the semantics of spoken languages?

As a first attempt to answer these questions, an investigation on the perception of isolated sounds would be interesting. One of the major issues that arises from the cognitive neuroscience point of view is whether similar neural networks are involved in the allocation of meaning in the case of language and that of sounds of other kinds. In a seminal study, Kutas and Hillyard (1980) showed that sentences that ended by words that were out of context (e.g., The fish is swimming in the river/carpet) elicited a larger negative amplitude in the evoked-response potential (ERP) measured on the scalp of the subjects 400ms after the onset of the incongruous word (N400) than when the last word was congruent. The N400 has been widely used since that time to study semantic processing in language. Authors of recent studies used a priming procedure with nonlinguistic stimuli such as pictures, odors, music, and environmental sounds (see Aramaki et al. 2009; Schön et al. 2009 for reviews). In this section, two priming experiments that use nonlinguistic stimuli to observe the negativity of ERP components for related versus unrelated stimuli are presented. In the first case priming effects induced by pairs of abstract sounds (favoring reduced listening) and written words were investigated, and pairs of impact sounds evoking different material categories were examined in the second case.

3.1 Priming with Abstract Sounds

Although the results of previous priming experiments have mostly been interpreted as

reflecting some kind of conceptual priming between words and nonlinguistic stimuli, they may also reflect linguistically mediated effects. For instance, watching a picture of a bird or listening to birdsong might both automatically activate the verbal label "bird." The conceptual priming cannot therefore be taken to be purely nonlinguistic because of the implicit naming induced by the processing of the stimulus.

Certain studies have attempted to reduce as far as possible the likelihood that a labeling process of this kind takes place. To this end, "abstract" sounds, which have the advantage of not being easily associated with an identifiable physical source, are useful (Schaeffer 1966; Merer et al. 2011, 2013). Sounds of this kind include environmental sounds that cannot be easily identified by listeners or can give rise to many different interpretations, depending on the context. They also include synthesized sounds and laboratory-generated sounds in general if their origin is not clearly detectable. Note that alarm or warning sounds do not qualify as abstract sounds. In practice, making recordings with a microphone close to the sound source, using musical instruments in untraditional ways, or using everyday objects such as tools or toys are common ways of creating abstract sounds. Sound synthesis methods such as granular synthesis, which consists of adding lots of very short (typically 1 to 50ms) sonic grains to form larger acoustic events (Curtis Roads 1988), are also efficient means of creating abstract sounds.

In the present study conceptual priming tests using word/sound pairs, for which the level of congruence between the prime and the target was varied were conducted. In the first experiment, a written word (the prime) was presented visually before an abstract sound (the target), and participants had to decide whether or not the sound and the word matched. In the second experiment, the order of presentation was reversed. Results showed that participants were able to assess the relationship between the prime and the target in both sound/word and word/sound presentations, showing low inter-subject variability and good consistency. The contextualization of the abstract sound facilitated by the presentation of a word reduced the

variability of the interpretations and led to a consensus between participants in spite of the fact that the sound sources were not easily recognizable. Electrophysiological data showed the occurrence of an enhanced negativity in the 250–600 ms latency range in response to unrelated as compared to related targets in both experiments suggesting that similar neural networks are involved in the allocation of meaning in the case of language and sounds. In addition differences in scalp topography were observed between word and sound targets (from fronto-central to centro-parietal distributions), which can be taken to argue that the N400 effect encompasses different processes and may be influenced by both the high-level cognitive processing of the conceptual relation between two stimuli and lower-level perceptual processes linked with the specific acoustic features of the sounds such as attack time, spectral centroid, spectral variation, and others (Schön et al. 2009). This means that a combination of sound features, so-called invariants (cf. Sec. 4) might be used by the listeners to determine specific aspects of sounds.

3.2 Priming with Material Categories

Pursuing this topic farther in a subsequent study, Aramaki et al. (2009) sought to completely avoid the use of words as primes or targets. Conceptual priming was therefore studied using a homogeneous class of nonlinguistic sounds, such as impact sounds, as both primes and targets. The degree of congruence between the prime and the target was varied in the following three experimental conditions: related, ambiguous and unrelated. The priming effects induced in these conditions were then compared with those observed with linguistic sounds in the same group of participants. Results showed that the error rate was highest with ambiguous targets, which also elicited larger N400-like components than related targets in the case of both linguistic and nonlinguistic sounds. The finding that N400-like components were also activated in a sound-sound design showed that linguistic stimuli were not necessary for this component to be elicited. This component may therefore reflect a search for meaning that

is not restricted to linguistic meaning. This study showed the existence of similar relationships in the congruity processing of both nonlinguistic and linguistic target sounds, thereby confirming that sounds can be considered as an interesting way to convey meaningful messages.

4 Towards Intuitive Controls of Sounds

The identification of perceptually relevant signal morphologies is of great interest in the domain of sound synthesis, since it opens a new world of control possibilities. In the 1960s when computer music was at its very beginning, the famous scientist John Pierce made the following enthusiastic statement about sounds made from computers: "wonderful things would come out of that box if only we knew how to *evoke* [emphasis added] them" (Pierce 1965, p. 150). In spite of the many synthesis algorithms that have been developed so far and that provide perfect resynthesis of sounds, meaning that no difference is perceived between the original and the synthesized sound (Kronland-Martinet et al. 1997; Cook and Scavone 1999; Bensa et al. 2003, Bilbao and Webb 2013), the issue of control is still a great challenge that prevents many potential users from considering sound synthesis in their applications. This issue has always interested composers and musicians, and a large number of interfaces and control strategies for digital sound synthesis have already been proposed in the musical domain (Moog 1987; Cook 2001; Gobin et al 2003, Miranda and Wanderley 2006). The first works on perceptual control were presented by David Wessel (1979) who proposed a new way to navigate within a perceptual sound space based on Grey's (1977) timbre space, defined by Krimphoff et al. (1994) as *"the mental organization of sound events at equal pitch, loudness and duration. The geometric distance between two timbres corresponds to their degree of perceived dissimilarity."* By manipulating an additive synthesis algorithm according to the audio descriptors, in particular the evolving spectral energy distribution and various temporal features (either the attack rate or the extent of synchronicity among the various components) defining the perceptual space, a perceptual control could be obtained. Such a control was more intuitive

than the control of basic signal parameters that define synthesis algorithms such as frequencies and amplitudes of spectral components.

However, a control device that focusses on signal properties of the sounds rather than the evoked sound sources, remains difficult to use. Deeper analyses that allow one to identify more complex sound structures that are responsible for evocations of sources and events are therefore necessary to propose more intuitive synthesis controls, for instance from verbal labels describing the perceived event. Such an approach necessitates a confrontation between distinct scientific domains such as experimental psychology, cognitive neuroscience, acoustics, physics and mathematics.

This section presents several studies that aim to identify various sound morphologies responsible for the evocation of sound sources, actions, and movements in order to develop new synthesis devices that enable more intuitive control of sounds. In the first part, evocations of quality, solidity, and sportiness induced by car doors and motors are described. A new sound-synthesis paradigm, called the "action/object" paradigm is then presented that considers any sound as the result of an action on an object, and that is based on a semantic description of the sound. Examples of investigations leading to the identification of perceptually relevant sound morphologies and the development of intuitive sound controls are given in this part.

4.1 Evidence of Actions on Objects in Everyday Timbres

In this subsection, two studies on sound quality are presented. In the first case, evocations of solidity and quality of cars induced by car-door noises are discussed, while the second case describes the link between evocations of sportiness and timbre of motor noise.

4.1.1 Door-Closure Sounds

This study was initiated by a car company that had noticed that the brief sound produced when slamming the car door was responsible for customers' mental image of the quality and

the solidity of the car and that this sound even was important for car sales! It was quite surprising that such a short sound that lasted for less than 250 ms could have any influence on the customers. To understand how the sound influenced the customers, signal morphologies responsible for the evocation of solidity and quality of the car had to be found in order to propose a predictive model of this relationship (Bezat 2007). For this purpose, door-closure sounds obtained from recordings of car doors (different brands and car categories) were analyzed and evaluated. The perceptual judgments of car doors were obtained from different types of listening tests. Following Gaver's (1993) definition, both analytical and ecological listening were considered.

In the case of analytical (or musical) listening, the sound signal is described without reference to the event in order to reveal perceptually relevant signal morphologies useful for signal analysis. To incite participants to characterize sounds in this rather unnatural way, a sensory analysis method was used (Roussarie et al. 2004). For this purpose, naïve subjects were trained during several sessions to define a minimal number of descriptors to qualify the sound they heard. Then they evaluated the stimuli. Through this method, a sensory profile for each sound could be obtained. The success condition of a sensory panel is sound discriminability, the judges' repeatability in time and their consensus. This is the reason why such a long procedure is necessary to transform naive subjects into efficient judges in order to obtain their consensus on all the descriptors. This approach revealed that the door-closure sound is described mainly by the intensity and by the onomatopoeia, BONM (pronounced [bɔ̃m]) and KE (pronounced [kø]). By comparing the analytical properties with expert listening, the BONM ([bɔ̃m]) descriptor could be related to the low-frequency closure sound and the KE ([kø]) descriptor to the high-frequency contribution that characterizes the lock component in the car door signal.

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

In the case of ecological (or everyday) listening, the event associated with the sound characterized by a set of natural properties as well as the evoked associations was described. Listening tests with both naive and expert listeners were performed, revealing that naive listeners were able to discriminate the doors by their quality, solidity, energy of closure, door weight, and door-closure effectiveness in a coherent manner. This is in line with previous studies (Kuwano et al. 2006) that resulted in coherent quality evaluations of car-door sounds across participants from semantic differential tests based on a predefined adjective scale. The expert listeners identified the elements of the car door that contribute to the sound such as the joints, the lock mechanism and the door panel in contrast to the more macroscopic evaluations of the naive listeners.

These different listening tests allowed the establishment of a network of perceptual properties of door-closure sounds (Fig. 3) that illustrates the links between the sensory panel's description of the sounds, (i.e. its analytical properties) and the notion of weight and closure linked to the natural properties and finally the evocations of quality and solidity of the car. It was thereby shown that the impression of solidity and quality was linked to the sensation of a heavy door and a gentle gesture, which in turn was characterized by the sensory panel as sounds without flaws (e.g., vibrations), low- pitched with little lock presence (strong BONM and weak KE), and of low intensity. In line with these results, the influence of the weight of the car door on the perceived quality was also confirmed by Scholl and Amman (1999) in a study in which car-door noises were evaluated after physical modifications of the car-door sources.

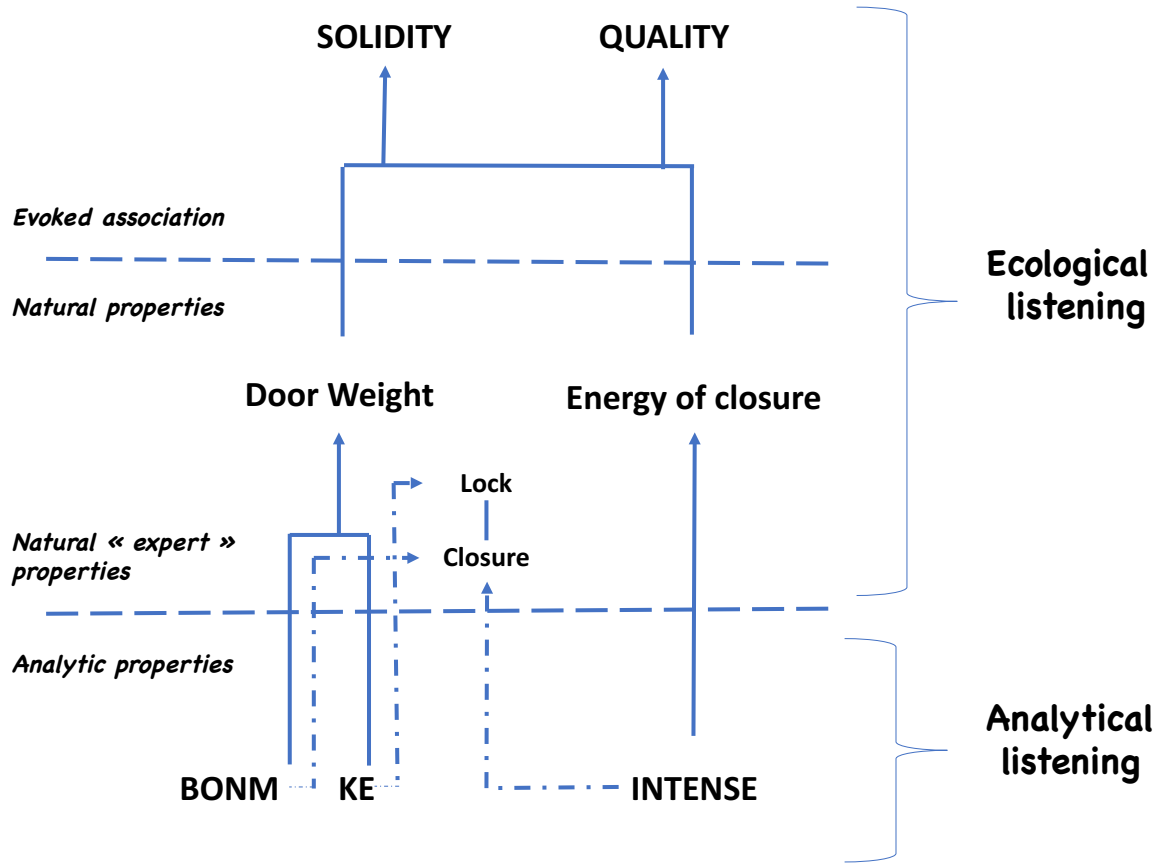

Fig. 3 *Relation between the sensory descriptors (BONM, KE, INTENSE), the natural properties (weight and energy closure) and the evocations of quality and solidity of the car.*

An analysis–synthesis approach based on empirical mode decomposition (EMD) (Huang et al. 1998) was then applied to separate the perceived source contributions (lock and closure) contained in the sound. This method consists in identifying iteratively intrinsic mode functions (IMFs) of the signal, both amplitude and frequency modulated, separating locally (on an oscillatory scale) a "fast" component from a slower pattern. This is done by pointing out the located maxima and minima of the signal, then constructing the superior and inferior envelopes, and finally the mean envelope. The first mode is thus obtained. The algorithm is then processed on the rest of the signal, until the second mode is obtained. By finding EMD modes, the rest of the signal (the residue) has less and less extrema. The decomposition process stops when the

last residue has only three extrema. The signal can then be perfectly reproduced by simple addition of the modes.

This signal analysis combined with the perceptual analyses revealed that an acceptable car-door noise should contain three impacts that evoke the latch mechanism (characterized as KE by the sensory panel) and one low-frequency impact that characterizes the door impact (characterized as BOMN by the sensory panel).

An additive synthesis model based on exponentially damped sinusoids was then used to synthesize the sounds. By adjusting the amplitudes, the damping coefficients, and the time between the different impacts, car-door-closure sounds corresponding to different vehicle qualities could then be generated. Further listening tests were then run, using the synthesized stimuli in order to relate the signal parameters to the perceived quality. Results showed that the energy and the damping of the door impact linked to the evocation of the weight of the door, and the elapsed time between the four impacts related to the evocation of a well-closed door was found to mediate the perception of solidity and quality of the car (Bezat et al. 2014). This study confirms that signal invariants evoking the solidity and quality of a car can be identified.

4.1.2 Motor Sounds

Another challenging study that aimed at relating evocations and sound structures was proposed by the same car company and concerned perceived motor noise quality during acceleration (Sciabica 2011). The aim was here to characterize the dynamic behavior of motor noise timbre in terms of sportiness and further propose a perceptual control of a synthesis model related to the degree of sportiness.

Sounds perceived in car passenger compartments are the result of three acoustic sources: the engine sound, the tire-road source and the aerodynamic source. The sound from tire-road source is due to the interaction between the tires and the road and depends on three main

parameters: car speed, tire texture and road texture. The contact between tire and road generates low frequency noise. The sound from the aerodynamic source is a broadband noise whose global sound level increases with speed. It mainly has a low frequency energy distribution (below 400 Hz), but its perceptual contribution is also important in the high frequency domain. Indeed, aerodynamic noise mainly masks high engine orders, but its impact can also be observed at low engine orders. The engine sound is a complex sound rich in overtones. Its fundamental frequency varies with the engine rotation speed, and the level of each harmonic depends on the multiple resonances inside the car. When the engine sound is sufficiently audible in the car, it can be described by perceptual attributes such as booming, brightness and roughness. Booming is associated with a resonant low- frequency harmonic and can be considered as annoying for the driver (Chaunier et al. 2005). Increased brightness reflects the presence of audible high-order harmonics, whereas increased roughness reflects audible secondary harmonics that interact with the main harmonics. The resulting signal is therefore a mixture of several harmonics and a low-frequency broadband noise.

Although these perceptual attributes can be clearly identified at a given instant, they fail to properly characterize the dynamic variation of the car sounds, during acceleration for instance. Hence, the perception of motor noise must be investigated, more precisely the evocation of identity and perceived quality induced by timbre variations during acceleration. Several methods can be used to elucidate such timbre variations. In this study a sensory analysis similar to the one used in the case of car door noise was first performed on various motor noises [Roussarie et al., 2004]. Among the descriptors that were identified by the panel, three were considered essential to characterize the acceleration, namely "ON" (pronounced [ɔ̃] characterizing the booming of the motor determined by the audibility of low order even harmonics, "REU"(pronounced [rœ]) that characterizes the roughness of the sound and "AN"(pronounced [ɑ̃]) that translates the spectral richness of the motor noise provided by an

increased intensity level of odd harmonics. It was hypothesized that the transition between "ON" and "AN" was linked to an increased impression of sportiness comparted to a monotonous "ON" sound.

In addition to standard Fourier analyses of the signals, an auditory model that focuses on the perceptually relevant parts of the motor noise was applied to the motor noise (Pressnitzer and Gnansia 2005). This model revealed an energy transfer from one group of harmonics towards another during acceleration (Fig. 4). To investigate the dynamic aspect of this energy transfer more thoroughly, vocal imitations in which subjects were asked to imitate an accelerating car were performed. Such approaches have for instance been used in the synthesis of guitar sounds using vocal imitation (Traube and Depalle 2004), to extract relevant features of kitchen sounds (Lemaitre et al. 2011, Lemaitre and Susini, Chap. 9) and are currently used to identify perceptually relevant sound structures of evoked movements and materials (Bordonné et al. 2017). The vocal imitations provided a simplified description of the dynamic evolution that enabled to link the perceived sportiness to the "ON/AN" transitions. A subtractive source/filter synthesis method (which consists in filtering an input signal that generally is either a noise or a pulse train) was then developed that enabled to control both the roughness (determined by the source) and the formantic structure of the sound (determined by the filter). The perceived sportiness could then be intuitively controlled by varying the characteristics of the filter to simulate the ON/AN transition and by modifying the density of the pulse train that constitutes the source related to the roughness (Sciabica et al. 2010, 2012).

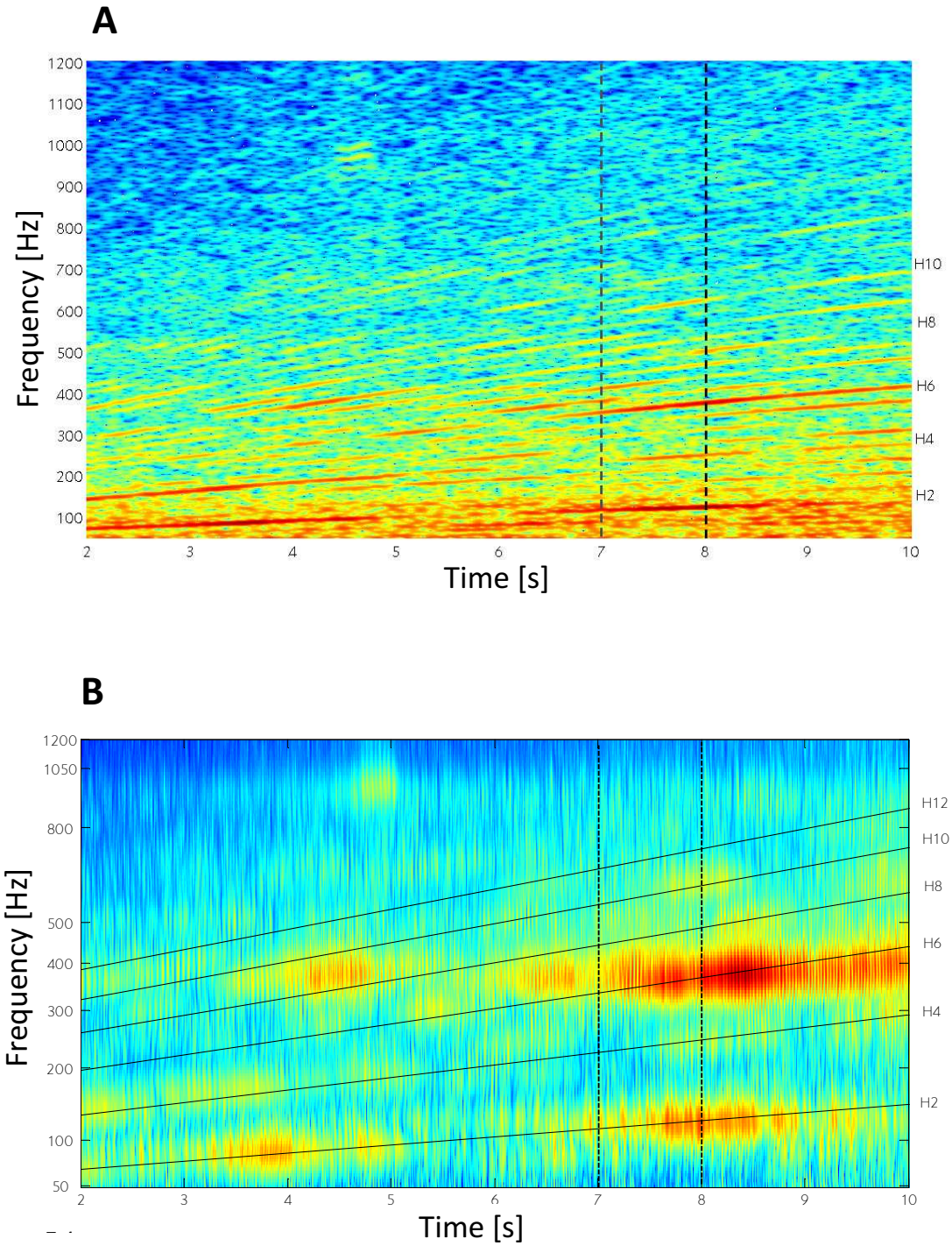

Fig. 4 *Spectrogram of an engine noise during an increase in engine speed (A) and cochleogram of the same engine noise (B). In the lower part of the figure harmonics are indicated by solid lines. The dotted lines at 7 s and 8 s indicate a beating effect between 300 and 400 Hz revealed by the cochleagram. Images extracted from Figs. 7.4 and 7.5 in Sciabica (2011).*

Ystad et al., SHAR 69, 2019, https://doi.org/10.1007/978-3-030-14832-4_13

4.2 Proposing a New Action/Object Paradigm for Sound Synthesis

The last part of this chapter presents the development of new synthesis tools based on perceptual and cognitive studies that unveil perceptually relevant sound morphologies, and the construction of synthesis algorithms that are based on these morphologies and thereby enable intuitive sound control. Previous studies on the perception of various sound categories have led to a new sound synthesis paradigm called the "action/object" paradigm, which considers any sound as the result of an action on an object, and which is based on a semantic description of the sound. This paradigm is coherent with the ecological approach to perception, initially proposed by Gibson (1979) in the case of vision, which suggests the existence of invariant morphological structures associated with the recognition of objects (structural invariants) and actions (transformational invariants). In this paradigm, the notions of action and object can be considered in a broad sense. Hence, the action can be associated with the dynamics of a sound (temporal evolution) and the object with a sound texture. This paradigm is in line with the phenomenological approach to sound listening adopted by Schaeffer (1966) who proposed a classification system of sounds that he called "typology of sound objects." In this typology, Schaeffer proposes a general classification of sounds (both musical and environmental), which relates to the maintenance (*facture* in French) of sounds, that is the way the energy spreads over time, and the spectral content (*masse* in French). Maintenance distinguishes sustained, iterative and impulsive sounds and can be linked to the perceived action, whereas the spectral content distinguishes sounds with constant, varying or undefinable pitch and can be linked to the object.

The action/object paradigm allows for the development of synthesizers that offer sound control from semantic descriptions of the events that created the sound, such as scraping a wooden plate, rubbing a metallic string, rolling on stones (Conan et al. 2014b). Such synthesizers make it possible to continuously navigate in a sound space based on perceptual

invariants of the acoustic signal. This new sound synthesis approach constitutes a radical methodological change and offers new research perspectives in the domains of human perception and cognition, sound design, and musical creation.

This section is divided into three parts. In the first part, invariant structures responsible for the recognition of three different material categories are identified by associating analysis-synthesis techniques with behavioral and brain-imaging approaches. The construction of a timbre space of perceived materials and the mapping between low-level signal parameters, audio descriptors, and semantic control parameters is then described. Further studies on perceived shapes and actions are presented. Finally, a synthesizer that enables the generation of sound metaphors is described.

4.2.1 Perception of Material Categories

In this study the perceptual identification of different materials based on impact sounds was investigated. Particular attention was paid to three different materials: wood, metal, and glass. For this purpose, natural sounds were recorded, analyzed, resynthesized, and tuned to the same pitch class ignoring octave to obtain sets of synthetic sounds representative of each material category. A sound morphing process was then applied to obtain sound continua simulating progressive transitions between materials. This morphing process consisted in mixing the spectra between sounds from different material categories and interpolating the damping laws of the two extreme sounds. Each progressive transition between materials was composed of 22 hybrid sounds. Participants were asked to categorize all the randomly presented sounds as wood, metal or glass in a categorization task. Based on the response rates, ‘‘typical’’ sounds were defined as sounds that were classified by more than 70% of the participants in the same material category and ‘‘ambiguous’’ sounds, those that were classified by less than 70% of the participants in a given category.

While performing the categorization task, reaction times and electrophysiological data were collected using a standard ERP protocol. Analysis of the participants' ERPs showed that the processing of metal sounds differed significantly from that of glass and wood as early as 150 ms after the sound onset. These early differences most likely reflect the processing of spectral complexity (see Shahin et al. 2005;Kuriki et al. 2006) while the later differences observed between the three material categories are likely to reflect differences in sound duration (i.e., differences in damping; see Alain et al. 2002 and McAdams 1999).

The association between the results of the acoustic and electrophysiological analyses suggested that spectral complexity, and more precisely the roughness and both global and frequency-dependent damping, are relevant cues explaining the perceptual distinction between categories (Aramaki et al. 2011). In particular, both global and frequency dependent damping differed between categories with metal sounds that had the weakest damping and sounds from the wood category that were most strongly damped. Metal sounds also had the largest number of spectral components that introduced roughness in these sounds. Glass sounds had the smallest number of spectral components, but a weaker damping than wood sounds.

These results can be linked to the physical behavior of the sound sources (in line with previous studies, see for instance Lutfi and Oh 1997; Klatzky et al. 2000; McAdams et al. 2004, 2010). When the material changes, the wave propagation process is altered by the characteristics of the media. This process leads to dispersion (due to the stiffness of the material) and dissipation (due to loss mechanisms). Dispersion, which introduces inharmonicity in the spectrum, results from the fact that the wave propagation speed varies depending on the frequency. The dissipation is directly linked to the damping of the sound, which is generally frequency-dependent (high-frequency components are damped more quickly than low-frequency components). These results made it possible to determine the acoustic invariants associated with various sound categories and propose a timbre space of material categories (Fig.

5).

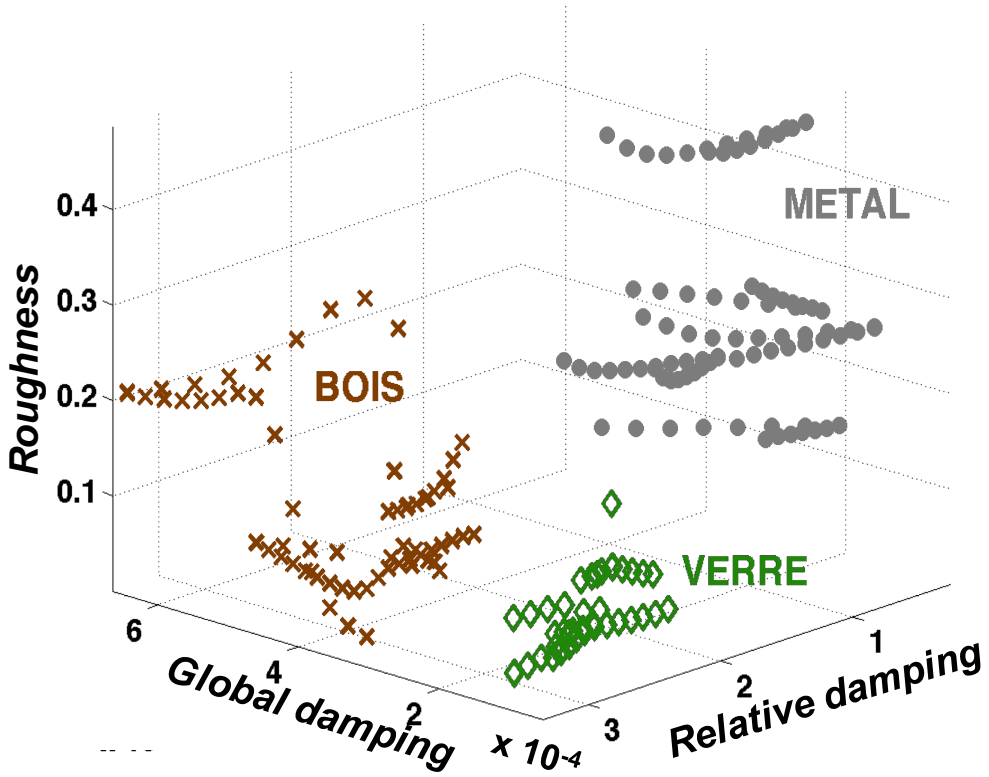

Fig. 5 *Timbre space of material categories*

In addition, results showed that ambiguous sounds were associated with slower reaction times than typical sounds. As might be expected, ambiguous sounds are therefore more difficult to categorize than typical sounds. This result is in line with previous findings in the literature showing slower RTs for nonmeaningful than for meaningful sounds (e.g., Cummings et al. 2006).

The differences between typical and ambiguous sounds were smaller in the wood–metal and glass–metal continua than in the wood–glass continuum. This is interesting from an acoustic perspective because metal sounds typically present higher spectral complexity (related to the density and repartition of spectral components) than both wood and glass sounds that

show closer sound properties. Thus, ambiguous sounds in wood–metal and glass–metal continua were easier to categorize than those in the wood–glass continuum and the ambiguity effect was smaller.

The same categorization protocol was used in a more recent study with participants diagnosed with schizophrenia. Results interestingly revealed that the transitions between material categories were shallower for these participants than for control participants, suggesting the existence of perceptual impairments in such patients due to sensory processing dysfunction (Micoulaud-Franchi et al. 2011).

The timbre space of material categories is particularly interesting from synthesis and control perspectives, since it provides clues for establishing links between low-level synthesis parameters (amplitudes, frequencies, etc.), acoustic descriptors describing pitch and timbre, and semantic labels (wood, metal, glass) that can be used in high-level control interfaces. The positions of the sounds in this material space tell us for instance how a sound that evokes wood can be transformed into a sound that evokes metal by decreasing the damping factors and increasing the roughness. Mapping strategies that enable intuitive controls can therefore be proposed (Fig. 6).

4.2.2 Perception of Shapes and Excitation

Previous acoustic studies on the links between perception and the physical characteristics of sound sources have brought to light several important properties that can be used to identify the perceived effects of action on an object and the properties of the vibrating object itself (see Aramaki et al. 2009, 2010 for a review). In addition to the frequency-dependent damping and roughness that were found to be important for the perceptual identification of material properties (see McAdams, Chap.2 and Agus, Suied and Pressnotzer, Chap.3), the

perceived hardness of a mallet striking a metallic object is predictable from the characteristics of the attack time. From a physical point of view, the shape of the impacted object determines the spectral content of the impact sound. The frequencies of the spectral components correspond to the so-called eigenfrequencies, which characterize the modes of the vibrating object and convey important perceptual information about the shape. Previous studies have investigated the auditory perception of physical attributes linked to shape, hollowness or material. In particular, studies on the geometry of objects have demonstrated that height-width ratios and lengths could be recovered from sounds with reliable accuracy (Lakatos et al. 1997; Carello et al. 1998). In terms of cavity, Lutfi (2001) showed that the perception of hollowness could be related to frequency judgments and to some extent (depending on the subjects) to acoustic parameters such as damping. Rocchesso (2001) revealed that spherical cavities sounded brighter than rounded cavities, since sounds were more strongly absorbed in cubes than in spheres.

In a previous study, Rakovec et al. (2013) investigated the perception of shapes, and found that three dimensional shapes (bowls, tubes, …) were easier to recognize than one dimensional objects (bars, strings, …). The Hollow and Solid attributes appeared to be quite evocative, since no confusion between Hollow and Solid was made. The results also revealed a mutual influence between the perceived material and the perceived shape, in line with Tucker and Brown (2002) and Giordano (2003), who found that shape recognition abilities were limited and strongly depended on the material composition.

The perception of the size of the object is mainly correlated with pitch: large objects generally vibrate at lower eigenfrequencies than do small ones. In the case of quasi-harmonic sounds, we assume the pitch to be related to the frequency of the first spectral component. All of these observations lead to the hypothesis of a three-layer mapping strategy that links basic signal parameters via acoustic descriptors to high-level semantic control parameters (see Fig.

6).

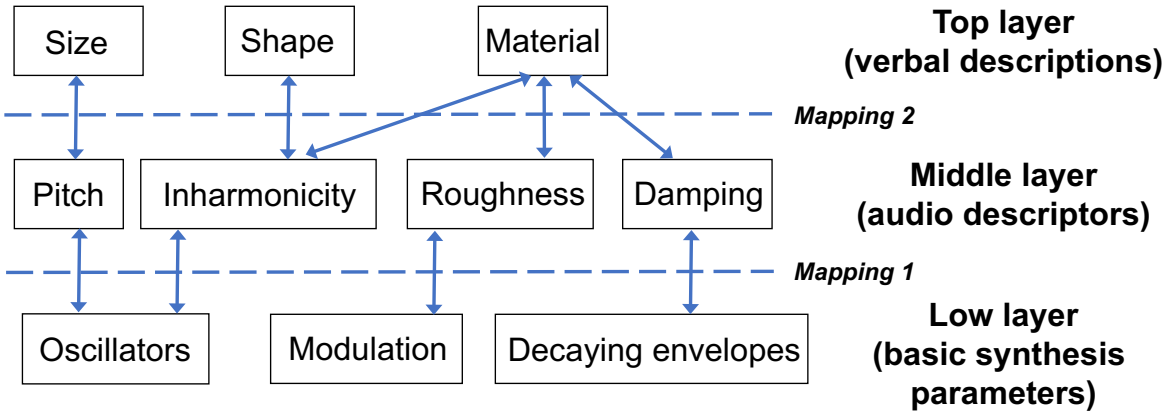

Fig. 6 *Three-level mapping strategy between basic synthesis parameters (Low Layer), audio descriptors (Middle Layer) and verbal descriptions (Top Layer).*

4.2.3 Perception of Various Actions

Previous studies have revealed that invariant sound structures linked to the evoked object, such as combinations of specific damping factors and spectral density in the case of material perception, pitch in the case size perception, or harmonic structure in the case shape perception, can be identified. The next question is whether invariants linked to the sound-producing action, so-called transformational invariants, can also be found. For this purpose, several studies on continuous interactions between solid objects were performed by considering a subset of continuous interaction sounds, that is, rubbing, scratching, and rolling. Synthesis models for such sounds have already been proposed in previous studies, some based on physical modeling or physically informed considerations (van den Doel et al. 2001; Rath and Rocchesso 2004, Stoelinga and Chaigne 2007; Houben 2002), others on analysis-synthesis schemes (Lagrange et al. 2010; Lee et al. 2010). Like the control space described in section 4.2.1 that allows the user to control the perceived material and to morph continuously from one material

to another (e.g., from glass to metal through a continuum of ambiguous materials), a control space that enables continuous control of evoked interactions was developed in this study (e.g., being able to synthesize a rubbing sound and slowly transform it into a rolling one). For this purpose, Conan et al. (2014a) first identified invariants related to the auditory perception of interactions. Phenomenological considerations, physical modeling, and qualitative signal analysis were investigated. They concluded that the interaction forces conveyed the relevant perceptual information regarding the type of interaction. In particular, the interaction force associated with rubbing and scratching sounds could be modeled as an impact series in which impacts are separated by shorter time intervals for rubbing than for scratching. To evoke rolling sounds, it is necessary to take into account the strong correlation between the amplitudes of each impact and the time interval that separates them. These findings led to the design of a generic synthesis model that sought to reproduce these interaction forces. An intuitive control space that enables continuous transitions between these interactions was designed. By combining this "action space" with the simulation of the properties of the object (material and shape as described in sections 4.2.1 and 4.2.2), an interface that enables intuitive and interactive real-time control of evoked actions and objects could be designed, as illustrated in Fig.7. This interface could easily be connected to various control devices.

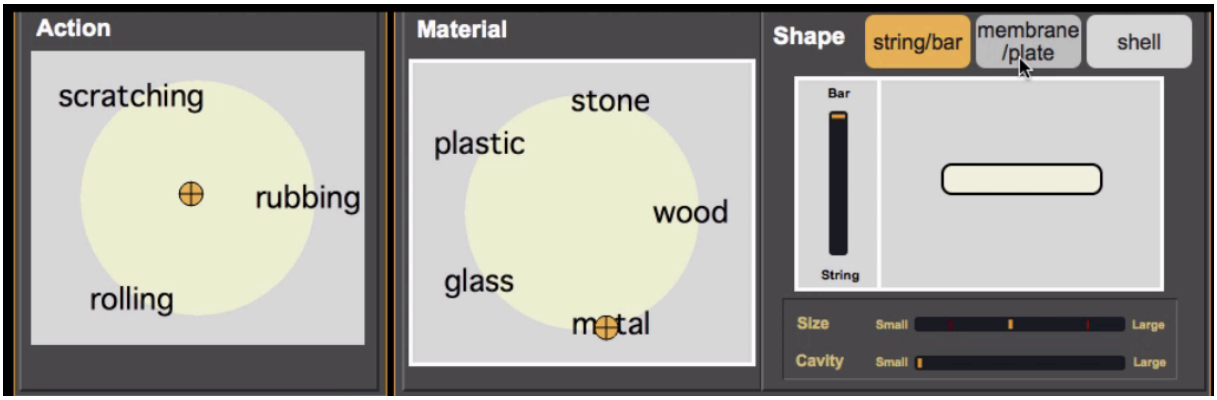

***Fig.** 7 Synthesis interface that enables intuitive and interactive real-time control of evoked* actions and objects.

In addition to the navigation in this "action space," the gesture can be taken into account in the control strategy. Indeed, for such continuous interactions, the underlying gesture is a fundamental attribute that can be conveyed in the dynamics of the sound (Merer et al. 2013; Thoret et al. 2014). Following the synthesis process discussed by van den Doel and colleagues (2001), the resulting interaction force is low-pass filtered with a cutoff frequency directly related to the relative transversal velocity between the objects that interact (hand, plectrum, etc.) and the surface. When associated with a biological law, a specific calibration of the velocity profile enables the evocation of a human gesture (Thoret et al. 2014). Such a synthesis tool has been used in several applications, for instance for video games (Pruvost et al. 2015) and associated with a graphic tablet to sonify handwriting as a remediation device for dysgraphic children (Danna et al. 2015).

4.2.4 Sound Metaphors

As we have seen previously, the association between labels describing the sound source and perceptually invariant sound structures makes it possible to envisage new sound spaces in which the users can easily create and transform synthesis sounds. Each label describing the sound source is associated, via invariant sound structures, to low-level synthesis parameters. Sounds can thereby be continuously transformed between different materials, shapes or actions. Sound synthesis thus offers a large field of sound investigations in which the verbal descriptions of actions and objects only constitute intuitive support to the expression of the composer's imagination.

Even if the action/object approach is naturally adapted to the control of realistic sounds produced by objects belonging to our surroundings, one might wonder if it would also be

possible to satisfy Varèse's old dream about the creation of "a whole new world of unsuspected sounds." Listening tests have shown that actions and objects even can be evoked through abstract sounds for which the source cannot be easily recognized (Merer et al. 2011, 2013). Also, as mentioned earlier, the typology of sounds proposed by Schaeffer, which is well adapted to analytic listening, is based on the notions of *facture* and *masse*, two attributes that are strongly correlated to dynamics (energy, gesture) and texture (matter), closely linked to the notion of action and object. Hence, the unexpected association between objects and actions might be a means to guide composers towards their search for unsuspected or unheard sounds. The sound space dedicated to the intuitive control of solid objects and their interactions presented earlier in this section, makes it possible to freely associate actions and objects. This means that it is possible to simulate physically impossible situations and for instance rub the wind, make a water drop bounce, or make an orchestra squeak. Even if it is difficult to describe sounds that we imagine with words from our language or with metaphoric expressions, new experiences reveal that such a control space opens the door to the creation of unsuspected sounds that conserve the cognitive references of objects and actions thanks to the invariant structures on which the control space is founded.

5 Summary

This chapter describes how timbre and more generally perceptually relevant sound morphologies can be identified using the analysis-by-synthesis approach. In Sect 1.2, dedicated to musical sounds, perceptual cues used by a xylophone maker to select the optimal wood species for xylophones are identified. Frequency-dependent damping as well as the spectral bandwidth turned out to be the most salient descriptors indicating that the xylophone maker searched for highly resonant and crystal-clear sounds. The role of timbre in musical performance was also investigated, revealing that musicians consciously used timbre variations

to enhance expressiveness. A study on cello performance also showed that ancillary gestures were important to produce a round (as opposed to harsh) timbre.

In the second part of the chapter, particular attention was given to environmental sounds, both in order to better understand how meaning is conveyed by such sounds and to extract sound morphologies that enable the construction of synthesizers that offer easy, intuitive, and continuous sound controls. For this purpose, electrophysiological measurements were performed to investigate how sense is attributed to environmental sounds and whether the brain activity associated with the interpretation of such sounds is similar to the brain activity observed in the case of language processing. These studies confirmed the existence of a semiotics of isolated sounds, thus suggesting that a language of sounds might be drawn up based on invariant sound structures.

In the last part of the chapter, perceptually salient signal morphologies were identified and associated with evocations of quality, solidity, and sportiness for sounds produced by cars. Invariant structures linked to evocations of solid sounds and their interactions could also be extracted and used to develop synthesizers that enable intuitive control from semantic descriptions of sound events. It should be mentioned that other invariant structures that are not described in this chapter have been identified in the case of evoked motion (Merer et al. 2013) and that a synthesizer of environmental sounds that offers intuitive control of auditory scenes (rain, waves, wind, fire, footsteps) has been developed by Verron and collaborators (2010).

These developments open the way to new and captivating possibilities for using nonlinguistic sounds for communication. Further extending our knowledge in this field, including for instance new approaches linked to machine learning (see Peeters, Chap. 11) and neural responses obtained by spectro-temporal receptive fields (see Elhilali, Chap. 12), should make it possible to develop new tools for generating sound metaphors (see Saitis and Weinzierl, Chap. 5) based on invariant signal structures that can be used to evoke specific mental images

via selected perceptual and cognitive attributes. These metaphors can either be constructed from scratch or obtained by shaping initially inert sound textures using intuitive (high-level) control approaches.

References

Alain C, Schuler B M, McDonald K L (2002) Neural activity associated with distinguishing concurrent auditory objects. *J. Acoust. Soc. Amer.* 111 (2): 990–995

Angster J, Rucz P, Miklós A (2017) Acoustics of Organ Pipes and Future Trends in the Research. Acoustics Today 13(1), pp.10-18, Spring

Aramaki M, Baillères H, Brancheriau L et al (2007) Sound quality assessment of wood for xylophone bars. *J. Acoust. Soc. Am.* 121(4): 2407–2420

Aramaki M, Marie C, Kronland-Martinet R et al (2009) Sound Categorization and Conceptual Priming for Non Linguistic and Linguistic Sounds. *Journal of Cognitive Neuroscience* 22(11): 2555-2569

Aramaki M, Gondre C, Kronland-Martinet R et al (2010) Imagine the sounds : an intuitive control of an impact sound synthesizer. *Computer Music Modeling and Retrieval - Auditory Display*, LNCS 5954 Springer-Verlag Heidelberg, p 408-422

Aramaki M, Besson M, Kronland-Martinet R. et al (2011) Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2): 301-314

Avanzini F, Rocchesso D (**2001)** Controlling material properties in physical models of sounding objects. In *Proceedings of the International Computer Music Conference,* 17–22 September 2001, Hawana, p 91–94

Barthet M, Guillemain Ph, Kronland-Martinet R. et al (2010a) From Clarinet Control to Timbre Perception. *Acta Acustica*, 96:678-689

Barthet M, Depalle Ph, Kronland-Martinet R et al (2010b) Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception: An Interdisciplinary Journal* 28 (2):135-154

Barthet M, Depalle Ph, Kronland-Martinet R et al (2011) Analysis- by-Synthesis of Timbre, Timing, and Dynamics in Expressive Clarinet Performance. *Music Perception: An Interdisciplinary Journal*, 28 (3): 265-279

Bedos De Celles D. F. (1766) L'art du facteur d'orgues,("The art of the organ builder") *Reprint of the original edition from Paris by Slatkine, 2004*, (ISBN 2051019398)

Bensa J, Bilbao S, Kronland-Martinet R et al (2003) The Simulation of Piano String Vibration: From Physical Models to Finite Difference Schemes and Digital Waveguides. *J. Acoust. Soc. Am.*, 114(2):1095-1107

Bezat, M. (2007). "Perception des bruits d'impact: Application au bruit de fermeture de porte automobile" ("Perception of impact noises: Application to car-door closure noise"), Ph.D. dissertation, Université Aix-Marseille I, Provence, France

Bezat M C, Kronland-Martinet R, Roussarie V et al (2014) From acoustic descriptors to evoked quality of car-door sounds. *J. Acoust. Soc. Am.* 136(1): 226-241

Bilbao S, Webb C (2013) Physical Modeling of Timpani Drums in 3D on GPGPUs. Journal of the Audio Engineering Society, 61(10):737–748

Bordonné T, Dias-Alves M, Aramaki M et al (2017) Assessing sound perception through vocal imitations of sounds that evoke movements and materials. In Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research CMMR "Music Technology with Swing", Matosinhos, Portugal, September 25-28

Carello C, Anderson K L, Kunkler-Peck A J (1998) Perception of Object Length by Sound. Psychological Science 26(1): 211–214

Chadefaux D, Le Carrou J-L, Wanderley M M et al (2013) Gestural strategies in the harp performance. Acta Acust.. 99(6): 986–996

Chaunier L, Courcoux P, Della ValleG. et al (2005) Physical and sensory evaluation of cornflakes crispness. *Journal of Texture Studies* 36:93-118

Conan S, Derrien O, Aramaki M, et al (2014a) A synthesis model with intuitive control capabilities for rolling sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 22(8): 1260–1273

Conan S, Thoret E, Aramaki M, et al (2014b) An intuitive synthesizer of continuous interaction sounds : Rubbing, Scratching and Rolling. *Computer Music Journal* 38(4):24-37, doi :10.1162/COMJ_a_00266.

Cook P R (2001) Principles for Designing Computer Music Controllers. In: Proceeding of New Interfaces for Musical Expression, NIME-01, Seattle, WA, USA, April 1-2, 2001

Cook P R, Scavone G (1999) The Synthesis Toolkit. International Computer Music Conference, Beijing, China, October 22-27, 1999

Cummings A, Ceponiene R, Koyama A, et al (2006) Auditory semantic networks for words and natural sounds. Brain Research, 1115: 92–107

Danna J, Paz-Villagrán V, Gondre C, et al (2015) Let me hear your handwriting! Evaluating the movement fluency from its sonification. *PLoS One*, 10(6)

Desmet F, Nijs L, Demey M, et al (2012) Assessing a clarinet player's performer gestures in relation to locally intended musical targets. J. New Music Res. 41(1): 31–48

De Wilde L (2016) Les fous du son, d'Edison à nos jours. *Ed. Grasset*, 560p.

Fletcher N H, Rossing T D **(1998)** The Physics of Musical Instruments, 2nd ed. Springer-Verlag, Berlin.

Fourier J (1878) The Analytical Theory of Heat. Cambridge University Press.

Gaver W W (1993) What in the world do we hear? An ecological approach to auditory event perception. Ecol. Psychol., 5(1): 1–29

Gibson J J (1979) The Ecological Approach to Visual Perception. *Boston: Houghton Mifflin*.

Giordano B L (2003) Material Categorization and Hardness Scaling in Real and Synthetic Impact Sounds. Rocchesso, D, Fontana F (eds.) The Sounding Object, pp. 73–93. Mondo Estremo, Firenze

Giordano B L, McAdams S. **(2006)** Material identification of real impact sounds: Effects of size variation in steel, wood, and Plexiglass plates. J. Acoust. Soc. Am. **119**(2):1171–1181

Gobin P, Kronland-Martinet R, Lagesse G A, et al (2003) From Sounds to Music : Different Approaches to Event Piloted Instruments. Lecture Notes in Computer Science, LNCS 2771, Springer Verlag, pp 225-246

Goodchild M, Wild J, McAdams S (2017) Exploring Emotional Responses to Orchestral Gestures. Musicae Scientiae, 1-25. https://doi.org/10.1177%2F1029864917704033

Grey J M (1977) Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61:1270-1277.

Guillemain P, Kergomard J, Voinier T (2005) Real-time synthesis of clarinet-like instruments using digital impedance models. *J. Acoust. Soc. Am.* 118: 483–494

Hajda J M, Kendall R A, Carterette E C et al (1997) Methodological issues in timbre research. *In* Deliége I, Sloboda J(ed): *Perception and Cognition of Music*. Psychology Press, New York, 2nd edition, pp 253–306

Helmholtz H (1868) On the Sensations of Tone as a Physiological Basis for the Theory of Music. *Longmans, Green, and co.*

Houben M (2002) The Sound of Rolling Objects, Perception of Size and Speed. PhD dissertation, Technische Universiteteit, Eindhoven.

Huang N E, Shen Z, Long S, et al (1998). "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. In: *Proc. R. Soc. A London*, 454(1971): 903–995

Klatzky R L, Pai D K, Krotkov E P **(2000)** Perception of material from contact sounds. Presence: Teleoperators and Virtual Environments, **9**(4):399–410

Krimphoff J, McAdams S, Winsberg S (1994) Caractérisation du timbre des sons complexes, II Analyses acoustiques et quantification psychophysique (Characterization of complex sounds timbre, II Acoustical analyses and psychophysical quantification). *Journal de Physique IV, Colloque C5*, 4:625–628

Kronland-Martinet R, Guillemain P, Ystad S (1997) Modelling of Natural Sounds Using Time-Frequency and Wavelet Representations. Organised sound, Cambridge University Press 2(3): 179-191

Kuriki S, Kanda S, Hirata Y (2006) Effects of musical experience on different components of meg responses elicited by sequential piano- tones and chords. *J. Neurosci.*, 26(15): 4046–4053

Kutas M, Hillyard S A (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207: 203–204

Kuwano S, Fastl H, Namba S. et al (2006) Quality of door sounds of passenger cars. Acoust. Sci. Technol. 27(5): 309–312

Lagrange M, Scavone G, Depalle P. (2010) Analysis/synthesis of sounds generated by sustained contact between rigid objects. IEEE Transactions on Audio, Speech, and Language Processing, 18(3): 509–518

Lakatos S, MacAdams S, Caussé R (1997) The Representation of Auditory Source Characteristics: Simple Geometric Form. Attention, Perception, & Psychophysics. 59(8): 1180–1190.

Lembke S A, Levine S, McAdams S. (2017) Blending between bassoon and horn players: an analysis of timbral adjustments during musical performance. Music Perception, 35 (2) : 144-164

Lemaitre G, Dessein A, Susini P et al (2011) Vocal imitations and the identification of sound events. Ecol. Psychol. 4(23): 267–307

Lee J S, Depalle P, Scavone G (2010) Analysis/synthesis of rolling sounds using a source-filter approach. In: 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria.

Lutfi R A (2001) Auditory Detection of Hollowness. *J. Acoust. Soc. Am.* 110(2): 1010–1019

Lutfi R A, Oh E L (**1997)** Auditory discrimination of material changes in a struck-clamped bar. J. Acoust. Soc. Am. **102**(6): 3647– 3656

Mathews, M V (1963) The digital computer as a musical instrument. Science, 142 (3592): 553-557

Mathews M V, Miller J E, Pierce J R et al (1965) Computer study of violin tones. J. Acoust. Soc. Am., 38(5): 912-913

McAdams, S. & Bigand (1993), E. Thinking in Sound: The cognitive psychology of human audition. *Oxford University Press.*

McAdams S, Winsberg S, Donnadieu S, et al (1995) Perceptual scaling of synthesized musical timbres : common dimensions, specificities, and latent subject classes. Psychological Research, 58:177–192

McAdams S (1999) Perspectives on the contribution of timbre to musical structure. Comput. Music J. 23(3):, 85–102

McAdams S, Chaigne A, Roussarie V (**2004)** The psychomechanics of simulated sound sources: Material properties of impacted bars. J. Acoust. Soc. Am. **115**(3):1306–1320

McAdams S, Roussarie V, Chaigne A, Giordano B L (2010) The psychomechanics of simulated sound sources: Material properties of impacted thin plates. J. Acoust. Soc. Am. 128 (3):1401-1413. doi:10.1121/1.3466867

Merer A, Ystad S, Kronland-Martinet R et al (2011) Abstract sounds and their applications in audio and perception research. In Exploring music contents, LNCS vol. 6684, Springer-pp. 176-187

Merer A, Aramaki M, Ystad S et al (2013) Perceptual characterization of motion evoked by sounds for synthesis control purposes. Association for Computing Machinery, Transactions on Applied Perception (TAP), 10(1): 1-24

Micoulaud-Franchi J A, Aramaki M, Merer A, et al (2011) Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research* 189(1): 149-152

Miranda R, Wanderley M (2006) New Digital Musical Instruments: Control and Interaction Beyond the Keyboard. A-R Editions, 286pp.

Moog R (1987) Position and force sensors and their application to keyboards and related controllers. In: Proceedings of the AES 5th International Conference: Music and Digital Technology, A. E. S. New York, Ed., pp. 179–181.

Moore T R (2016) The Acoustics of Brass Musical Instruments. Acoustics Today, 12(4): 30-37, Winter 2016

Peeters G, Giordano B, Susini P, et al (2011) The Timbre Toolbox: Audio descriptors of musical signals. *J. Acoust. Soc. Am.*, 130 (5)

Pierce J (1965) Portrait of the machine as a young artist. Playboy 12(6), 124-5 and 150 and 182 and 184

Pressnitzer D, Gnansia D (2005) Real-time auditory models. Proceedings of International Computer Music Conference, Barcelona, Spain, September 5-9, p 295-298

Pruvost L, Scherrer B, Aramaki M, et al (2015). Perception-Based Interactive Sound Synthesis of Morphing Solids' Interactions. In: Proceedings of the Siggraph Asia 2015, Kobe, Japan, 2-5 November.

Rakovec C E, Aramaki M, Kronland-Martinet R (2013) Perception of Material and Shape of Impacted Everyday Objects. In: Proc. of the 10th International Symposium on Computer Music Multidisciplinary Research, Marseille, France, October 15-18, p 943-959

Rath M, Rocchesso D (2004) Informative sonic feedback for continuous human–machine interaction—controlling a sound model of a rolling ball. IEEE Multimedia Special on Interactive Sonification, 12(2): 60–69

Risset J C (1965) Computer study of trumpet sounds. J. Acoust. Soc. Am., 38(5): 912

Roads C (1988) Introduction to Granular Synthesis. Computer Music Journal, 12(2): 11-13. doi:10.2307/3679937

Roussarie V, Richard F, Bezat M C (2004) Validation of auditory attributes using analysis synthesis method. Congrès Francais d'Acoustique/DAGA (Strasbourg, Germany).

Rocchesso D (2001) Acoustic Cues for 3-D Shape Information. In: Proc. of the 2001 Int. Conf. on Auditory Display, Espoo, Finland, July 29-August 1, pp 175–180

Rozé J, Aramaki M, Kronland-Martinet R, et al (2016) Exploring the effects of constraints on the cellist's postural displacements and their musical expressivity. In Music, Mind and Embodiment, LNCS 9617, Springer-Verlag Heidelberg.

Rozé J, Aramaki M, Kronland-Martinet R et al (2017) Exploring the perceived harshness of cello sounds by morphing and synthesis techniques. J. Acoust. Soc. Am., 141(3): 2121-2136

Saitis C, Giordano B, Fritz C et al (2012) Percpetual evaluation of violins: A quantitative analysis of preference judgments by experienced players. J. Acoust. Soc. Am. 132 (6):4002-4012

Saitis C, Fritz C, Scavone G, et al (2017) A psycholinguistic analysis of preference verbal descriptors by experienced musicians. J. Acoust. Soc. Am. 141 (4):2746-2757

Schaeffer P (1966) Traité des objets musicaux (Treatise of musical objects). *Éditions du Seuil. (An English translation by Christine North and John Dack is available at* https://books.google.de/books?id=q9oMvgAACAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepag&q&f=false*)*

Scholl D, Amman S (1999) A new wavelet technique for transient sound visualization and application to automotive door closing events. In: Proceedings of the SAE Noise and Vibration Conference and Exposition, Traverse City, MI, USA

Schön D, Ystad S, Kronland-Martinet R et al (2009) The evocative power of sounds: Conceptual priming between words and nonverbal sounds. Journal of Cognitive Neuroscience, 22: 1026–1035

Sciabica J F, Bezat M C, Roussarie V, et al (2010). Timbre Characteristics of Interior Car Sound. Auditory Display, Springer Verlag Berlin Heidelberg, pp 377-391

Sciabica J F (2011) Modélisation perceptive du bruit habitacle et caractérisation de son ressenti (Perceptual modelling of interior car noise and characterization of induced sensation), Ph.D. dissertation, Aix-Marseille Univ., France, September 19th. 2011

Sciabica J F, Olivero A, Roussarie V, et al (2012) Dissimilarity Test modelisation applied to Interior Car Sound Perception. Proceedings AES 45th International Conference on Applications of Time-Frequency Processing in Audio, Helsinki, Finland, March 1- 4, 2012

Shahin A, Roberts L E, Pantev C, et al (2005) Modulation of p2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport, 16 (16): 1781–1785

Stoelinga C, Chaigne A (2007) Time-Domain Modeling and Simulation of Rolling Objects. Acta Acustica United with Acustica 93(2):290–304

Sundberg J (2000) Grouping and differentiation two main principles in the performance of music. In Integrated Human Brain Science : Theory, Method Application (Music), 299–314. Elsevier Science B.V.

Thompson M R, Luck G (2012) Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. Musicae Sci. 16(1): 19–40

Traube C, Depalle P (2004) Timbral analogies between vowels and plucked strings tones. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP04) (Montreal, Quebec), 4:293–296

Tucker S, Brown G J (2002) Investigating the Perception of the Size, Shape and Mateial of Damped and Free Vibrating Plates. Technical Report, Department of Computer Science, University of Sheffield

van den Doel K, Kry P G, Pai, D K (2001) FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, Los Angeles, CA, USA, August 12-17, 2001, pp 537–544

Van Zijl A G, Luck G (2013) Moved through music: The effect of experienced emotions on performers' movement characteristics. Psychol. Music 41(2): 175–197

Varèse E (1917) Que la Musique Sonne. *391*, n°5, June 1917

Verron C, Aramaki M, Kronland-Martinet R et al (2010) A 3D Immersive synthesizer for environmental sounds. IEEE Transactions on Audio, Speech, and Language Processing, 18(6): 1550-1561

Wanderley M M, Vines B W, Middleton N, et al (2005) The musical significance of clarinetists' ancillary gestures: An exploration of the field. J. New Music Res. 34(1): 97–113

Warren W, Verbrugge R (1984) Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics. Journal of Experimental Psychology: Human Perception and Performance, 10(5): 704–712

Wessel D L (1979) Timbre Space as a Musical Control Structure. Computer Music Journal, 3(2): 45-52, June 1979

Wildes R P, Richards W A **(1988)** Recovering Material Properties from Sound. MIT Press,

Cambridge, MA , Chap. 25, pp. 356–363
Wolfe J (2018) The Acoustics of Woodwind Musical instruments. Acoustics Today, Spring 2018, 14(1):50-56