Perceptual Control of Environmental Sound Synthesis

Mitsuko Aramaki, Richard Kronland-Martinet, and Sølvi Ystad

Laboratoire de Mécanique et d'Acoustique, 31,
31, Chemin Joseph Aiguier
13402 Marseille Cedex 20
`{aramaki,kronland,ystad}@lma.cnrs-mrs.fr`

Abstract. In this article we explain how perceptual control of synthesis processes can be achieved through a multidisciplinary approach relating physical and signal properties of sound sources to evocations induced by sounds. This approach is applied to environmental and abstract sounds in 3 different experiments. In the first experiment a perceptual control of synthesized impact sounds evoking sound sources of different materials and shapes is presented. The second experiment describes an immersive environmental synthesizer simulating different kinds of environmental sounds evoking natural events such as rain, waves, wind and fire. In the last example motion evoked by abstract sounds is investigated. A tool for describing perceived motion through drawings is proposed in this case.

Keywords: perceptual control, synthesis, analysis, acoustic descriptors, environmental sounds, abstract sounds

1   Introduction

The development and the optimization of synthesis models have been important research issues since computers produced the first sounds in the early sixties [20]. As computers became increasingly powerful, real-time implementation of synthesis became possible and new research fields related to the control and the development of digital musical instruments appeared. One of the main challenges linked to such fields is the mapping between the control parameters of the interface and the synthesis parameters. Certain synthesis algorithms such as additive synthesis [17, 16, 4] allow for a very precise reconstruction of sounds, but contain a large number of parameters (several hundreds in the case of piano synthesis), which makes the mapping between the control device and the synthesis model complicated. Other synthesis approaches such as global or non-linear approaches (e.g. frequency modulation (FM) or waveshaping [9, 7]) are easier to implement and to control since they contain fewer synthesis parameters, but do not allow for a precise resynthesis. This means that the control device cannot be dissociated from the synthesis model when conceiving a digital musical instrument and even more, a genuine musical interface should go past the technical stage to integrate the creative thought [14]. So far, a large number of control

devices have been developed for musical purposes [25], but only a few are being actively used in musical contexts either because the control is not sufficiently well adapted to performance situations or because they do not offer an adequate sound control. This means that the control of digital musical instruments is still an issue that necessitates more investigations. Nowadays, sounds are used in a large number of applications (e.g. car industry, video games, radio, cinema, medicine, tourism, ...) since new research domains where sounds are investigated to inform or guide persons (e.g. auditory display, sound design, virtual reality, ...) have developed. Researchers within these domains have traditionally made use of prerecorded sounds, but since important progress has been achieved concerning the development of efficient and realistic synthesis models, an increasing interest for synthesis solutions has lately been observed [8, 29, 33]. The control requirements related to such applications differ from musical control devices since the role of the sounds in this case is to provide specific information to the end users. Hence, a perceptual control that makes it possible to control sounds from semantic labels, gestures or drawings would be of great interest for such applications. Such control implies that perceptual and cognitive aspects are taken into account in order to understand how a sound is perceived and interpreted. Why are we for instance able to recognize the material of falling objects simply from the sounds they produce, or why do we easily accept the ersatz of horse hooves made by the noise produced when somebody is knocking coconuts together? Previous studies [6, 28] have shown that the processing of both linguistic and non-linguistic target sounds in conceptual priming tests elicited similar relationships in the congruity processing. These results indicate that it should be possible to draw up a real *language of sounds*. A certain number of questions have to be answered before such a language can be defined, in particular whether the identification of a sound event through the signal is linked to the presence of specific acoustic morphologies, so-called invariants that can be identified from signal analyses [22]. If so, the identification of signal invariants should make it possible to propose a perceptual control of sound synthesis processes that enables a direct evocative control.

To develop perceptual control strategies of synthesis processes, it is in the first place necessary to understand the perceptual relevance of the sound attributes that characterize the sound category that is investigated. The sound attributes can be of different types and can either be linked to the physical behavior of the source [13], to the signal parameters [18] or to timbre descriptors obtained from perceptual considerations [21]. In this paper we focus on the perceptual control of environmental sounds and evoked motion by describing how such control can be defined from the identification of signal invariants obtained both from the considerations of physical behavior of the sound generating sources and the perceptual impact of the sounds on the listeners. The general approach proposed to obtain perceptual control strategies is shown in Figure 1.

 In the first section of this article, we describe how the perceptual control of an impact sound synthesizer enabling the definition of the sound source through verbal labels can be defined. Then a tool for controlling 3D environmental im-
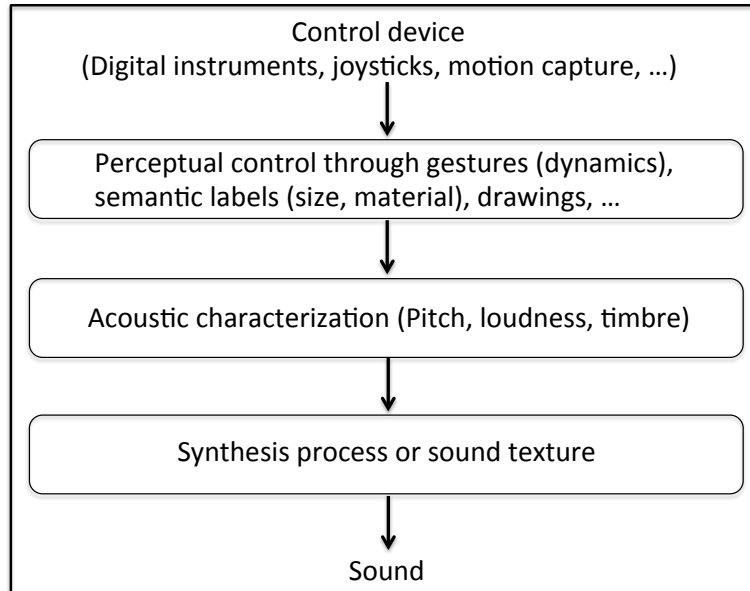
Fig. 1. Synoptics of the perceptual control strategy.

mersive auditory scenes with verbal labels based on a synthesizer adapted to environmental sounds is described. Finally an investigation on perceived motion and how intuitive control parameters for this specific type of evocation can be defined is presented.

2 Impact sound synthesizer

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. These vibrations are governed by a wave equation and the natural frequencies of the system are obtained from the solution of this equation. These natural frequencies correspond to the frequencies for which the objet is capable of undergoing harmonic motion. The wave propagation depends on the characteristics of the object that influences two physical phenomena, i.e. dispersion (due to the stiffness of the material) and dissipation (due to loss mechanisms). Dispersion results from the fact that the wave propagation speed varies with the frequency and introduces inharmonicity in the spectrum. Dissipation is directly linked to the damping of the sound and is generally frequency-dependent. The perceptual relevance of these phenomena and how they contribute to the identification of impact sounds will be discussed in the next section.

3

2.1 Invariant sound structures characterizing impact sounds

Impact sounds have been largely investigated in the literature. In particular, some links between the physical characteristics of actions (impact, bouncing . . .) and sound sources (material, shape, size, cavity . . .) and their perceptual correlates were established (see [2, 1] for a review). For instance, the perception of the hardness of a mallet impacting an object is related to the characteristics of the attack time. The perception of material seems to be linked to the characteristics of the damping that is generally frequency-dependent: high frequency components are damped more heavily than low frequency components. In addition to the damping, we concluded that the density of spectral components which is directly linked to the perceived roughness, is also relevant for the distinction between metal versus glass and wood categories [2, 1]. The perceived shape of the object is related to the distribution of the spectral components of the produced sound. It can therefore be assumed that both the inharmonicity and the roughness determine the perceived shape of the object. From a physical point of view, large objects vibrate at lower eigenfrequencies than small ones. Hence, the perceived size of the object is mainly based on the pitch. For complex sounds, the determination of pitch is still an open issue. In some cases, the pitch may not correspond to an actual component of the spectrum and both spectral and virtual pitches are elicited [30]. However, for quasi-harmonic sounds, we assume that the pitch is linked to the fundamental frequency.

These considerations allowed us to identify signal morphologies (i.e. invariants) conveying relevant information on the perceived material, size, shape and type of impact on an object. A mapping strategy defining a link between synthesis parameters, acoustic descriptors and perceptual control parameters can then be defined, as described in the next section.

2.2 Control of the impact sound synthesizer

To develop a perceptual control of impact sounds based on semantic description of the sound source, a mapping strategy between synthesis parameters (low level), acoustic descriptors (middle level) and semantic labels (high level) characterizing the evoked sound object was defined. The mapping strategy that was chosen is based on a three level architecture as seen in Figure 2) [5, 3, 2]. The top layer is composed of verbal descriptions of the object (nature of the material, size and shape, etc.). The middle layer concerns the control of acoustic descriptors that are known to be relevant from the perceptual point of view as described in section 2.1. The bottom layer is dedicated to the control of the parameters of the synthesis model (amplitudes, frequencies and damping coefficients of components). The mapping strategy between verbal descriptions of the sound source and sound descriptors is designed with respect to the previous considerations described in section 2.1. The control of the perceived material is based on the manipulation of damping but also that of spectral sound descriptors such as inharmonicity and roughness. Since the damping is frequency dependent, a damping law was arbitrarily defined and we proposed an exponential function:
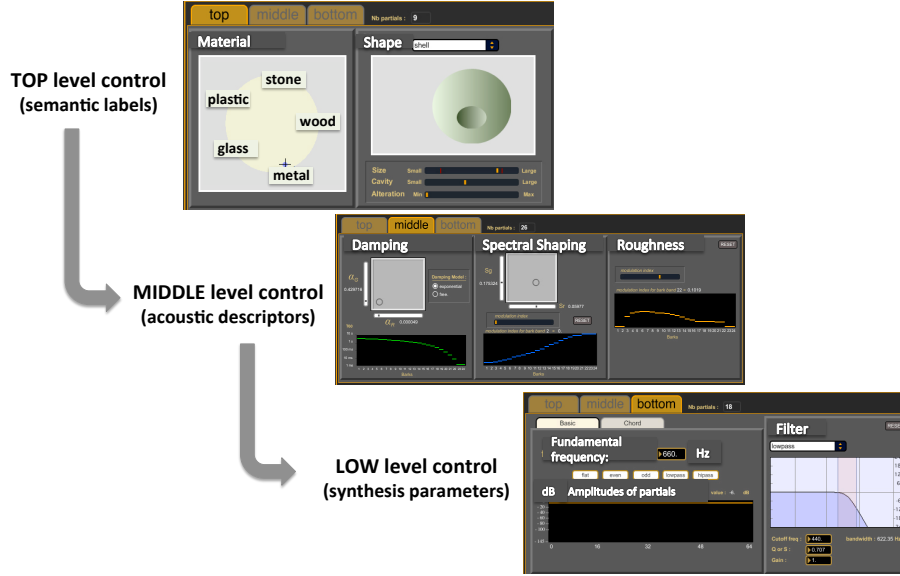
Fig. 2. Three level control strategy of impact sounds

$\alpha(\omega) = e^{(\alpha_G + \alpha_R \omega)}$ characterized by two parameters: a global damping α_G and a relative damping α_R. The choice of an exponential function enables us to reach various damping profiles characteristic of physical materials by acting on a few control parameters. Hence, the control of the damping was effectuated by two parameters. The perception of size is controlled by the frequency of the first component and the perception of shape by the spectral distribution of components defined from inharmonicity and roughness. As for damping, an inharmonicity law characterized by a few parameters was proposed. Some pre-defined presets give direct access to typical inharmonicity profiles, such as those of strings, membranes and plates. The roughness is created by applying amplitude and frequency modulations on the initial sound and can be controlled separately for each Bark band.

The mapping between sound descriptors and synthesis parameters is organized as follows. The damping coefficients of the components are determined from the damping law $\alpha(\omega)$ and their amplitudes from the envelope modulations introduced by the excitation point. The spectral distribution of components (frequency values) are defined from the inharmonicity law and the roughness. A direct control at low level allows for readjustments of this spectral distribution of components by acting separately on the frequency, amplitude and damping coefficients of each component. This mapping between middle and bottom layer

5

depends on the synthesis model and should be adapted with respect to the chosen
synthesis process.

What the control of action is concerned, the hardness of the mallet is con-
trolled by the attack time and the brightness while the perceived force is related
to the brightness: the heavier the applied force is, the brighter the sound. The
timbre of the generated sound is also strongly influenced by the excitation point
of the impact that creates envelope modulations in the spectrum due to the can-
cellation of modes presenting a node at the point of excitation. From a synthesis
point of view, the location of the impact is taken into account by shaping the
spectrum with a feed forward comb filter.

3   Immersive environmental sound synthesizer

Impact sounds constitute a specific category of environmental sounds. In this sec-
tion an immersive synthesizer simulating various kinds of environmental sounds
is proposed. These sounds are divided in three main categories according to
W. W. Gaver's taxonomy of everyday sound sources: vibrating solids (impact,
bouncing, deformation ...) liquid (wave, drop, rain ...) and aerodynamic (wind,
fire ...) objects [12, 11]. Based on this taxonomy, we proposed a synthesizer to
create and control immersive environmental scenes intended for interactive vir-
tual and augmented reality and sonification applications. Both synthesis and
spatialization engines were included in this tool so as to increase the realism and
the feeling of being immersed in virtual worlds.

3.1   Invariant sound structures characterizing environmental sounds

In the case of impact sounds, we have seen that physical considerations reveal
important properties that can be used to identify the perceived effects of the
generated sounds (cf. section 2.1). For other types of environmental sounds such
as wave, wind or explosion sounds, the physical considerations involve com-
plex modeling and can less easily be taken into account for synthesis perspective
with interactive constraints. Hence the identification of perceptual cues linked to
these sound categories was done by the analyses of sound signals representative
of these categories. From a perceptual point of view, these sounds evoke a wide
range of different physical sources, but interestingly, from a signal point of view,
some common acoustic morphologies can be highlighted across these sounds. To
date, we concluded on five elementary sound morphologies based on impacts,
chirps and noise structures [32]. This finding is based on a heuristic approach
that has been verified on a large set of environmental sounds. Actually, granular
synthesis processes associated to the morphologies of these five grains have en-
abled the generation of various environmental sounds such as solid interactions
and aerodynamic or liquid sounds. Sounds produced by solid interactions can be
characterised from a physical point of view. When a linear approximation applies
(small deformation of the structure), the response of a solid object to external
forces can be viewed as the convolution of these forces with the modal response

of the object. Such a response is given by a sum of exponentially damped sinusoids, defining the typical "tonal solid grain". Nevertheless, such a type of grain cannot itself account for all kinds of solid impact sounds. Actually, rapidly vanishing impact sounds or sounds characterized by a strong density of modes may rather be modelled as exponentially damped noise. This sound characterization stands for both perceptual and signal points of views, since no obvious pitch can be extracted from such sounds. Exponentially damped noise constitutes the so-called "noisy impact grain". Still dealing with physical considerations, we may design a "liquid grain" that takes into account cavitation phenomena occurring in liquid motion. Cavitation leads to local pressure variations that, from an acoustic point of view, generate time varying frequency components such as exponentially damped linear chirps. Exponentially damped chirps then constitute our third type of grain: the "liquid grain". Aerodynamic sounds generally result from complicated interactions between solids and gases. It is therefore difficult to extract useful information from corresponding physical models. The construction of granular synthesis processes was therefore based on heuristic perceptual expertise defining two kinds of aerodynamic grains: the "whistling grain" consisting in a slowly varying narrow band noise; and the "background aerodynamic grain" consisting in a broadband filtered noise. By combining these five grains using an accurate statistics of appearance, various environmental sounds can be designed such as rainy ambiances, seacoast ambiances, windy environments, fire noises, or solid interactions simulating solid impacts or footstep noises. We currently aim at extracting the parameters corresponding to these grains from the analysis of natural sound, using matching pursuit like methods.

3.2   Control of the environmental sound synthesizer

To develop a perceptual control of the environmental sound synthesizer based on semantic labels, a mapping strategy that enabled the design of complex auditory scenes was defined. In particular, we took into account that some sound sources such as wind or rain are naturally diffuse and wide. Therefore, the control included the location and the spatial extension of sound sources in a 3D space. In contrast with the classical two-stage approach, which consists in first synthesizing a monophonic sound (timbre properties) and then spatializing the sound (spatial position and extension in a 3D space), the architecture of the proposed synthesizer yielded control strategies based on the overall manipulation of timbre and spatial attributes of sound sources at the same level of sound generation [31].

For that purpose, we decided to bring the spatial distribution of the sounds to the lowest level of the sound generation. Indeed, the characterization of each elementary time-localized sound component, that is generally limited to its amplitude, frequency and phase, was augmented by its spatial position in the 3D space. This tremendous addition leads to an increasing number of control possibilities while still being real time compatible thanks to an accurate use of the granular synthesis process in the frequency domain [34]. We then showed that

the control of the spatial distribution of the partials together with the construction of decorrelated versions of the actual sound allowed for the control of the spatial position of the sound source together with the control of its perceived spatial width. These two perceptual spatial dimensions have shown to be of great importance in the design of immersive auditory scenes.

Complex 3D auditory scenes can be intuitively built by combining spatialized sound sources that are themselves built from the elementary "grain" structures (cf. section 3.1).
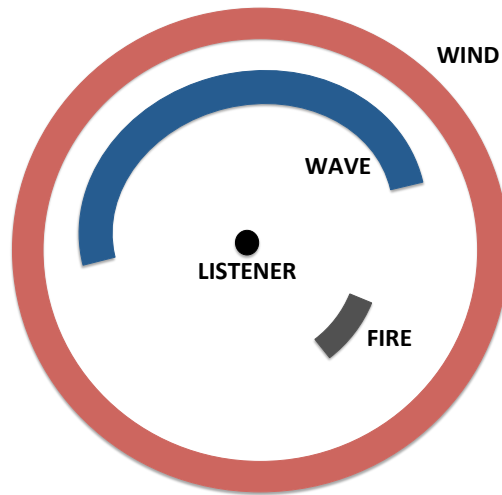

Fig. 3. Auditory scene of a windy day (wind source surrounding the listener) on a beach (wave coming towards the listener) and including a BBQ sound (fire located at the back right of the listener).

The fire is for instance built from three elementary grains that are a whistling grain (simulating the hissing), a background aerodynamic grain (simulating the background combustion) and noisy impact grains (simulating the cracklings). The grains are generated and launched randomly with respect to time using an accurate statistical law that can be controlled. A global control of the fire intensity, mapped with the control of the grain generation (amplitude and statistical law), can then be designed. The overall control of the environmental scene synthesizer is effectuated through a graphical interface (see figure 3) where the listener is positioned in the center of the scene. Then the user selects the sound sources to be included in the auditory scene among a set of available sources (fire, wind, rain, wave, chimes, footsteps ...) and places them around the lis-

tener by graphically defining the distance and the spatial width of the source. In
cases of interactive uses, controls can be achieved using either MIDI interfaces,
from data obtained from a graphical engine or other external data sources.

4 Synthesis of evoked motion

A third approach aiming at developing perceptual control devices for synthesized
sounds that evoke specific motions is presented in this section. The definition of
perceptual control necessitates more thorough investigations in this case than
in the two previous cases due to the rather vague notion of perceived motion.
Although physics of moving sound sources can to some extent give indications on
certain morphologies that characterize specific movements [19], it cannot always
explain the notion of perceived motion. In fact, this notion does not only rely
on the physical displacement of an object, but can also be linked to temporal
evolutions in general or to motion at a more metaphoric level. It is therefore
necessary to improve the understanding of perceived dimension of motion linked
to the intrinsic properties of sounds. Therefore, an investigation of perceived
motion categories obtained through listening tests was effectuated before signal
morphologies that characterize the perceptual recognition of motion could be
identified.

4.1 Invariant structures of evoked motion

As already mentioned, motion can be directly linked to physical moving sound
sources, but can also be considered in more metaphoric ways. Studies on the
physical movement of a sound source and the corresponding signal morphologies
have been widely described in the literature [10, 27, 26, 35, 19]. One aspect that
links physics and perception is the sound pressure that relates the sound intensity
to the loudness. The sound pressure is known to vary inversely with the distance
between the source and the listener. This rule is highly important from the
perceptual point of view [27], 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 should be considered in this context. Another important aspect is the
timbre and more specifically the brightness variations, which can be physically
accounted for in terms of the air absorption [10]. A third phenomenon which
is well known in physics is the Doppler effect which explains why frequency
shifts can be heard while listening to the siren of an approaching police car
[26]. 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. Finally, the reverberation
is another aspect that enables the distinction between close and distant sound
sources [15]. A close sound source will produce direct sounds of greater magnitude
than the reflected sounds, which means that the reverberation will be weaker for
close sound sources than for distant ones.

When considering evoked motion at a more metaphoric level, like for instance
in music and cartoon production processes, signal morphologies responsible for
the perceived motion cannot be directly linked to physics and must be identi-
fied in other ways, for instance through listening tests. The selection of stimuli
for such investigations is intricate, since the recognition of the sound produc-
ing source might influence the judgement of the perceived motion. For instance,
when the sound from a car is presented, the motion that a listener will associate
to this sound will most probably be influenced by the possible motions that the
car can make, even if the sound might contain other interesting indices that
could have evoked motions at more metaphoric levels. To avoid this problem,
we therefore decided to investigate motion through a specific sound category,
so-called "abstract sounds" which are sounds that cannot be easily associated to
an identifiable sound source. Hence, when listeners are asked to describe evoca-
tions induced by such sounds, they are forced to concentrate on intrinsic sound
properties instead of the sound source. Such sounds, that have been explored
by electroacoustic music composers, can be obtained from both recordings (for
instance with a microphone close to a sound source) or from synthesized sounds
obtained by for instance granular synthesis [23]. In a previous study aiming at
investigating semiotics of abstract sounds [28], subjects often referred to various
motions when describing these sounds. This observation reinforced our convic-
tion that abstract sounds are well adapted to investigate evoked motion.

As a first approach toward a perceptual control of evoked motion, perceived
motion categories were identified through a free categorization test [24]. Subjects
were asked to categorize 68 abstract sound and further give a verbal description
of each category. Six main categories were identified through this test, i.e. "rotat-
ing", "falling down", "approaching", "passing by", "going away", "going up".
The extraction of signal features specific to each category revealed a system-
atic presence of amplitude and frequency modulations in the case of sounds
belonging to the category "turning", a logarithmic decrease in amplitude in the
category "passing" and amplitude envelopes characteristic of impulsive sounds
for the category "falling". Interestingly, several subjects expressed the need to
make drawings to describe the perceived motions. This tends to indicate that
a relationship between the dynamics of sounds and a graphic representation is
intuitive. This observation was decisive for our control strategy investigation
presented in the next section.

4.2 Control of evoked motion

In the case of evoked motion the definition of a perceptual control is as pre-
viously mentioned less straightforward than in the case of impact sounds and
environmental sounds. From the free categorization test described in the previous
section, categories of motion 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 correspond-
ing to evoked dynamic patterns, further experiments were conducted in which

subjects were asked to describe the evoked trajectories by drawings. Since hand
made drawings would have been difficult to analyze and would have been influ-
enced by differences in people's ability to draw, a parametrized drawing interface
was developed, meaning that subjects were given identical drawing tools that
required no specific skills. The control parameters available in the interface were
based on the findings obtained in the free categorization test, and the accuracy
of the drawing was limited to prevent the interface from becoming too complex
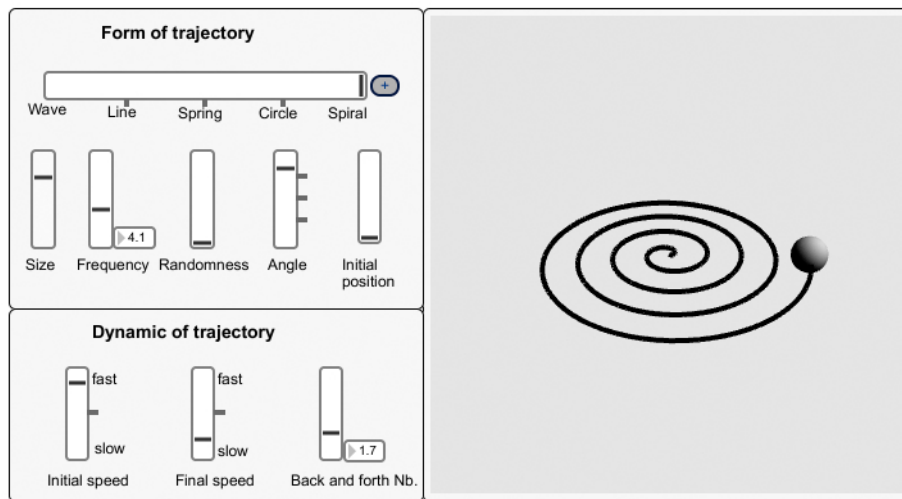to handle. The interface is shown in Figure 4.

Fig. 4. Graphical User Interface

Two aspects, i.e. shape and dynamics, enabled the subjects to define the mo-
tion. Six parameters were available to draw the shape of the trajectory (shape,
size, frequency oscillation, randomness, angle, initial position) and three param-
eters were available to define the dynamics (initial and final velocity and number
of returns). Each time a sound was presented, the subject made a drawing that
corresponded to the trajectory he or she had perceived. No time constraint was
imposed and the subject could listen to the sound as often as he/she wanted.
The dynamics was illustrated by a ball that followed the trajectory while the
sound was played.

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 showed good agreement on the dis-
tinction between linear and oscillating movements and between wave-like and
circular oscillations. This means that these three aspects give a sufficiently ex-
act control of the perceived shape of sound trajectories. As far as the orientation
of the trajectory was concerned, only the distinction between horizontal and

11

vertical seems to be relevant. While there was agreement among subjects about
the distinction between the upward/downward direction, the difference between
the left/right direction 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. This might have been related to the graphical user interface which
did not provide a sufficiently precise control of the dynamics according to several
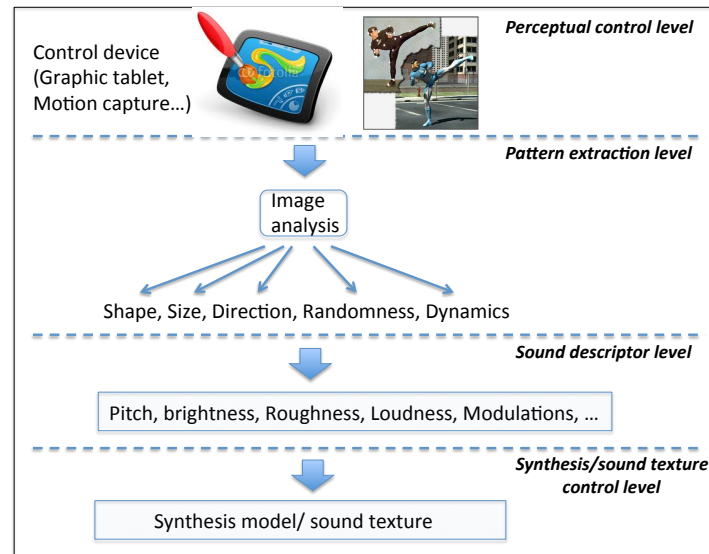subjects.

Fig. 5. Generic motion control

The identification of perceptually relevant parameters enabled the definition
of a reduced number of control possibilities. Hence 3 kinds of shapes (linear,
circular and regular), 3 different directions (south, north and horizontal), and
various degrees of oscillation frequencies (high and low), randomness (non, lit-
tle, much), size (small, medium, large) and dynamics (constant, medium an high
speed) were found to be important control parameters that enabled the defini-
tion of perceived trajectories. Based on these findings, a generic motion control
strategy could hereby be defined as shown in Figure 5. This strategy could be
separated in three parts, i.e. a perceptual control level based on drawings, an
image processing level dividing the drawings in elementary patterns (i.e. waves,
lines, direction, etc) and a third level containing the synthesis algorithm or a
sound texture.

5   Conclusion and Discussion

This article describes perceptual control strategies of synthesis processes obtained from the identification of sound structures (invariants) responsible for evocations induced by sounds. In the case of impact sounds, these sound structures are obtained by investigating the perceptual relevance of signal properties related to the physical behavior of the sound sources. Variations of physical phenomena such as dispersion and dissipation make perceptual distinctions possible between different types of objects (i.e. strings versus bars or plates versus membranes) or materials (wood, glass, metal,...). The spectral content of the impact sound, in particular the eigen-frequencies that characterize the modes of a vibrating object, is responsible for the perception of its shape and size. In cases where the physical behavior of sound sources are not known (*e.g.* certain categories of environmental sounds) or cannot explain evocations (*e.g.* metaphoric description of motion), recorded sounds are analyzed and linked to perceptual judgements. Based on the invariant signal structures identified (chirps, noise structures, ...), various control strategies that make it possible to intuitively control interacting objects and immersive 3-D environments are developed. With these interfaces, complex 3-D auditory scenes (the sound of rain, waves, wind, fire, etc.) can be intuitively designed. New means of controlling the dynamics of moving sounds via written words or drawings are also proposed.

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. This makes it for instance possible to transform an initially stationary sound into a sound that evokes a motion that follows a specific trajectory.

References

1. Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S.: 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-Verlag Berlin Heidelberg (2009)
2. Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S.: Controlling the perceived material in an impact sound synthesizer. IEEE Transactions on Audio, Speech, and Language Processing 19(2), 301–314 (2011)
3. Aramaki, M., Gondre, C., Kronland-Martinet, R., Voinier, T., Ystad, S.: Imagine the sounds : an intuitive control of an impact sound synthesizer. In: Ystad, Aramaki, Kronland-Martinet, Jensen (eds.) Auditory Display, Lecture Notes in Computer Science, vol. 5954, pp. 408–421. Springer-Verlag Berlin Heidelberg (2010)
4. Aramaki, M., Kronland-Martinet, R.: Analysis-synthesis of impact sounds by real-time dynamic filtering. IEEE Transactions on Audio, Speech, and Language Processing 14(2), 695–705 (2006)

5. Aramaki, M., Kronland-Martinet, R., Voinier, T., Ystad, S.: A percussive sound synthetizer based on physical and perceptual attributes. Computer Music Journal 30(2), 32–41 (2006)
6. Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., Besson, M.: Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. Journal of Cognitive Neuroscience 22(11), 2555–2569 (November 2010)
7. Brun, M.L.: Digital waveshaping synthesis. JAES 27(4), 250–266 (1979)
8. Bzat, M., Roussarie, V., Voinier, T., Kronland-Martinet, R., Ystad, S.: Car door closure sounds : Characterization of perceptual properties through analysis-synthesis approach. In: International Conference on Acoustics, (ICA 2007). Madrid (2007)
9. Chowning, J.: The synthesis of complex audio spectra by means of frequency modulation. JAES 21(7), 526–534 (1973)
10. Chowning, J.: The simulation of moving sound sources. Journal of the Audio Engineering Society 19(1), 2–6 (1971)
11. Gaver, W.W.: How do we hear in the world? explorations in ecological acoustics. Ecological Psychology 5(4), 285–313 (1993)
12. Gaver, W.W.: What in the world do we hear? an ecological approach to auditory event perception. Ecological Psychology 5(1), 1–29 (1993)
13. Giordano, B.L., McAdams, S.: Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates. Journal of the Acoustical Society of America 119(2), 1171–1181 (2006)
14. Gobin, P., Kronland-Martinet, R., Lagesse, G.A., Voinier, T., Ystad, S.: From sounds to music: Different approaches to event piloted instruments. In: Uffe Kock Wiil (ed.) Computer music modeling and retrieval, pp. 225–246. Lecture Notes in Computer Science, Springer Berlin / Heidelberg (2003)
15. Jot, J.M., Warusfel, O.: A real-time spatial sound processor for music and virtual reality applications. In: Proceedings of the International Computer Music Conference (ICMC '95). pp. 294–295 (1995)
16. Kleczkowski, P.: Group additive synthesis. Computer Music Journal 13(1), 12–20 (1989)
17. Kronland-Martinet, R.: The use of the wavelet transform for the analysis, synthesis and processing of speech and music sounds. Computer Music Journal 12(4), 11–20 (1989)
18. Kronland-Martinet, R., Guillemain, P., Ystad, S.: Modelling of natural sounds by time-frequency and wavelet representations. Organised Sound 2(3), 179–191 (1997)
19. Kronland-Martinet, R., Voinier, T.: Real-time perceptual simulation of moving sources: Application to the leslie cabinet and 3d sound immersion. EURASIP Journal on Audio, Speech, and Music Processing 2008 (2008)
20. Mathews, M.: The digital computer as a musical instrument. Science 142(3592), 553–557 (1963)
21. McAdams, S.: Perspectives on the contribution of timbre to musical structure. Computer Music Journal 23(3), 85–102 (2011)
22. McAdams, S., Bigand, E.: Thinking in Sound: The cognitive psychology of human audition. Oxford University Press (1993)
23. Merer, A., Ystad, S., Aramaki, M., Kronland-Martinet, R.: Abstract Sounds and Their Applications in Audio and Perception Research, chap. Exploring Music Contents, pp. 269–297. Springer-Verlag Berlin Heidelberg (2011)
24. Merer, A., Ystad, S., Kronland-Martinet, R., Aramaki, M.: Computer Music Modeling and Retrieval. Sense of Sounds, chap. Semiotics of Sounds Evoking Motions:

Categorization and Acoustic Features, pp. 139–158. Springer Berlin / Heidelberg (2008)

25. Miranda, E., R., Wanderley, M.: New Digital Musical Instruments: Control And Interaction Beyond the Keyboard. A-R Editions (2006)

26. Neuhoff, J., McBeath, M.: The doppler illusion: the influence of dynamic intensity change on perceived pitch. Journal of Experimental Psychology: Human Perception and Performance 22(4), 970–985 (1996)

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

28. Schn, D., Kronland-Martinet, R., Ystad, S., Besson, M.: The evocative power of sounds: Conceptual priming between words and nonverbal sounds. Journal of Cognitive Neuroscience 22(5), 1026–1035 (2010)

29. Sciabica, J., Bezat, M., Roussarie, V., Kronland-Martinet, R., S., Y.: Towards the timbre modeling of interior car sound. In: 15th International Conference on Auditory Display. Copenhagen (2009)

30. Terhardt, E., Stoll, G., Seewann, M.: Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. Journal of Acoustical Society of America 71(671-678) (1982)

31. Verron, C., Aramaki, M., Kronland-Martinet, R., Pallone, G.: A 3d immersive synthesizer for environmental sounds. IEEE Transactions on Audio, Speech, and Language Processing 18(6), 1550–1561 (2010)

32. Verron, C., Pallone, G., Aramaki, M., Kronland-Martinet, R.: Controlling a spatialized environmental sound synthesizer. In: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). pp. 321–324. New Paltz, NY (2009), 18-21 October 2009

33. Verron, C., Aramaki, M., Kronland-Martinet, R., PALLONE, G.: Spatialized additive synthesis. In: Acoustics '08. Paris, France (Jun 2008), `http://hal.archives-ouvertes.fr/hal-00463365`, oR 20 OR 20 CIFRE

34. Verron, C., Aramaki, M., Kronland-Martinet, R., Pallone, G.: Analysis/synthesis and spatialization of noisy environmental sounds. In: Proc. of the 15th International Conference on Auditory Display. pp. 36–41. Copenhague, Danemark (2009)

35. Warren, J., Zielinski, B., Green, G., J.P., R., Griffiths, T.: Perception of sound-source motion by the human brain. Neuron 34(1), 139–148 (2002)