AI & Soc (2012) 27:245–255
DOI 10.1007/s00146-011-0340-8

ORIGINAL ARTICLE

High-level control of sound synthesis for sonification processes

Richard Kronland-Martinet · Sølvi Ystad · Mitsuko Aramaki

Received: 15 October 2010 / Accepted: 1 August 2011 / Published online: 6 September 2011

Abstract Methods of sonification based on the design and control of sound synthesis is presented in this paper. The semiotics of isolated sounds was evidenced by performing fundamental studies using a combined acoustical and brain imaging (event-related potentials) approach. The perceptual cues (which are known as invariants) responsible for the evocations elicited by the sounds generated by impacts, moving sound sources, dynamic events and vehicles (car-door closing and car engine noise) were then identified based on physical and perceptual considerations. Lastly, some examples of the high-level control of a synthesis process simulating immersive 3-D auditory scenes, interacting objects and evoked dynamics are presented.

Keywords High-level control · Invariants · Synthesis · Semiotics

R. Kronland-Martinet (✉) · S. Ystad
Laboratoire de Mécanique et d'Acoustique, CNRS,
31, Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
e-mail: kronland@lma.cnrs-mrs.fr

S. Ystad
e-mail: ystad@lma.cnrs-mrs.fr

M. Aramaki
Institut de Neurosciences Cognitives
de la Méditerranée, CNRS,
31, Chemin Joseph Aiguier, 13402
Marseille Cedex 20, France
e-mail: aramaki@incm.cnrs-mrs.fr

M. Aramaki
Aix-Marseille University,
38, Boulevard Charles Livon, Marseille, France

1 Introduction

Sonification consists in associating the data obtained by performing measurements of various kinds with sounds. The main idea is to use human beings' natural perceptual ability to identify sound structures to detect hidden coherences in a set of data. Sonification is often based on "music-like" strategies, whereby data are attributed to elementary sound sequences in which the tempo, rhythm and the duration of the sounds are controlled. Few studies have dealt so far with the use of sounds as a "language" based on timbre.

In this paper, we focused on the design and control of synthesis processes for sonification purposes. In the first step, a fundamental approach based on brain imaging techniques was used to investigate the semiotics of sounds, i.e., how we attribute a meaning to specific sounds. Experiments based on the use of categorization and priming tasks were then conducted in order to determine how isolated and contextualized sounds are processed by normal subjects. The results obtained suggest the existence of structural invariants which endow sounds with specific meanings. To determine these invariants, we further analyzed both the physical behavior of the sound-generating sources and the perceptual impact of the sounds on the listeners. Examples of the invariants associated with the physical properties of impact sounds and moving sound sources are presented. In the case of other less clearly defined situations such as the evoked dynamics or quality of industrial sounds, the invariants associated with the signal morphologies were identified by performing listening tests. Based on the invariant signal structures found to be specific to given sound categories, high-level control of real-time synthesis processes on the basis of the parameters mainly responsible for perceptual evocations was developed. These

control strategies can be used as the basis of sonification processes, since the control parameters can easily be mapped onto external data. The following three concrete examples of high-level control applications are presented: a real-time spatialized synthesizer designed to generate immersive 3-D auditory scenes, an impact synthesizer simulating sounds generated by interacting objects and an evoked dynamics strategy whereby sounds can be controlled on the basis of either written words or drawings. In this review, it is proposed to summarize several projects and papers published during the last 4 years by members of our research group, in partnership with researchers from the Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France (Besson and Schön) and the company PSA-Peugeot Citroën Automobiles, Velizy-Villacoublay, France (Roussarie and Bezat). First, we will present the results obtained on the investigation of semiotics of sound, using a combined sound modeling and cognitive neuroscience approach. Methods of identifying structural invariants are then described, and their application to developing high-level control strategies for sound synthesizers is presented. Lastly, we discuss the potential of these methods of sound synthesis as means of conveying specific meanings via different sound textures (sound metaphors), which constitute a promising new means of sonifying data.

2 On the existence of sound semiotics

Sound is a vital carrier of information, since we are able to categorize various types of sounds and construct exact mental representations of the environment from auditory scenes. Here, we present the results of several experimental studies on the nature of sound semiotics, based on the use of categorization and conceptual priming procedures, in which special care was taken to design an appropriate sound corpus. In these studies, we used either synthetic stimuli using analysis/transformation/synthesis processes or sounds of a specific kind called "abstract" sounds promoting "musical" listening, which involves the perception of the quality of the sound, as opposed to "everyday" listening, which is a more source-oriented kind of listening (Gaver 1993a, b). The behavioral data recorded, consisting of the participants' responses and reaction times (RTs), provided objective measurements to the processing of stimulus complexity in the various tasks.

When appropriate, we also investigated the neural bases of the brain processes involved by analyzing the event-related potentials (ERPs) time-locked to the stimulus onset during the various information processing stages. In general, the ERPs elicited by a stimulus (a sound, a light, etc.) are characterized by a series of positive (P) and negative (N) deflections relative to a baseline. These deflections (called components) were characterized in terms of their polarity, their maximum latency (relative to the stimulus onset), their distribution among several electrodes placed in standard positions on the scalp and by their functional significance. Components P100, N100 and P200, which are known to reflect the sensory and perceptual information processing stages, were consistently activated in response to the auditory stimuli (Rugg and Coles 1995). Several late ERP components (N200, P300, N400, etc.) were subsequently elicited, which may be associated with specific brain processes depending on the experimental design and the task in hand.

2.1 Sound categorization: the perception of sounds from impacted materials

Environmental sounds constitute the permanent auditory world we live in. Previous authors have defined these sounds as everyday sounds other than speech, music, animal communications and electronic sounds. Several studies have dealt with the identification and classification of such sounds (Ballas 1993, Gygi et al. 2007, Gygi and Shafiro 2007, Vanderveer 1979). These studies support the ecological theories proposed by Gibson (1986), according to whom listeners tend to identify their environment by perceiving the properties of a sound event rather than the acoustical properties of the signal. A hierarchical taxonomy of everyday sounds based on the physics of sound events was proposed by Gaver (1993a). The first level in this hierarchy contains three main categories: vibrating solids (impacts, deformation, etc.), aerodynamic sounds (wind, fire, etc.) and liquid sounds (drops, splashes, etc.). The latter author pointed out that as previous studies have shown, nobody ever mistakes the sounds belonging to these categories at the perceptual level.

We have studied the perception of the class consisting of impact sounds, especially that of impacts with materials (wood, metal and glass) (Aramaki et al. 2009, 2010b, 2011). For this purpose, natural sounds were recorded, analyzed, resynthesized and tuned to the same chroma to obtain sets of synthetic sounds representative of each category of material selected. A sound morphing process was then applied to obtain sound continua simulating progressive transitions between materials. The main aim here was to create typical and ambiguous impact sounds in order to be able to vary the difficulty of the categorization task. Although sounds located at the extreme positions on the continua were indeed perceived as typical exemplars of their respective categories, sounds in intermediate positions were synthesized by interpolating the acoustic parameters characterizing sounds at extreme positions and were consequently expected to be perceived as ambiguous (e.g., to be neither wood nor metal). Participants were asked to categorize all the randomly

presented sounds as wood, metal or glass (in a constrained categorization task). Based on the response rates, "typical" sounds were defined as sounds that were classified by more than 70% of participants in the same category of material and "ambiguous" sounds, those that were classified by less than 70% of the participants in a given category.

Analysis of the subjects' 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. The results of the acoustic and electrophysiological analyses suggested that spectral complexity and sound duration are relevant cues explaining this early differentiation. These results are relevant to determining the acoustic invariants associated with various sound categories (cf. Sect. 3.1). In addition, they showed that ambiguous sounds were associated with slower RTs 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 that slower RTs were associated with non-meaningful than meaningful sounds. Electrophysiological data showed that ambiguous sounds elicited more negative ERPs (a negative component, N280, followed by a negative slow wave, NSW) in fronto-central brain regions and less positive ERPs (P300 component) in parietal regions than typical sounds. This difference may reflect the difficulty to access information from long-term memory. Lastly, it is worth noting that no significant differences were observed on P100 and N100 components. These components are known to be sensitive to sound onset and temporal envelope, reflecting the fact that the categorization process occurs in the later sound processing stages. From the sonification point of view, this experiment suggests new ways of conveying information using sounds with different rates of congruence triggering different brain processing patterns.

2.2 Contextualized sounds: conceptual priming

The existence of sound semiotics is based on the human ability not only to identify isolated sounds but also to combine them in coherent ways. For instance, a comprehensible linguistic message is conveyed by associating words in keeping with the rules of syntax and grammar. Can similar links be generated between non-linguistic sounds so that any variations will change the global information conveyed? One of the major issues that arises here 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) established using a priming procedure that the amplitude of a negative ERP component, the N400 component, is larger when final words do not relate to the context of the previous sentence than otherwise (e.g., *The fish is swimming in the river/carpet*). The N400 has been widely used since that time to study semantic processing in language. The authors of recent studies used a priming procedure with non-linguistic stimuli such as pictures, odors, music and environmental sounds (see Aramaki et al. 2010b; Schön et al. 2010 for a review). Although the results of these experiments have mostly been interpreted as reflecting some kind of conceptual priming between words and non-linguistic 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 non-linguistic because of the implicit naming induced by the processing of the stimulus.

The aim of our first study on these lines (Schön et al. 2010) was to attempt to reduce as far as possible the likelihood that a labeling process of this kind takes place. To this end, we worked with a specific class of sounds called "abstract" sounds, which have the advantage of not being easily associated with an identifiable physical source (Merer et al. 2010). 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 and some methods of synthesis such as granular synthesis are particularly efficient means of creating abstract sounds. We then conducted conceptual priming tests using pairs of word/sound, and the level of congruence between the prime and the target was varied. In the first experiment, a written word (the prime) was presented visually before an abstract sound (the target), and subjects 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 subjects 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 and the presence of a more fronto-central distribution in response to word targets and a more centro-parietal distribution in response to sound targets.

Pursuing this topic farther in a subsequent study (Aramaki et al. 2010b), we sought to completely avoid the use of words as primes or targets. Conceptual priming was therefore studied using a homogeneous class of non-linguistic sounds, i.e., impact sounds (synthesized for the categorization experiment, cf. the Sect. 2.1), 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 non-linguistic sounds (Fig. 1). 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 non-linguistic and linguistic target sounds. From the sonification point of view, this study clearly means that it is possible to draw up a real language of sounds. In addition, it indicates that the sounds used for sonification processes do not have to be clearly identifiable, thus opening new possibilities for sonifying based on the use of abstract sounds. At this point, sonification processes join up with the rules of musical interpretation.

2.3 Musical context: timbre variations underlying interpretation

The fundamental studies presented in the previous section focused on the nature of sound semiotics, depending on whether the sounds are combined in a congruent or incongruent way. A well-known case where sounds are arranged in a specific way to transmit a message to listeners is that of music. Music expresses various emotions and moods that can transcend the possibilities of language in some cases. This scope for expression is based on the semiotic relationships between sounds and on the musicians' interpretation, which introduces variations between notes.

We here present a study that dealt with the acoustical factors characterizing the timbre variations liable to account for expressiveness in clarinet performances (Barthet et al. 2011). Timbre is defined as a perceptual attribute which makes it possible to discriminate between different sounds having the same pitch, loudness and duration (McAdams and Bigand 1993). 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 the timbre. A strong interaction between the expressive intentions and the timbre descriptors (attack time, spectral centroid, odd/even ratio) was observed in the case of both musical excerpts. The changes in the timbre descriptors were found to depend on the position of the notes in the musical phrases.

In a companion study (Barthet et al. 2010), the perceptual influence of some acoustical correlates of timbre (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 acoustical energy. Twenty skilled musicians were asked to select which version they preferred in a paired-comparison

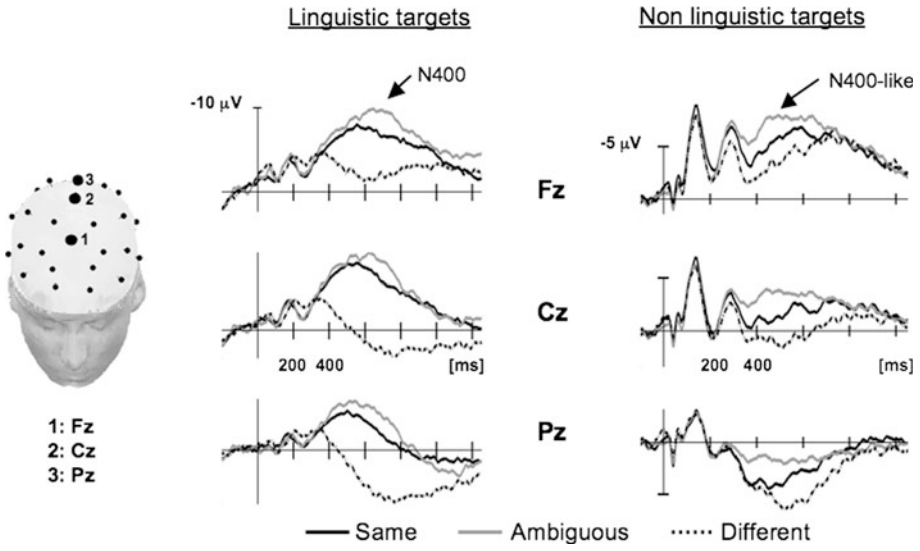

Fig. 1 Conceptual priming experiment: ERPs to same (*black line*), ambiguous (*gray line*) and different (*dashed line*) targets at selected midline electrodes (Fz, Cz, Pz) for linguistic (*left*) and non-linguistic (*right*) stimuli. (ERP traces taken from Aramaki et al. 2010b)

task. Results showed that the removal of the SC variations led to the greatest loss of musical expressiveness. These studies showed that in the case of clarinet playing, the musician consciously uses timbre variations and that these variations contribute greatly to the perceived quality of the performance. This suggests that the "meaning" of musical sounds is linked to their acoustical morphology, which depends strongly on the context (i.e., the previous and following notes, a fast or slow or musical tempo, etc.).

A question that is naturally raised by these findings is whether acoustical morphology also constitutes an important contribution to the significance of non-musical sounds such as environmental sounds and whether the morphology may mediate the mental representations elicited by sounds. To address this question, some signal properties specific to certain sound categories were studied in order to identify the invariant signal structures specific to these categories.

3 Searching for acoustical invariants: the analysis-by-synthesis method

As previously mentioned, identifying the invariant structures mediating the evocations induced by sounds is an important step toward the use of sounds in the form of sound metaphors to convey meaning. These invariants can be related either to so-called timbre properties or to the physics of the generating sources. However, when the sensation produced by sounds involves emotive aspects, specific protocols based on perceptual judgments are necessary.

In this section, we will first present some physical factors that mediate a certain number of evocations in the case of impact sounds. We will then describe the perceptual cues that are relevant in the case of moving sources, and describe a fundamental study in which abstract sounds were used to identify the signal morphologies responsible for conveying evoked dynamics. Lastly, we will present two examples of automobile industry applications, where it was proposed to identify the signal structures responsible for the suggestions of vehicle quality conveyed by car-door sounds and motor noise.

3.1 When physics can help

3.1.1 Environmental sounds: case of impact sounds

From the physical point of view, impact sounds are typically generated by an object undergoing free oscillations after being excited by an impact, or by a collision with other solid objects. In the simplest cases, the vibratory response of a system of this kind (viewed as a mass-spring-damper system) can be described by a set of linear partial differential equations. When the material changes, the wave propagation process is altered by the characteristics of the media. The 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).

Previous acoustic studies on the links between perception and the physical characteristics of sound sources have brought to light several important properties which can be used to identify the perceived effects of action on an object and the properties of the object itself (see Aramaki et al. 2009 for a review). In particular, frequency-dependent damping was found to be an important factor in the perception of sounds evoked by impacted materials. In the sound categorization experiments described in Sect. 2.1, in which electrophysiological and acoustical methods were combined (in a "neuro-acoustic" approach), we also established that the roughness is an important factor for distinguishing between the sound produced by metal versus glass and wood (Aramaki et al. 2009). The perceived hardness of a mallet striking a metallic object is predictable from the characteristics of the attack time. The shape of the impacted object determines the spectral content of the impact sound from the physical point of view. The frequencies of the spectral components correspond to the so-called eigenfrequencies, which are characteristics of the modes of the vibrating object and convey important perceptual information about the shape. Therefore, we assumed that the perceived shape depended on both the inharmonicity and the roughness. The perception of the size of the object is mainly correlated with the pitch: large objects generally vibrate at lower eigenfrequencies than small ones. In the case of quasi-harmonic sounds, we assume the pitch to be related to the frequency of the first spectral component, whereas complex sounds can elicit both spectral and virtual pitches (Terhardt et al. 1982).

3.1.2 Moving sound sources

The simulation of moving sources features importantly in many audio sound applications, including sonification and musical applications. Perception of moving sound sources obeys different processes from those mediating the localization of static sound events, and the evocation of sounds in motion can be achieved without any spatialization processes, providing relevant cues are taken into account. Based on previous studies (see, e.g., Väljamäe et al. 2005) and the references therein, Chowning 1971;

Kronland-Martinet and Voinier 2008), four important perceptual cues can be said to be the main components of a motion invariant.

Sound pressure: This parameter relates to the sound intensity and, in a more complex way, to the loudness. It varies inversely with the distance between the source and the listener. This rule is of great importance from the perceptual point of view (Rosenblum et al. 1987), and it is possibly decisive in the case of slowly moving sources. It is worth noting that only the relative changes in the sound pressure have to be taken into account in this context.

Timbre: Changes in the timbre of moving sound sources, which can be physically accounted for in terms of the air absorption, play an important perceptual role. Composers such as Maurice Ravel used cues of this kind in his Bolero in addition to intensity variations to make a realistic sensation of an incoming band: the orchestra starts in a low-frequency register to simulate the band playing far away, and the brightness gradually increases to give the impression that the musicians are getting closer. Schaeffer (1966) also used changes of timbre in a radiophonic context to simulate different speakers' positions in virtual space.

Doppler effect: As everyone has experienced while listening to the siren of a moving police car, moving sound sources induce changes in intensity as well as frequency shifts. Actually, depending on the relative speed of the source with respect to the listener, the frequency measured at the listener's position varies and the specific time-dependent pattern seems to be a highly relevant cue enabling the listener to construct a mental representation of the trajectory (Rosenblum et al. 1987).

Reverberation: The perceptual effects of reverberation cause distant sound sources to produce more highly reverberated signals than nearby sound sources because both the direct and reflected sound paths of distant sound sources are of similar orders of magnitude, whereas nearby sources produce direct sounds of greater magnitude than the reflected sounds. Moving sound sources therefore involve a time-dependent direct-to-reverberated ratio, the value of which depends on the distance between source and listener.

Figure 2 gives time-frequency representation of a moving sound source. Transformations (i.e., sound pressure, Doppler effect and timbre variations) were applied to an original recorded sound emitted by a static source. These sound transformations may be of great interest for sonification processes, since they make it possible to associate virtual particle motions with external data.

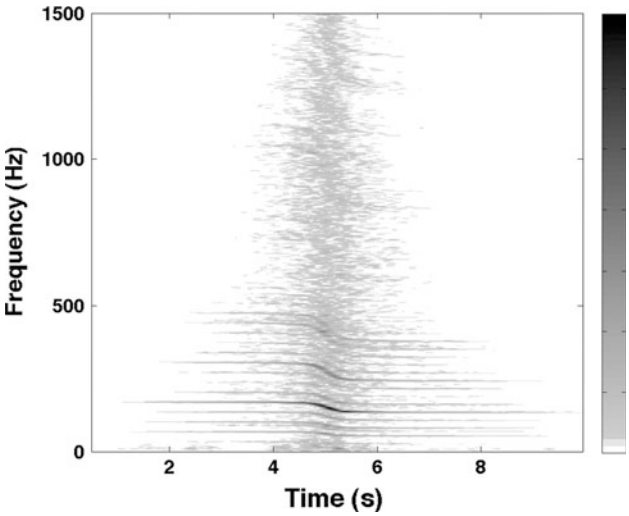

Fig. 2 Time-frequency representation of a moving sound source. The original sound was a recording of a concrete mixer producing a stationary, wide band noise, in which each effect of the sound processing can be assessed (i.e., sound pressure, Doppler effect and timbre variations). The sound source moved along a *straight line* at about 60 km/h and passed about 10 m away from the listener. This sound example is available on the webpage cited in Kronland-Martinet and Voinier (2008)

3.2 Beyond physics: case of the dynamics induced by sounds

To design a synthesizer that is able to generate specific motions via intuitive control devices, the signal morphologies conveying the moving sound have to be identified. As seen in the previous sections, physics can guide us to some extent but cannot always account alone for evocations of movement. The notion of perceived movement is vague and does not necessarily rely on the physical characteristics of the sound sources and/or the gestures involved. This is the case when listening to a falling object, for example. Before the object hits the ground, there is no noise although it is moving, and when it hits the floor, the resulting sound is an impact sound that does not necessarily evoke any movement.

The first step in this study consisted in choosing a sound corpus which could be used to identify the movement categories corresponding to various signal morphologies. The choice of the sound corpus to be used for this purpose was not easy, since we did not necessarily want people's judgments to be influenced by the source of the sound. For instance, if we listen to the sound of a clearly identified moving source such as a car or a motorcycle, the corresponding sound will be described as something that passes by, approaches or leaves, since we know that movements of this kind are typical of these sources. Since only a few movement categories would have been obtained if only easily recognizable sound sources had been used, we decided to use "abstract" sounds, the sources of which

cannot easily be determined Merer et al. (2010) (see Sect. 2.2 for more details).

The abstract sounds selected were used in a free categorization test to identify movement categories (Merer et al. 2008). This test consisted in presenting sounds which were arbitrarily displayed as small squares on a screen and asking subjects to group together the sounds that evoked similar movements. They could choose the number of groups freely and listen to the sounds as many times as they wanted. The analysis of the results led to five main movement categories: "turning", "falling", "approaching", "passing" and "rising". Interestingly, several subjects spontaneously made drawings to illustrate the categories they had created. This tends to show that a relationship between the dynamics of sounds and a graphic representation is intuitive. This result has important consequences for both sonification and control issues, as will be seen in Sect. 4.3. Signal features specific to each movement category were then extracted by performing signal analysis. Results systematically showed the presence of amplitude and frequency modulations in the case of sounds belonging to the category "turning", a logarithmic decrease in amplitude in that of sounds in the category "passing" and impulse characteristics in that of sounds in the category "falling".

The identification of the signal morphologies evoking movement is still in progress, but the preliminary results support the involvement of acoustical invariants in addition to physical considerations.

3.3 Industrial sounds

In the automobile industry, some sounds are known to affect people's appreciation of a vehicle. The sound generated by a car motor can influence the driver's impression of power and speediness. More surprisingly, studies have shown that even car-door noise can influence potential buyers' impression of quality and solidity. By linking the clients' judgments to the signal properties of these sounds, some perceptual cues mediating the notion of car quality were extracted from the signal. Both of these aspects were studied in the framework of a collaboration with the company PSA (Peugeot Citroën Automobiles). This study can be viewed as the analysis of a sonification process in ecological context. Motor sounds are actually the natural expression of the engine's behavior, and door sounds can reveal some hidden aspects of the car's structure. Analyzing such sonification processes can therefore highlight relevant concepts which can then be applied to virtual situations.

3.3.1 Door-closure sounds

The aim of this study was to analyze customers' perception of the sound of car doors closing and to use their judgments to improve these sounds by adapting the dimensions of the mechanical parts of the car door appropriately. For this purpose, car-door-closure sounds recorded with different makes and categories of cars were presented to subjects who were asked to judge the quality of the vehicle from the sounds they heard. An analysis–synthesis approach was then adopted. The signal was decomposed into two parts by performing empirical mode decomposition (EMD) (Huang et al. 1998): a low-frequency contribution from the impact of the door itself and a high-frequency contribution involving the latch mechanism were defined. Analysis of each contribution showed that the latch mechanism could be characterized by three impacts, while a single impact seemed to suffice to characterize the door impact. 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 elapsing between the different impacts, car-door-closure sounds corresponding to different qualities of vehicle 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 as well as the time elapsing between the four impacts mediated the perception of solidity and quality of the car (Bezat et al. 2006, 2007). This study confirms that signal invariants evoking the solidity and quality of a car can be identified.

3.3.2 Motor sounds

The aim of the next study was to determine the influence of the dynamic evolution of the noise of a car engine on subjects' perception of the quality and the controllability of the vehicle. The noise perceived inside a vehicle results from interactions between three main sound sources: engine noise, aerodynamic noise and tire/road noise. Depending on the speed of the car, each of these three contributions are variably predominant and masking phenomena also occur between these different sound sources. To obtain a representation of the signal that reflects the perceived sound, an auditory model was used, which focused on the most perceptually important parts of the motor noise (Pressnitzer and Gnansia 2005). Among the most perceptually relevant signal parameters thus identified, a time-evolving formant structure as well as roughness factors was identified (Sciabica et al. 2010). The formant structure, which depends strongly on the structure of the passenger cell, is now being used to define a high-level control (linked to the sportiness and the quality of the vehicle) of engine noise using an additive synthesis model. Results obtained by performing listening tests on the synthesized sounds are being analyzed, and the formant level ranging around 800 Hz was found to have a strong

influence on the sensation of sportiness of a car. These studies address some sonification issues in the context of the automobile industry and provide valuable information about designing new sounds for electric and hybrid motorcars.

4 Control of synthesis processes

Based on the identification of invariant signal structures specific to given sound categories, control processes mediating various perceptual evocations (i.e., types of material, perceived movement, quality, etc.) could possibly be designed for sonification purposes.

In this section, we present several synthesis tools that we have developed for generating and controlling sounds. These synthesis models make it possible to relevantly resynthesize natural sounds. At signal level, the sound generation problem required processing hundreds of parameters and the procedure was therefore only intended for experts. In spite of the efficiency of these models, the control issue, and the so-called mapping strategy, is an important aspect that has to be taken into account when constructing a synthesizer. In addition, this aspect is of fundamental importance when using the synthesizer for sonification purposes.

In practice, we adopted hierarchical levels of control making it possible to route and dispatch the parameters from an intuitive level to the signal and algorithmic level. As the parameters that allow intuitive controls are not independent and might be linked to several signal characteristics at a time, the mapping between levels was far from being straightforward. The real-time implementation of all the synthesis tools was carried out with MaxMSP[1].

4.1 Immersive auditory scenes

Nowadays, interactive 3-D environments tend to include both synthesis and spatialization processes so as to increase the realism and the feeling of being immersed in virtual scenes. In this context, we designed a real-time spatialized synthesizer to generate 3-D immersive auditory scenes, which were intended to be used in the framework of interactive virtual reality and sonification applications. The system simulates various environmental sound sources as defined by the taxonomy proposed by W.W. Gaver (i.e., vibrating solids, liquids, aerodynamics ; cf. the Sect. 2). The synthesis engine was based on an efficient combination of additive synthesis using a frame-by-frame approach and 3-D positional audio modules (Verron et al. 2010). Sound synthesis and spatialization were implemented at the same level of sound generation, contrary to what occurs with the classical two-stage approach, which consists in first synthesizing a monophonic sound (generation of the intrinsic timbre properties) and then spatializing the sound (defining its spatial position in the environment).

The new architecture yielded control strategies based on the overall manipulation of the timbre and spatial attributes (position, perceived width) of the sound sources. Concerning the timbre, environmental sounds include a wide range of sounds but interestingly, their acoustic morphology shows the existence of common characteristics which call for a granular-like synthesis process. Therefore, environmental sounds were designed using a suitable set of elementary signals based on impact, chirp or noise structures and spatialized by positioning them in 3-D space (Verron et al. 2009). The relevance and sufficiency of these "grains" were tested empirically by analyzing and resynthesizing various environmental sounds in each category. This atomic dictionary can be completed in the future without detracting from the method used here. Complex 3-D auditory scenes (e.g., rainy or windy weather sounds) could even be intuitively designed by appropriately combining spatialized environmental sources. In practice, these controls can be achieved either using MIDI interfaces in an interactive way or automatically from data obtained using a video game engine or other external data sources.

4.2 Sounds produced by interacting objects

We have also developed synthesis tools for simulating interactions between objects. We focus here on the case of an impact sound synthesizer for which we developed an intuitive control strategy based on a three-level architecture (Aramaki et al. 2010a). The top layer was composed of verbal descriptions of the object (nature of the material, size and shape, etc.) and of the excitation (input force, hardness of the mallet, position of the excitation, etc.). The middle layer was based on sound descriptors that are known to be relevant from the perceptual point of view. The bottom layer consisted of parameters of the synthesis model. The following two mapping strategies were implemented between the layers.

The first mapping strategy focused on the relationships between verbal descriptions of the sound source and sound descriptors based on the acoustical invariants presented in the Sect. 3.1. For example, the control of the perceived material involved the control of the damping processes but also that of spectral sound descriptors such as the inharmonicity and the roughness. The perceived size of the object was directly linked to the fundamental frequency of the sound and the perceived shape to the control of the inharmonicity along with the pitch. The perceived hardness of the mallet was controlled by the attack time and the

[1] Cycling '74, http://www.cycling74.com/downloads/.

brightness and the perceived force by the brightness, so that the heavier the force applied was, the brighter the sound became. The excitation point, which strongly influences the amplitude of the components by causing envelope modulations in the spectrum, was also taken into account by shaping the spectrum with a feedforward comb filter.

The second mapping strategy focused on the relationships between sound descriptors and synthesis parameters. Since the damping was frequency dependent, we defined a damping law expressed as an exponential function so that the control of the damping was reduced to two parameters: an overall damping and a frequency-dependent damping. The choice of an exponential function enables us to efficiently simulate various damping profiles characteristic of different materials by acting on a few control parameters. The possibility of readjusting the distribution of the spectral components as required opens the way to many possible strategies. We have suggested a means of controlling the inharmonicity that allows users to adapt the spectral relationships between the initial harmonic components using an inharmonicity law involving only a few parameters. Some pre-defined presets give direct access to typical inharmonicity profiles, such as those of membranes and plates. The roughness was controlled separately for each Bark band based on amplitude and frequency modulations. Synthesizers of this kind should provide valuable tools for sonification purposes, when associating materials with sound data is relevant to the context.

4.3 Movements induced by sounds

As described previously, impact sounds belong to a class of sounds that is generally well identified, and we presented a control strategy based on physical and perceptual information. The dynamics of moving sounds are a more intricate problem, and defining suitable parameters for controlling the dynamic aspects requires a special approach. In the free categorization test described in Sect. 3.2, five categories of movement were identified, along with suitable signal invariants corresponding to each category. However, this test did not directly yield any perceptual cues as to how these evocations might be controlled in a synthesis tool. Therefore, to identify perceptually relevant control parameters corresponding to the dynamic patterns evoked, further experiments were conducted in which subjects were asked to show the trajectory evoked by sounds by drawing them with a graphical interface on a computer, using the six parameters available linked to the shape of the trajectory (shape, size, frequency oscillation, randomness, angle and initial position) and three parameters linked to the dynamics (initial and final velocity and number of returns). This idea was based on the fact that in the previous (free categorization) test, some of the subjects spontaneously made drawings to describe the movements of the sounds they had perceived.

Results showed that although the subjects used various drawing strategies, equivalent drawings and common parameter values could still be discerned. As far as the shape was concerned, subjects' drawings showed good agreement on the distinction between linear and oscillating movements and between wave-like and circular oscillations. This means that these three aspects give a sufficiently exact control of the perceived shape of sound trajectories. As far as the orientation of the trajectory was concerned, only the distinction between horizontal and vertical seems to be relevant. While there was agreement among subjects about the distinction between the upward and downward directions, the difference between the left and right directions was not relevant. As far as the velocity was concerned, the subjects distinguished between constant and varying velocities, but they did not show good agreement in the way they specified the velocity variations they perceived. Based on these findings, a high-level dynamic control strategy was developed to control sound trajectories using labels (Fig. 3).

5 Toward sound metaphors for sonification processes

Electrophysiological measurements have confirmed the existence of a semiotics of isolated sounds, thus suggesting

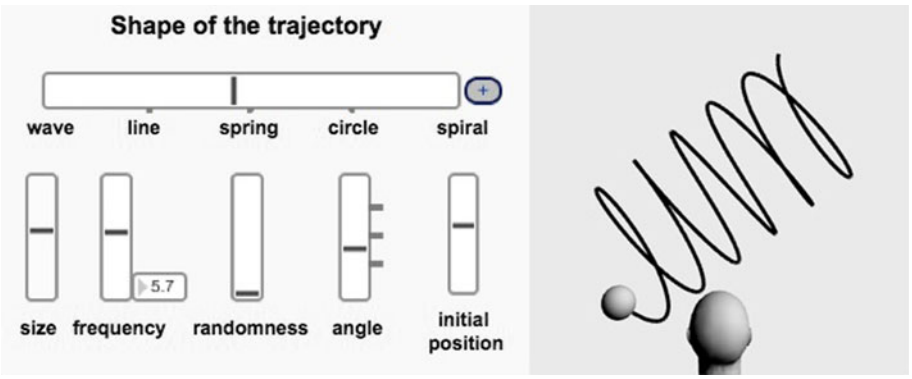

Fig. 3 User interface for controlling evoked sound trajectories. The high-level control parameters (shape, size, etc.) are defined in the *left window* and the corresponding trajectory of the moving sound source (i.e., the *gray sphere*) is shown in the *right window*

that a language of sounds might be drawn up based on the timbre attribute. Analysis of the associations between sounds in musical contexts (during clarinet performances) has shown that the context has important effects on the acoustic morphologies of the notes, which contribute importantly to the "significance" of musical sounds. Based on these findings, the existence of sound structures (which are called invariants) allowing the attribution of significance to sounds has been hypothesized.

The first step to developing a general tool for sonification processes by providing synthesis models with means of high-level control consisted in identifying invariant signal structures. These structures have both physical and perceptual sound properties. Variations of physical properties such as dispersion and dissipation make perceptual distinctions possible between different types of objects (i.e., strings versus bars, plates versus membranes, etc.) or materials. The spectral content of the impact sound, in particular the eigenfrequencies that characterize the modes of a vibrating object, is responsible for the perception of its shape and size. The perception of moving sources (independently of the spatialization process) can also be explained by physical considerations involving air absorption and source velocity inducing loudness, brightness and reverberation variations, as well as frequency fluctuations due to the Doppler effect. In other cases, properties that cannot be explained entirely in physical terms were determined by carrying out listening tests: the time intervals and energy variations between the various impacts composing a door-closure sound are responsible for the listeners' perception of the quality of the vehicle.

Based on the invariant signal structures identified, various control strategies have been developed for designing sounds that can be controlled by the external data to be sonified. The real-time synthesis platform designed and constructed makes it possible to intuitively control interacting objects and immersive 3-D environments. With these interfaces, complex 3-D auditory scenes (the sound of rain, waves, wind, fire, etc.) can be intuitively designed as well as sounds involving various types of material, objects of various sizes and shapes and their interactions with the environment. New means of controlling the dynamics of moving sounds via written words or drawings are currently being added to the platform.

These developments open the way to new and captivating possibilities for using non-linguistic sounds as a means of communication. Further extending our knowledge in this field will make it possible to develop new tools for generating sound metaphors based on invariant signal structures which can be used to evoke specific mental images via selected perceptual and cognitive attributes. These metaphors will be obtained by shaping initially inert sound textures using intuitive (high-level) control methods designed on the lines described above. The development of these innovative tools will be accomplished thanks to the METASON project (ANR) that will start in November 2010.

References

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2009) Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad S, Kronland-Martinet R, Jensen K (eds) Computer music modeling and retrieval—genesis of meaning of sound and music, *LNCS*, vol 5493. pp 1–17 Springer, Berlin

Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2011) Controlling the perceived material in an impact sound synthesizer. IEEE Trans Audio Speech Lang Process 19(2):301–314

Aramaki M, Gondre C, Kronland-Martinet R, Voinier T, Ystad S (2010a) Imagine the sounds: an intuitive control of an impact sound synthesizer. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen J (eds) Auditory display, lecture notes in computer science, vol 5954. pp 408–421 Springer, Berlin

Aramaki M, Marie C, Kronland-Martinet R, Ystad S, Besson M (2010b) Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. J Cogn Neurosci 22(11):2555–2569

Ballas JA (1993) Common factors in the identification of an assortment of brief everyday sounds. J Exp Psychol Hum Percept Perform 19(2):250–267

Barthet M, Depalle P, Kronland-Martinet R, Ystad S (2010) Acoustical correlates of timbre and expressiveness in clarinet performance. Music Percept Interdiscip J 28(2):135–154

Barthet M, Depalle P, Kronland-Martinet R, Ystad S (2011) Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. Music Percept Interdiscip J 28(3):265–278

Bezat MC, Roussarie V, Kronland-Martinet R, Ystad S, McAdams S (2006) Perceptual analyses of action-related impact sounds. In: Proceedings of the 6th European conference on noise control euronoise 2006. Tampere, Finland

Bezat MC, Roussarie V, Voinier T, Kronland-Martinet R, Ystad S (2007) Car door closure sounds: characterization of perceptual properties through analysis-synthesis approach. In: Proceedings of the 19th international congress on acoustics. Madrid, Spain

Chowning JM (1971) The simulation of moving sound sources. J Audio Eng Soc 19(1):2–6

Gaver WW (1993a) How do we hear in the world? Explorations of ecological acoustics. Ecol Psychol 5(4):285–313

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

Gibson JJ (1986) The ecological approach to visual perception. Lawrence Erlbaum Associates, Hillsdale

Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. Percept Psychophys 69(6):839–855

Gygi B, Shafiro V (2007) General functions and specific applications of environmental sound research. Front Biosci 12:3152–3166

Huang NE, Shen Z, Long S, Wu M, Shih H, Zheng Q, Yen N, Tung C, Liu HH (1998) The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Roy Soc Lond A 454:903–995

Kronland-Martinet R, Voinier T (2008) Real-time perceptual simulation of moving sources: application to the leslie cabinet and 3d sound immersion. EURASIP J Audio Speech Music Process. doi:10.1155/2008/849696

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Sci Agric 207:203–204

McAdams S, Bigand E (1993) Thinking in sound: the cognitive psychology of human audition. Oxford University Press, Oxford

Merer A, Ystad S, Kronland-Martinet R, Aramaki M (2008) Semiotics of sounds evoking motions: categorization and acoustic features. In: Kronland-Martinet R, Ystad S, Jensen K (eds) Computer music modeling and retrieval—sense of sounds, lecture notes in computer science, vol 4969. pp 139–158 Springer, Berlin

Merer A, Ystad S, Kronland-Martinet R, Aramaki M (2010) On the potentiality of abstract sounds in perception research. In: Proceedings of the 7th international symposium on computer music modeling and retrieval, CMMR 2010—music is in the sound

Pressnitzer D, Gnansia D (2005) Real time auditory model. In: Proceedings of international computer music conference. Barcelona, Spain

Rosenblum LD, Carello C, Pastore RE (1987) Relative effectiveness of three stimulus variables for locating a moving sound source. Perception 16(2):175–186

Rugg MD, Coles MGH (1995) Electrophysiology of mind. Event-related brain potentials and cognition, chap. The ERP and cognitive psychology: conceptual issues, No. 25 in Oxford psychology. Oxford University Press, Oxford, pp 27–39

Schaeffer P (1966) Traité des objets musicaux. Ed. du Seuil

Schön D, Ystad S, Kronland-Martinet R, Besson M (2010) The evocative power of sounds: conceptual priming between words and nonverbal sounds. J Cogn Neurosci 22(5):1026–1035

Sciabica JF, Bezat MC, Roussarie V, Kronland–Martinet R, Ystad S (2010) Towards timbre modeling of sounds inside accelerating cars. In: Ystad S, Aramaki M, Kronland-Martinet R, Jensen J (eds) Auditory display, lecture notes in computer science, vol 5954. pp 377–392 Springer, Berlin

Terhardt E, Stoll G, Seewann M (1982) Pitch of complex signals according to virtual-pitch theory: tests, examples, and predictions. J Acoust Soc Am 71:671–678

Väljamäe A, Larsson P, Västfjäll D, Kleiner M (2005) Travelling without moving: auditory scene cues for translational self-motion. In: Proceedings of the 11th international conference on auditory display (ICAD '05)

Vanderveer NJ (1979) Ecological acoustics: human perception of environmental sounds. Ph.D. thesis, Georgia Institute of Technology

Verron C, Aramaki M, Kronland-Martinet R, Pallone G (2010) A 3d immersive synthesizer for environmental sounds. IEEE Trans Audio Speech Lang Process 18(6):1550–1561

Verron C, Pallone G, Aramaki M, Kronland-Martinet R (2009) Controlling a spatialized environmental sound synthesizer. In: Proceedings of the IEEE workshop on applications of signal processing to audio and acoustics (WASPAA). New Paltz, NY. 18–21 October 2009, pp 321–324