Perceptual Characterization of Motion Evoked by Sounds for Synthesis Control Purposes

ADRIEN MERER, MITSUKO ARAMAKI, SØLVI YSTAD, and RICHARD KRONLAND-MARTINET,
LMA, CNRS UPR 7051, Aix-Marseille Univ, Centrale Marseille

This article addresses the question of synthesis and control of sound attributes from a perceptual point of view. We focused on an attribute related to the general concept of motion evoked by sounds. To investigate this concept, we tested 40 monophonic abstract sounds on listeners via a questionnaire and drawings, using a parametrized custom interface. This original procedure, which was defined with synthesis and control perspectives in mind, provides an alternative means of determining intuitive control parameters for synthesizing sounds evoking motion. Results showed that three main shape categories (linear, with regular oscillations, and with circular oscillations) and three types of direction (rising, descending, and horizontal) were distinguished by the listeners. In addition, the subjects were able to perceive the low-frequency oscillations (below 8 Hz) quite accurately. Three size categories (small, medium, and large) and three levels of randomness (none, low amplitude irregularities, and high amplitude irregularities) and speed (constant speed and speeds showing medium and large variations) were also observed in our analyses of the participants' drawings. We further performed a perceptual test to confirm the relevance of the contribution of some variables with synthesized sounds combined with visual trajectories. Based on these results, a general typology of evoked motion was drawn up and an intuitive control strategy was designed, based on a symbolic representation of continuous trajectories (provided by devices such as motion capture systems, pen tablets, etc.). These generic tools could be used in a wide range of applications such as sound design, virtual reality, sonification, and music.

Categories and Subject Descriptors: H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing—*Systems*; J.5 [**Computer Applications**]: Arts and Humanities—*Music*

General Terms: Design, Experimentation, Performance, Theory

Additional Key Words and Phrases: Description, motion, perception, trajectories, synthesis control, mapping, sound perception, drawing

ACM Reference Format:

Merer, A., Aramaki, M., Ystad, S., and Kronland-Martinet, R. 2013. Perceptual characterization of motion evoked by sounds for synthesis control purposes. ACM Trans. Appl. Percept. 10, 1, Article 1 (February 2013), 24 pages.
DOI = 10.1145/2422105.2422106 http://doi.acm.org/10.1145/2422105.2422106

1. INTRODUCTION

One of the central issues that arises when designing sound synthesizers is how to provide end-users with meaningful intuitive control parameters. Synthesizers based on intuitive control devices are of great importance in the fields of sound design and virtual reality, where sounds tend to be used to convey specific kinds of information. They could also be useful in other fields, such as post-production cinema applications and video games, as means of replacing sound databases indexed with verbal

Authors' address: 31 chemin Joseph Aiguier 13402 Marseille Cedex 20, France; email: merer@lma.cnrs-mrs.fr

DOI 10.1145/2422105.2422106 http://doi.acm.org/10.1145/2422105.2422106

labels. In addition, since synthesizers are able to produce an infinite variety of sounds (which are not subject to any physical constraints), they seem to have more to offer than prerecorded sound databases. To obtain synthesizers endowed with intuitive mapping abilities, it is necessary to take human perception into account, right from the very first steps in the design of the synthesizer. In practice, mapping between the basic signal parameters and the intuitive control device can be achieved by defining three layers, consisting of high-level parameters (describing the way sounds are perceived in terms of words, drawings, gestures etc.); middle-level parameters (relating to the characteristics of the signal); and low-level parameters (the synthesis parameters). A few recent attempts have been made to design synthesizers based on similar constraints. Some authors have presented intuitive control systems with which sound can be generated directly from semantic descriptions of the sound source or the sound quality in terms of timbre [Gounaropoulos and Johnson 2006; Le Groux and Verschure 2008; Aramaki et al. 2011]. Several studies have also focused on feature-based sound synthesis, where signal features (i.e., middle-level parameters) that are known to be relevant from the perceptual point of view (such as the spectral centroid, roughness, etc.) are used to control the process of synthesis [Hoffman and Cook 2006]. Note that in the musical context, the mapping procedure has to preserve the causal links between gesture and sound and achieve the complex relationships between the control and synthesis parameters (as occurs with acoustic instruments). Various approaches such as machine-learning techniques [Miranda 1998; McDermott et al. 2008] were based on an automatic mapping between the control and synthesis parameters, and many other tools are also available [Bevilacqua et al. 2005; Steiner 2006; Malloch et al. 2008] to facilitate the testing of intrinsic relationships of this kind. The aim of the present study was to draw up a framework for a mapping strategy for implementing an intuitive control of motion evoked by sounds, which is a fundamental auditory attribute. We focused here on monophonic sounds presented diotically, which ruled out the use of spatial cues. The concept of motion evoked by sounds is much more complex than it may seem, since it can be addressed from several perspectives. Motion is obviously associated with sound in the case of a physical moving sound source or some human gestures. But a sound can also evoke a motion in a more metaphoric way, and this aspect is widely used in music and in cartoon production processes, for instance. The method used consisted first in determining how motion evoked by sounds can be described and characterized from a perceptual point of view and then in identifying the high-level parameters involved in the mapping strategy. Hence we started by running listening tests using so-called “abstract” sounds (which are commonly used in electro-acoustic musical compositions), defined as sounds in which the physical sources cannot be easily recognized [Merer et al. 2010]. Sounds of this kind are believed to provide relevant keys for investigating the concept of motion and the dynamic aspects of sound in general, which go beyond physical considerations. We asked the subjects to describe how they perceived the motions evoked by these sounds by answering a questionnaire and drawing trajectories. It was established in previous studies that drawings are a relevant means of describing motion in an intuitive way. Based on the subjects’ answers to the questionnaire, their drawings were further analyzed to determine the most relevant variables describing the motion from a perceptual point of view. Since the design of a global control strategy depends on the prospective applications (generating either realistic or aesthetic sounds, for example), the mapping between high-level and low-level parameters on which the design of the synthesizer was based, was taken to be user-dependent. For present purposes, we designed a physically-based synthesis strategy to generate the stimuli to be used in the listening test performed to validate some of the variables related to the high-level control parameters describing the evoked motion. This validation test then led to drawing up a typology of the motion evoked by sounds reflecting the results obtained in this particular study. Lastly, we designed a generic sonification and control tool for evoking motion with sound from any input device generating continuous trajectories: this tool could be used in a wide range of applications.

2. MOTION EVOKED BY SOUNDS

The concept of evoked motion is not at all straightforward, and might involve actual physical motion as well as metaphoric descriptions of motion, such as those used in a musical context. We decided here to call this general concept "motion evoked by sound". In this section, various aspects of this concept will be described. Motion can be evoked by sounds when the actual physical sources are moving. In the case of silent sources, these sounds are due to the interactions between the sources and the environment (impact, bouncing, etc.) alone. In the case of sounding sources, they will result from changes in the initial sound (before the occurrence of any displacement) caused by the environment. Based on physical considerations, moving sound sources can be accurately characterized by a few acoustic attributes of the signal. For instance, the sound of a passing source can be simulated using frequency shifts along with intensity variations (Doppler effect). These sound effects can be produced satisfactorily under monophonic playback conditions, even with a poor quality loudspeaker. From the perceptual point of view, some studies have focused on the perception of motion cues with elementary stimuli (such as a harmonic comb, noise burst, and pure tone) [Lutfi and Wang 1999; Carlile and Best 2002; Kaczmarek 2005], but they have dealt with only a few motion attributes (speed and acceleration with sources moving in front of a listener). Other studies based on the use of complex stimuli from everyday life (the ecological approach [Gaver 1993]) have focused on specific types of motion such as breaking and bouncing [Warren and Verbrugge 1984], and have led to the identification of the temporal patterns mediating the perception of these movements. However, no exhaustive perceptual assessment of motion attributes has been proposed so far, and some aspects, such as the recognition of a source trajectory, still remain to be explored. Sound-generating interactions between humans and their environment constitute a particular case of evoked motion. Several studies have suggested that the perception of movements of this kind may be based on the identification of the underlying gestures, which might contribute to a motor theory of perception [Liberman and Mattingly 1985]. In some cases the perception of evoked motion can be addressed in a more metaphoric way, without resorting to physical considerations. Interestingly, sound designers for animated video films commonly use this approach to sonify visual movements with sounds that do not actually correspond to real situations, in order to produce comical effects. Some well-known examples of this process are those where a jumping cartoon character is accompanied by a series of chirps and a falling cartoon object is accompanied by a slowly descending chirp, although no such sounds occur in reality. In these cases the sonification strategy is based on the insertion of temporal variations (in the frequency, intensity, timbre, etc.), which are synchronized with the dynamics of the visual motion. Note that auditory and visual interactions seem to be particularly important and complex in the case of motion perception [Riecke et al. 2009]. Certain studies reveal that auditory-visual desynchrony might be tolerated to some extent [Dixon and Spitz 1980] as well as some audiovisual incompatibilities. For instance, we can tolerate incompatibilities between voice and lip position in vowel perception [Summerfield and McGrath 1984] and a temporal delay might even be necessary to avoid perceptual asynchrony [Arrighi et al. 2006]. Fortunately, it seems that bimodal integration occurs without domination of one modality, and in the case of motion detection, unimodal information is still available [Alais and Burr 2004]. Lastly, as suggested by the expression "motion without movement," the concept of evoked motion is more general than the mere perception of sound sources moving in space. Motion has been extensively explored in music analysis [Shove and Repp 1995; Eitan and Granot 2006; Fremiot et al. 1996] and in musical interpretation [Kronman and Sundberg 1987; Friberg and Sundberg 1999]. Some studies have focused on the relationships between music and motion in general [Honing 2003; Johnson and Larson 2003] and on the nature of these close relationships. In the musical context, the term "movement" is generally used to refer to part of a piece that is characterized by a specific type of musical dynamics (allegro, largo, etc.). In recent years, an

increasing interest in the use of sounds to convey information related to movement has been observed. Applications of such research can be found in a large variety of domains such as medicine related to motor training and sports (e.g., stroke rehabilitation [Chen et al. 2008], physiotherapy [Vogt et al. 2010], elite sports [Schaffert et al. 2010]), and the car industry [Larsson 2010].

3. THE USE OF ABSTRACT SOUNDS

We propose to focus here on a sound attribute that encompasses the aspects of motion mentioned above, from physical movement to motion at a more metaphorical level, such as that which occurs in music. The choice of a suitable sound corpus was therefore delicate, since it involved sounds allowing the exploration of these various aspects of motion. In particular, to obtain a more metaphoric level of evoked motion, we did not want people's judgments to be influenced by the fact that they could identify the actual sources. The perception of motion would be considerably biased if the sound sources were easily recognizable. For instance, if we listen to the sound of a car or a motorcycle (i.e., a sound emanating from a clearly identified source), the corresponding motion would be described as something that is passing by, approaching or moving away, since we know that this is what sources of this kind typically do. In these cases the evocation of motion may be influenced by what the listener knows about the physical source, in addition to the intrinsic acoustic attributes of the sounds themselves. Whereas abstract sounds were assumed to be suitable material for investigating the perceived motion induced by intrinsic acoustic attributes, since no particular physical source can be associated with these sounds. In a previous study (where no particular attention was paid to evoked motion), in which we asked the subjects to briefly describe abstract sounds played with a monophonic device, we observed that many of the words they used related to motion [Schön et al. 2010]. In the case of several sounds, the motion-related attributes were the only aspects the subjects were able to describe. To further investigate this point, we conducted a preliminary study on evoked motion, in which abstract sounds were presented using a free categorization procedure [Merer et al. 2008]. We asked the subjects to classify a set of 68 abstract sounds depending on the motion evoked (with no constraints on the number of categories used) and to label each subset of sounds with a word describing the motion. They produced six main categories of motion: "rotating," "falling down," "approaching," "passing by," "going away," and "going up". We also determined a number of acoustical patterns associated with these categories, such as two low-frequency amplitude modulations in the category "rotating". These preliminary findings showed the existence of some trends in the perception of motion evoked by sounds. The findings obtained in the present study using an original experimental procedure, including a questionnaire and a drawing test, made it possible to specify more closely the attributes conveying motion-related information.

4. SOUND ASSESSMENT BASED ON A QUESTIONNAIRE AND DRAWINGS

In this section we describe how evoked motion was assessed in a listening test divided into two parts. In the first part, listeners completed a questionnaire about specific aspects of a series of sounds and their sources and the motion they possibly evoked. In the second part, they were asked to describe this motion in the form of drawings. The idea of depicting motion in drawings arose from some findings made in our previous study (cf., Section 3), in which we carried out a free categorization test [Merer et al. 2008]. Interestingly, we observed that some listeners said they would have preferred to use drawings (cf., Figure 1) to describe their motion categories, although nothing was said in the instructions that might have suggested this mode of expression. Drawings are commonly used in contemporary music to either describe or create music [Thiebaut et al. 2008]. The use of drawings has also led to some interesting applications, including the development of new sonification strategies [Andersen and Zhai 2008]. Hence, drawing seemed to be a natural way of describing the motion evoked by sounds and of controlling perceptually relevant attributes for synthesis purposes. We are aware, of course, that this

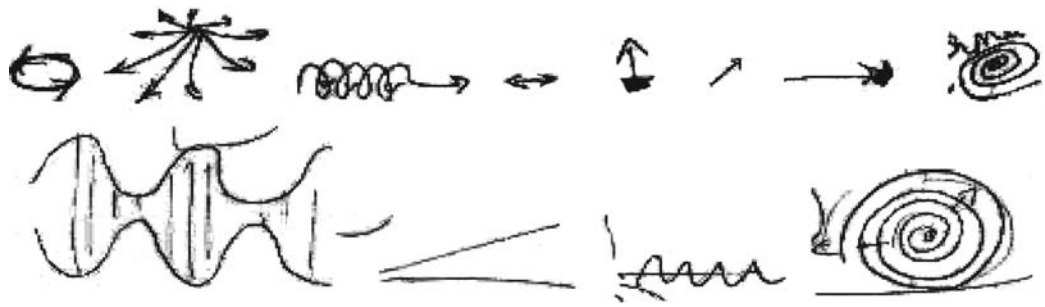

Fig. 1. Examples of drawings obtained in a previous study [Merer et al. 2008] to describe categories of motion.

procedure may result in audio-visual interactions, and this aspect will be discussed in connection with the validation of the synthesis strategy (cf., Section 7.5). Since hand-made drawings show considerable variability due to differences in people's ability to draw, we decided to create a parametrized drawing interface. "Parametrized" here means that all the subjects were given identical drawing tools requiring no specific skills. This meant that compromises had to be made between the possibilities of the interface, the accuracy of the description produced, and learning issues. An informal test on a few subjects showed that an interface for drawing 3-D sound trajectories around a listener would be extremely complex to handle. The scope of the interface was therefore reduced by defining parameters that could be easily handled by subjects and by fixing the viewing angle with the interface so that the sound trajectory was always drawn in front of the subjects. This means that certain aspects that might have been useful for motion description, such as source localization behind the listener, sensation of immersion, and so on, could not be taken into account with this interface. The control parameters available in the drawing interface were based on the findings previously obtained in the free categorization test. In particular, it was necessary to give the subjects the possibility of reproducing the main tendencies previously observed in their spontaneous drawings and to reproduce at least the six motion categories identified in the previous study [Merer et al. 2008] (cf., Section 3). In addition, we added means of controlling the temporal dynamics, angle and randomness, as these parameters were assumed to be perceptually relevant as far as evoked motion is concerned.

4.1 Subjects

Twenty-nine subjects (10 women and 19 men aged 21 to 53 years; mean age 29.8) participated in the experiment. Among the participants, 22 were musicians and/or acoustic and audio experts. Two subjects had no previous experience of listening tests. All the subjects were familiar with the use of computers, and 25 had obtained at least a Master's degree.

4.2 Drawing Interface

We designed a custom drawing interface for this study[1] for collecting and displaying the parametrized trajectories. We used the Max/MSP[2] graphic programming environment including OpenGL tools to obtain real-time 3-D image synthesis with automatic shading and curve perspective. The graphic user interface (GUI) is shown in Figure 2. Users can draw a trajectory by manipulating nine linear sliders

[1]Mac and Windows versions of this software are available at http://www.lma.cnrs-mrs.fr/~kronland/Motion/. The possibilities of the interface as well as all the stimuli used in the tests are also presented.
[2]http://cycling74.com.

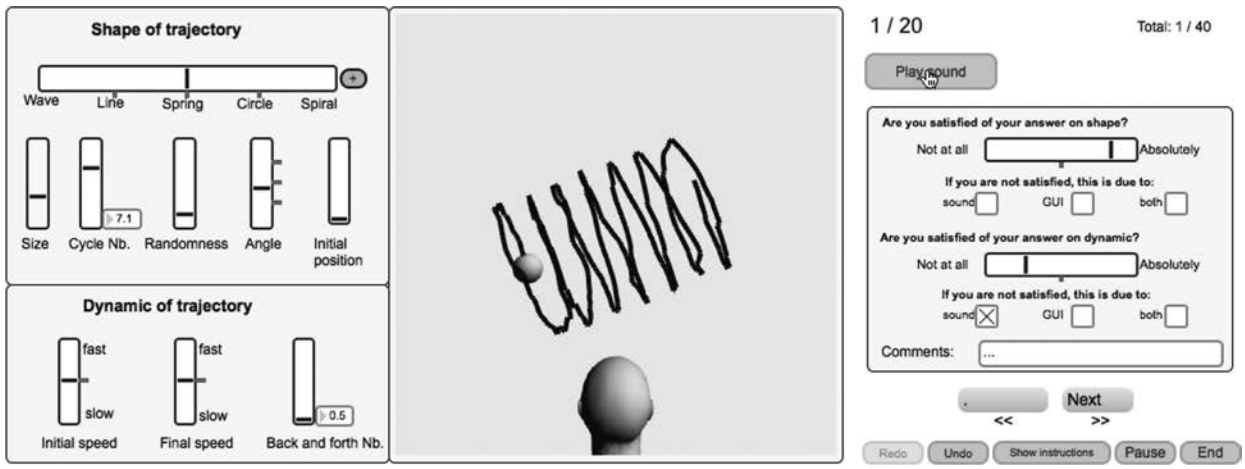

Fig. 2. Graphic user interface (GUI) designed for the listening test. Sliders on the left define the trajectory displayed in the central window. When the sound is played (in response to the "Play sound" button), the point source (grey sphere) moves along the trajectory (black curve) synchronously with the sound. At the end of each sound test, the participants are asked to assess the test on Shape and Dynamics separately (right window).

corresponding to the six variables characterizing the shape properties (*Shape*, *Size*, *Number of cycles*, *Randomness*, *Angle*, and *Initial position*) and three variables characterizing the temporal dynamics (*Initial speed*, *Final speed*, and *Back & Forth*). The resulting trajectory is displayed in perspective in front of the listener, who is represented virtually on the screen, viewed from behind. To account for the dynamic aspects, a point source (a grey sphere) moves along this trajectory (a black curve) synchronously with the sound. With the *Shape* slider, users can select five predefined shapes (*Wave*, *Line*, *Spring*, *Circle*, and *Spiral*) and make continuous transitions between them in a successive order. Continuous transitions are possible, since from the mathematical point of view, these five shapes can be obtained from the equation of a helix:

$$\begin{cases} x(u) = A_1(1 - A_4 u)cos\Big(2\pi(A_5 u + A_6)\Big) \\ y(u) = A_2(1 - A_4 u)sin\Big(2\pi(A_5 u + A_6)\Big) \\ z(u) = A_3 u, \end{cases} \tag{1}$$

where u is the spatial sampling parameter and ranges from 0 to 1. The *Shape* slider acts simultaneously on A_1, A_2, A_3, and A_4, and the continuous transitions between predefined shapes are based on a linear interpolation of these parameters. The *Number of cycles* slider acts directly on A_5 and varies continuously from 0 to 8 cycles. The *Initial position* slider acts directly on A_6 and varies continuously from 0 to 2π. Note that depending on the shapes, some controls become useless, such as the *Number of cycles* when a *Line* is being drawn. In these cases, the corresponding sliders are shaded in grey and made inaccessible to the user. In addition to the latter controls, users can also rescale the trajectory with the *Size* slider, rotate it with the *Angle* slider, and/or introduce stochastic variations with the *Randomness* slider. The *Size* slider varies continuously from 0 to 1, and the rescaled trajectory is the product of the Size parameter and the trajectory defined in Eq. (1). A *Size* equal to 1 (equal to 0, respectively) corresponds to a trajectory covering the whole surface (10% of the surface, respectively) of the display screen. The *Angle* slider can be continuously moved from $-\pi$ to π and acts differently depending on the shapes: the *Circle* and *Spiral* are rotated around the listener's horizontal axis (the ear axis), whereas

the *Wave*, *Line*, and *Spring* are rotated around the back/front axis. With a zero angle, *Circle* and *Spiral* are displayed in the vertical plane and *Wave*, *Line*, and *Spring* on the horizontal axis. The *Randomness* slider can also be continuously moved from 0 to 1, and is used to introduce stochastic variations in the initial shape, causing a change of trajectory by adding different sequences of a stochastic process to the helix equation. With a *Randomness* equal to 1, the magnitude of the stochastic variations added to the initial shape is equal to the size of the trajectory. As regards the dynamics, the *Initial speed* and *Final speed* sliders can be moved continuously from 0 (low speed) to 1 (high speed), and the *Back & Forth* slider gives access to up to 10 returns in 0.1 steps. A constant speed is obtained when *Initial speed* and *Final speed* are equal to 0.5. Since the duration of the sounds differed (see Section 4.3), the program automatically adjusted the speed of the point source with each sound so that the point source covered the whole trajectory within a time corresponding to the duration of the sound. Note that all the sliders on the GUI manipulated by the users were qualitative, that is, no values were displayed on the slider scales, except for *Number of cycles* and *Back & Forth*.

4.3 Stimuli

A useful way of collecting abstract sounds is to take the background used by electroacoustic music composers who have developed specific recording techniques and/or sound transformation methods to prevent sound sources from being too easily identified. In this study, we selected 40 stimuli from a sound database consisting of 200 sounds obtained from electroacoustic musical compositions (including some composed by one of the authors). We selected sounds using the four-step procedure described below. We first classified the 200 sounds using Schaeffer's typological criteria [Schaeffer 1966], which were suitable for our study since they were not intended *a priori* to be used for differentiating between types of sounds (and thus could be applied to abstract sounds). This classification is based on the spectral and temporal sound variations, with considerations on the complexity of those variations (see Dack [1999] for details of this typology). With this procedure, it is possible to ensure that the sounds selected cover a wide range of spectro-temporal characteristics. The 200 sounds selected were in fact distributed quite uniformly over Schaeffer's typology, although sounds with a varying pitch occurred more frequently. Since these sounds were initially generated to obtain specific musical effects, some of them were quite complex and carried several auditory streams. We transformed these sounds using a band reject filtering and/or transposition process in order to select a single auditory stream. We then conducted an informal listening test on four subjects in order to assess the level of identification associated with all the sounds selected, and removed any sounds that could still be associated in any way with an identifiable sound source. This left us with a set of 150 sounds matching the definition of "abstract" sounds (see the Introduction section). Lastly, we conducted a further listening test with the same four subjects in order to select the forty sounds which made use of all the control parameters available on the GUI. We used this criterion to increase the chances of obtaining the largest possible variety of trajectories during the test (cf., Section 4.2). This selection process was similar to that used in the formal design of the experimental (DOE) methods, where the number of stimuli is optimum in view of the hypotheses (the hypotheses were taken into account here in the choice of the interface control parameters). The choice of the stimuli (abstract and presenting a wide range of spectro-temporal characteristics) and the design of the GUI allowed the evocation of a large variety of motions. However, we are aware that certain types of motion could not be evoked—in particular, motions behind oneself (since the viewing angle of the central window was fixed and the trajectory was displayed in front of the listener) or motions of several sources simultaneously (since the selected sounds contained a single auditory stream). As the spectro-temporal characteristics of the 40 final stimuli differed considerably, we performed loudness equalization "by ear," based on the overall impression of loudness. Note that existing automatic methods such as that presented in Glasberg and Moore [2002] and Zwicker and

Fastl [1990] did not give satisfactory loudness equalization. An experimenter compared the sounds with each other, choosing the reference sound at random each time. We repeated the procedure three times to ensure that the gain obtained was consistent. We then windowed the stimuli by means of a fade in/ fade out function with a ramp time ranging from 20 to 200 ms, depending on the onset/offset envelope of the stimuli. The mean duration of the sounds was 2220 ms (it ranged from 540 to 5560 ms).

4.4 Procedure

The listening test took place in an audiometric room. Stimuli were presented diotically using Stax 3R202 headphones with a Stax SRM310 preamp Apple MacBook internal audio interface operating at 16 bits with a 44.1 kHz sampling frequency. We asked the participants to answer a questionnaire in a first session of the listening test and to produce drawings during the second session. Sounds were presented in random order at each session. Participants were free to adjust the sound level once at the beginning of the test. They could listen to the sounds several times and go back to their assessments of the previous sounds at any time during the test. We did not impose time constraints, and allowed pauses. In particular, the two sessions were conducted on two different days, with an average interval of 4.2 days between them. The printed instruction sheet, which included examples of possible answers to each question, could be consulted during the whole test. We emphasized that there were no "correct" responses and told the participants to be as spontaneous as possible and to take special care when assessing their own answers at the end of the session. Details of the two listening test sessions are given below.

4.4.1 *Questionnaire.* The questionnaire was a multiple choice test with exclusion rules (i.e., sometimes giving access to further questions or not, depending on the previous responses). The whole questionnaire was displayed via a graphic interface, and subjects gave their answers directly on the screen. The questions were as follows:

a) Could you recognize the sound source? (responses formed on a continuous linear scale from "not at all" to "clearly recognizable").

b) Is the sound natural or synthetic? (responses formed on a continuous linear scale from "synthetic" to "natural").

c) Does the sound evoke the displacement of an object? (three possible responses: Yes/No/I don't know).

d) *If response "No" to question c)*: Does the sound evoke a motion in any way? (Yes/No/I don't know).

There were four additional questions about the origin of the motion and how involved the subjects felt in the sound event:

e) *If response "Yes" to question c)*: Does the object produce sound itself (without moving)? (Yes/No/I don't know).

f) Does the object have an internal motion[3]? (Yes/No/I don't know).

g) Did the object's motion trigger a reaction from you[4]? (Yes/No/I don't know).

h) Could you imagine yourself producing this sound (with a gesture)? (Yes/No/I don't know).

[3]Below this question, it was specified: "Internal motion means an intrinsic motion of an object that is not necessarily moving. For instance, a pendulum clock is typically an object that does not move, but that produces a sound that evokes a motion due to its pendulum, which is moving from one side to the other."

[4]Below this question, it was specified: "Did you for example feel that the sound evoked an approaching object or any other action that incited you to avoid it or react in some way?"

We also collected response times (RTs), defined as the time elapsing from the onset of the sound the first time it was played until the participants had answered the questions, and the number of times each sound was played. We asked the subjects to assess the degree of difficulty of the test with each sound and invited them to give any comments they wanted to make.

4.4.2 *Assessment of Evoked Motions Based on Drawings.* The task applied in the second session consisted in drawing the motion evoked by each sound with the GUI described in Section 4.2. A two-step training session was run at the start to familiarize participants with the use of the GUI. The first step, which focused on how to handle the shape-related variables, consisted in copying several target shapes displayed on the screen. The second step, which focused on how to handle the dynamic variables, consisted in copying vertical and linear trajectories with various temporal patterns. During the training, no sounds were presented. After the training, the actual test was divided into 2 blocks, in each of which 20 sounds were presented. With each sound, we asked the participants to draw the trajectory they perceived, using the GUI. Initial values of each slider were set at zero (except for the speed parameters, which were set at 0.5, corresponding to constant speed) so that no specific trajectory was displayed the first time each sound was heard. Lastly, we asked the subjects to assess their satisfaction with their own responses (about the shape and dynamics separately) with a linear slider (on a continuous scale from "Not at all" to "Absolutely"). When low satisfaction ratings were given, they were asked whether this was due to the difficulty in assessing the sound, the GUI limitations or both. We also invited them to make any comments they had in mind. We recorded the nine slider values, the RTs and the number of times each sound was played. At the end of each session, the subjects completed a printed questionnaire designed to obtain an overall assessment of the test. This questionnaire included questions such as "Did you find the test complicated?", "Did you find the test too long?" or "Were certain controls too limited?"

5. GENERAL RESULTS

5.1 Duration of the Test and Participants' Comments

The first session lasted for 32 minutes on average (std: 10 min). We conducted a Kruskall—Wallis test on the RTs associated with each sound in order to determine which sounds were associated with significantly longer RTs ($\chi^2(39, 1120) = 70.29$, $p < 0.01$). These sounds may have been more difficult to assess than the others, although the subjects did not explicitly mention this fact in either their self-assessments or their comments (low correlation between the self-assessments ratings and the RTs: $R = 0.18$). We did not observe effects of habituation/tiredness (the RTs did not lengthen during the test), since the Kruskall—Wallis test on RTs with respect to the order of presentation was not significant ($\chi^2(39, 1120) = 32.28$, $p = 0.76$). In all, less than 5% of the subjects used the answer "I don't know". Sounds were played five times on average (std: 1.7). The second session lasted for 86 minutes on average, but there was considerable intersubject variability (ranging from 39 to 199 minutes). As in the first session, the order of presentation was not found to affect the RTs ($\chi^2(39, 1120) = 47.64$, $p = 0.16$), whereas the RTs depended on the sound ($\chi^2(39, 1120) = 151.73$, $p < 0.01$). After the interactive drawing session, we conducted a final interview with the subjects concerning their global impression of the test. This investigation revealed that 79% of the subjects thought that the test duration was reasonable, 62% that the task was not too complicated, and 82% that they would have preferred more subtle controls in the interface. Further details on their answers showed that the needs for more subtle controls concerned mainly the dynamics (30% for dynamics evaluations vs. 16% for shape evaluations). They said they would have liked to draw the velocity curve "by hand" instead of using the two sliders *Initial velocity* and *Final velocity*. Sounds were played 14 times on average (maximum of 25 times).

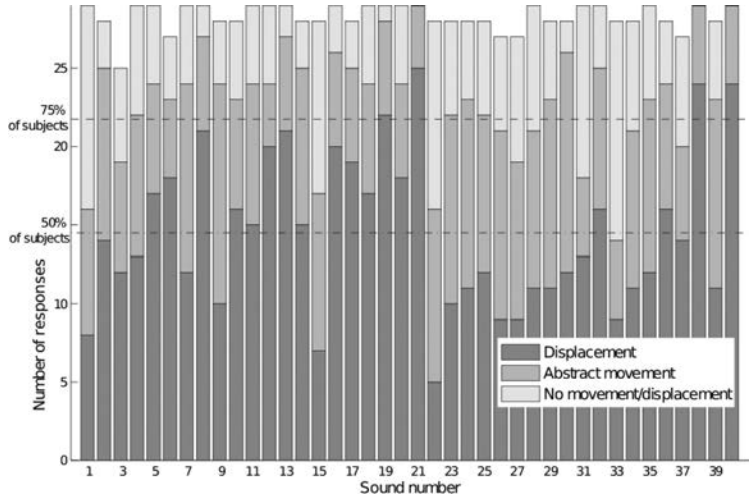

Fig. 3. Number of responses "Yes" to the questions "Does the sound evoke the displacement of an object" (in dark grey) and "Does the sound evoke a movement anyway?" (grey) and number of responses "No" to the first question (light grey). The sum of responses is not 100% since it was possible to answer "I don't know".

5.2 Sounds Evoking Motion

For each sound, we collected the scale ratings for the questions a) and b) and computed the percentages of "Yes," "No," and "I don't know" responses for the questions c) to h). Generally the percentage of "I don't know" response to each question was very low (less than 5%), revealing that sounds were clearly evocative of the perceptual attributes that were evaluated in the questionnaire. We selected a set of sounds that specifically evoked a motion, based on the responses to the items *c)* and *d)* in the questionnaire. We decided to keep the 29 sounds (out of the 40 tested in the experiments) that evoked motion in more than 75% of the listeners, for trajectory analysis (*cf.,* Figure 3). The threshold of 75% seemed to be a good compromise between the number of stimuli to be kept and the representativeness of the subjects' responses. Figure 3 shows that for the 29 selected sounds in average, 56% rated them as evoking a displacement (std 15%). When regrouping questions c) and d), an average of 79% (std 12%) of the users felt that each sound evoked a motion. Among these 29 sounds, 18 evoked an actual displacement of an object in more than 50% of the listeners (question *c)*), whereas only 4 sounds did so in more than 75% of them. Subjects did not recognize the source of the 29 sounds: the average answer to question *a)* was close to "Not at all". They also judged these 29 sounds to be "synthetic" (question *b)*), although most of them were obtained by recording natural sounds. Interestingly, a highly significant correlation was observed between the responses to questions a) and b) ($R = 0.89$), which means that the abstract and synthetic aspects of the sounds were intrinsically linked. Responses to questions *e)* to *h)* about the type of motion evoked were then analyzed. Based on the responses to the question *e)*, three sounds evoked a source producing sound itself, and four sounds evoked a silent object in more than 50% of the subjects. With the latter sounds, subjects presumably judged the sound event as being produced entirely by the interaction between the object and its environment (like a rolling or bouncing object). These findings show that for these three sounds, we are able to distinguish the sound produced by an object independently of its motion. It is worth noting that this was also the case with abstract sounds (where no physical source could be identified). This perceptual distinction could be taken into account in signal analysis, since different acoustic information is conveyed by sounds that are produced only by an object's displacement and those made by sounding objects that are transformed by their

own displacement. In response to the question *f*), six sounds elicited a positive answer in more than 50% of the listeners. This ability to differentiate between global and internal motion will be further discussed in Section 6.5. The questions *g)* and *h)* were designed to determine the subjective relationship between subjects and the auditory event. The sound events did not really affect the subjects, and we interpret this result as if the sound was localized outside the listeners' own perceptual space (in particular they did not feel that the sound or the event was associated with "might be dangerous"). In addition, we did not identify gesture-related sounds. This is in line with physiological evidence that the cognitive processing of human body motion has some specificities in comparison with the processing of nonhuman motion [Pellegrino et al. 1992]. The lack of gesture-related sound identification was also coherent with the definition of abstract sounds, since the ability to associate a sound with a human gesture may depend to some extent on the possibility of identifying the sound source.

6. TRAJECTORY ANALYSES

Next we focused on the 29 sounds that actually evoked a motion (Section 5.2). The drawings corresponding to these 29 sounds were analyzed in terms of the following attributes: perceived shape, direction and size of the trajectory, perception of oscillations in the sound, perceived randomness, and dynamic patterns during the trajectory. Here we selected, only drawings that were associated with high self-assessments. Self-assessment scores were scaled on the basis of each subject's average score for shape and dynamics separately (to reduce the interindividual differences and facilitate comparisons between the self-assessment scores). Data was defined as missing if the score was below the middle (default) value. We therefore analyzed the properties of 708 trajectories (133 of the initial set of 841 drawings - i.e., 29 subjects$\times$29 sounds - were judged by the subjects to be unsatisfactory), the dynamics of 678 trajectories and both the shapes and dynamics of 630 trajectories.

6.1 Shape

To characterize the various perceived shapes revealed by the subjects' drawings, eight basic shapes were defined, including the five initial ones (i.e., *Wave*, *Line*, *Spring*, *Circle*, and *Spiral*) and three additional shapes which occurred consistently in the selected set of drawings: *Hollow*, *Dome*, and *Bounce*. *Hollow* and *Dome* were produced using either *Wave* or *Spring* presets with *Number of cycles* values below 1 and specific combinations (four different combinations were possible) between *Angle* and *Initial position*. *Bounce* was produced using *Line* with a *Back & Forth* value greater than 1. We therefore classified drawings in one of the eight shape categories, so that each sound was associated with percentages of classification according to these shapes. We then performed factor analysis on these percentages, with Shape as the factor with maximum likelihood estimation and varimax rotation. Bartlett's Test of Sphericity was significant ($p < 0.01$) and the first two factors were selected (they explained 47% of the total variance). We rejected the third factor because it coincided with the inflection point on the scree plot, although its Eigenvalue was greater than 1. We also rejected the remaining factors, since the Eigenvalues were lower than 1. As shown in Figure 4, the x-axis (Factor I) singled out the linear category (mainly *Line* and, to a lesser extent, *Hollow* and *Dome*) from the other shapes, which were oscillating ones. In addition, among the oscillating shapes, the y-axis (Factor II) discriminated between two categories: "regular oscillations," which were mainly attributed to *Bounce* and "circular oscillations," which were attributed to *Spring* and *Spiral*. We therefore concluded that our set of stimuli involved three main shape categories: "linear," "regular oscillations," and "circular oscillations".

6.2 Oscillation Frequency

The above factor analysis on shapes differentiated mainly between oscillating and nonoscillating trajectories, which means that participants clearly detected the presence of oscillations in the sound. To

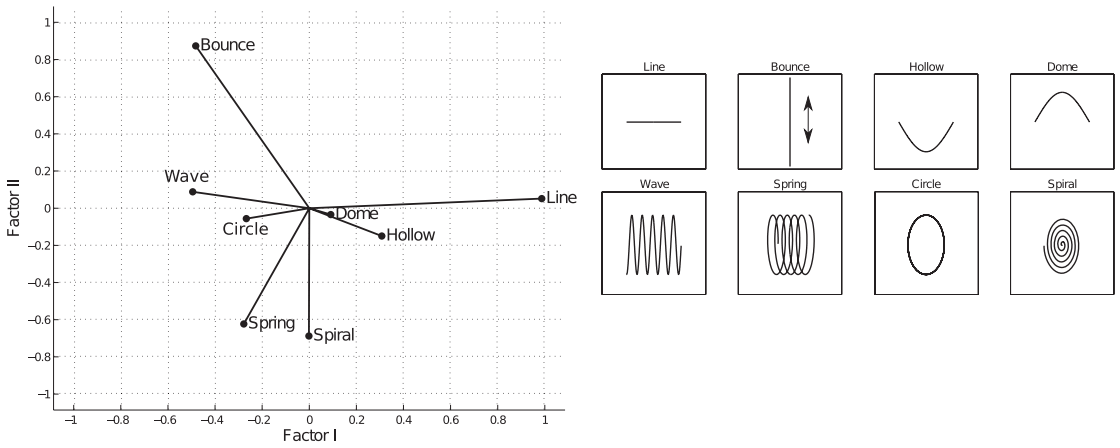

Fig. 4. Factor loading plot: the shapes projected onto the first two factors (left). Typical examples of these shapes are presented in the right part of the figure.

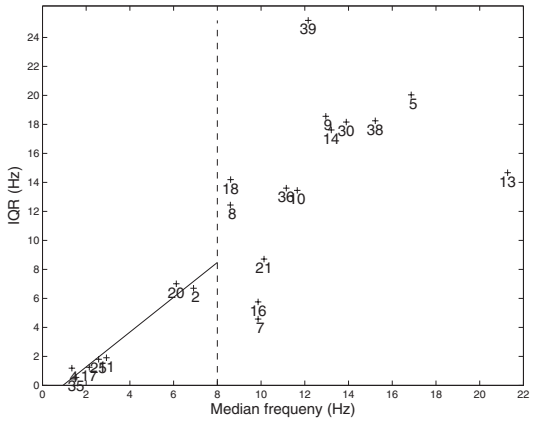

Fig. 5. Interquartile range (IQR) versus median values of *Oscillating frequency*. The regression line is shown in the case of sounds with a median value of less than 8 Hz.

test the subjects' ability to detect the number of oscillations, we focused on the 21 sounds (among the 29 initial sounds) that evoked an oscillating motion in more than 50% of the subjects. The GUI made it possible to control the oscillations in the trajectory via various strategies, that is, by setting *Number of cycles* at values higher than 1 and/or *Back & Forth* at values higher than 0.5. We defined a new variable called *Oscillation frequency* to quantify the overall contributions of these two strategies. This variable was defined from the product of *Number of cycles* and twice the *Back & forth* value divided by the sound duration, to make it homogeneous to a frequency (in Hz). We calculated median values and the interquartile range (IQR) of the *Oscillation frequency* of the 21 sounds selected. As illustrated in Figure 5, the highest correlations were obtained between median values and IQRs up to 8 Hz. Beyond 8 Hz, the correlation decreased. This indicates that subjects drew the number of oscillations below 8 Hz quite accurately and that the accuracy decreased beyond this frequency limit. Interestingly, trajectories with high *Oscillation frequency* values were also associated with high *Randomness* values, which means that subjects often added stochastic variations to highly oscillating trajectories. They may have

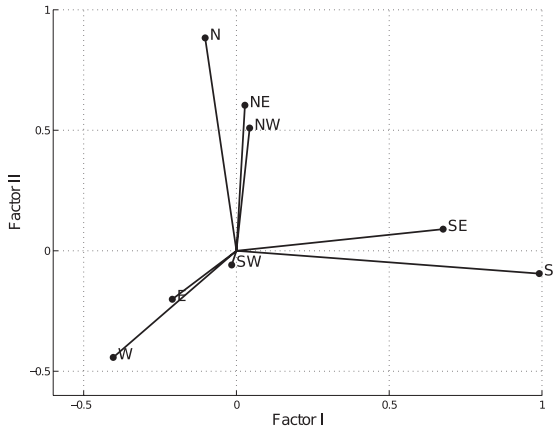

Fig. 6. Factor loading plot: the eight directions of the compass rose are projected onto the first two factors.

often perceived rapid oscillations as fuzziness in the sound. These findings therefore showed that the perception of "oscillations" differed between high and low frequencies and that for sound synthesis purposes, different mapping strategies should be designed, depending on the frequency range.

6.3 Direction

The perceived direction was given directly by the setting of the *Angle* slider. To reduce the high intersubject variability, we grouped the *Angle* values into the eight main *Directions* of a compass rose, namely, North-East (NE), North (N), North-West (NW), South-East (SE), South (S), South-West (SW), East (E), and West (W). In practice, with *Line*, *Wave*, and *Spring*, the *Angle* value was replaced by the nearest (among the eight) *Direction* value. The *Direction* of *Hollow* and *Dome* drawings was defined from the line between the initial and final points of the trajectory, since the actual angle displayed with these shapes depended on both *Shape* and *Initial position* settings (as discussed in the Section 6.1). No *Direction* was defined in the case of *Circle*, *Spiral*, and *Bounce* drawings because of their symmetry. We conducted a factor analysis on the percentage of classification in those eight directions for each sound, and three main factors emerged with Eigenvalues greater than 1, which explained 54.8% of the total variance. Results showed that general directions toward South and North were clearly discriminated: on the one hand, both S and SE directions were correlated with Factor I, and on the other hand, N, NE, and NW directions were correlated with Factor II (Figure 6). By contrast, E and W directions were positively correlated and grouped together, which means that these opposite directions were confused from the perceptual point of view. As was to be expected, since we were working with monophonic sounds presented diotically (i.e., with no interaural differences), the distinction between right-oriented (E *Direction*) and left-oriented (W *Direction*) trajectories was not relevant. Factor III was more difficult to interpret, since it was mainly supported by SE and SW. This factor probably corresponds to a distinction between sloping and more vertical trajectories. It can therefore be concluded that, in terms of *Direction*, three main categories were clearly distinguished by the subjects: South (descending trajectory), North (rising trajectory), and Horizontal (combined E and W *Direction*).

6.4 Size

The subjects' perception of the magnitude of the evoked trajectories was given directly by the setting of the *Size* slider. First of all, the differences between subjects' ratings were large and significant ($\chi^2(28, 679) = 102.24$, $p < 0.01$). In particular, the ranges used to depict the perceived sizes differed

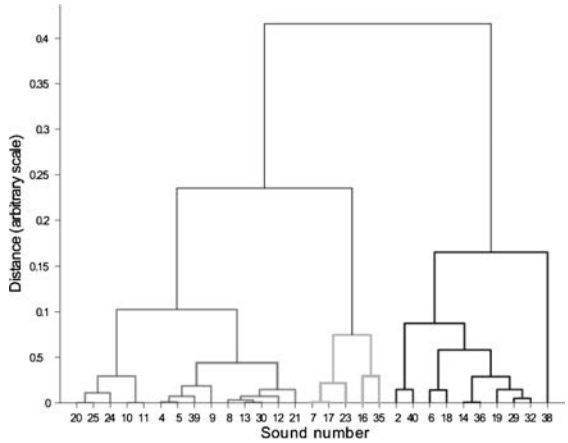

Fig. 7. Dendrogram for ratings of *Size* variable. Three sound clusters were identified, corresponding to different perceived sizes: medium (left - sounds 20-21), small (middle - sounds 7-35), and large (right - sounds 2-38).

Table I. Statistics from Kruskall–Wallis analysis conducted for Size, Randomness and Dynamics and from post-hoc tests with Bonferroni correction (the level of significance is indicated by $^{**}p < .01$)

	Cluster1 vs. Cluster2	Cluster1 vs. Cluster3	Cluster2 vs. Cluster3
Size	Small vs. Medium $\chi^2(1, 596) = 53^{**}$	Small vs. Large $\chi^2(1, 362) = 103^{**}$	Medium vs. Large $\chi^2(1, 452) = 38^{**}$
Randomness	None vs. Low amplitude $\chi^2(1, 513) = 174^{**}$	None vs. High amp. $\chi^2(1, 563) = 69^{**}$	Low amp. vs. High amp. $\chi^2(1, 334) = 45^{**}$
Dynamics	Constant vs. Somewhat variant $\chi^2(1, 329) = 12.9^{**}$	Constant vs. Greatly variant $\chi^2(1, 548) = 34.4^{**}$	Somewhat variant vs. Greatly variant $\chi^2(1, 351) = 49.6^{**}$

across subjects, which highlights the subjective nature of this attribute. Each subject's responses were therefore scaled by the difference between the individual median ratings and the overall median *Size*. This procedure decreased the mean IQR on sounds by 7.1% (from 0.28 to 0.26) and removed the significance of the effect of Size ($\chi^2(28, 679) = 9.71$, $p = 1$) on the subjects. We performed a hierarchical cluster analysis using the complete linkage method on the median values of *Size* (one value per sound). Results are presented in the dendrogram in Figure 7. Analysis of the groups obtained with different cut-off thresholds showed that the maximum number of groups in which the sounds differed significantly from each other was three: this finding was obtained with a threshold of around 0.2 (see Table I). The three groups were associated with the perception of small ($< 40\%$ of the window), medium ($[40\%, 60\%]$) and large ($> 60\%$) trajectory sizes. Note that we obtained similar results when focusing on the range occupied by the trajectory in the drawing window (i.e., the maximum value between the projections of the trajectory on the x- and y-axes).

6.5 Randomness

The perceived randomness of a motion is given directly by the setting of the *Randomness* slider. There was good agreement between subjects on sounds where no randomness was included in the trajectories (10 sounds obtained a median *Randomness* value equal to 0). There was also good agreement on sounds where subjects used the *Randomness* slider. With some of these sounds, subjects said Yes to the question *f*) *"Does the object have an internal motion?"*; *cf.,* Section 4.4.1). These findings tend to indicate that some subjects might have used the *Randomness* slider to characterize the internal motion evoked. We conducted a hierarchical cluster analysis using the complete linkage method on the mean

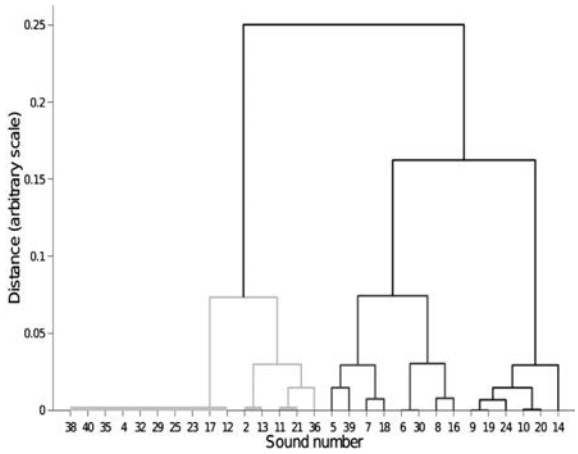

Fig. 8. Dendrogram of ratings on the *Randomness* variable. Three groups were identified, corresponding to various levels of randomness: none (left, sounds 38-36), low amplitude irregularities (middle, sounds 5-16), and high amplitude irregularities (right, sounds 9-14).

Randomness values (one value per sound). We didn't scale individual ratings because the differences were due to only one subject ($\chi^2(28, 679) = 54$, $p < 0.01$). Results are presented in the dendrogram in Figure 8. The largest number of groups was obtained at a threshold value of 0.1 with three clusters of sounds that differed significantly from each other (see Table I). The three groups corresponded to the amount of randomness: none, low amplitude irregularities (fluctuation magnitude between 5 and 15% of the size of the trajectory), and high amplitude irregularities (fluctuations $> 15\%$). We note that more than 95% of all the ratings were below half of the maximum value of the *Randomness* slider.

6.6 Dynamics

The speed of the point source moving along the trajectory was controlled by two sliders, that is, *Initial speed* (V_{init}) and *Final speed* (V_{final}) ranging from 0 to 1 (see Section 4.2). Subjects' perception of the dynamics of the motion were therefore given by the settings of these two sliders. High intersubject variability was observed between these ratings, which may have been partly attributable to the subjects' low satisfaction with the dynamic control possibilities (see Section 5.1). To reduce this variability, a new variable called *Dy* was introduced to quantify absolute deviations from constant speed (i.e., $V_{init} = 0.5$ and $V_{final} = 0.5$), corresponding to the default setting adopted for all sounds (see Section 4.4.2), defined as $Dy = |V_{init} - 0.5| + |V_{final} - 0.5|$. We performed a hierarchical cluster analysis using the complete linkage method on the mean *Dy* values obtained with each sound. Results are presented in the dendrogram in Figure 9. The largest number of groups was obtained at a threshold value of 0.3 with three clusters of sounds that differed significantly from each other (see Table I). The three categories corresponded to trajectories with a constant speed, somewhat variant, and greatly variant speed.

7. PERCEPTUAL VALIDATION BY SYNTHESIS

The above analysis of the drawings yielded a set of variables describing how sounds carrying motion are perceived. Among them, it was proposed to test the perceptual relevance of *Shape*, *Direction*, and *Oscillation frequency* on a new set of synthesized sounds. For this purpose, a sonification process based on moving sound source synthesis was used to generate synthetic sounds associated with a set of visual trajectories characterized by these variables. We then performed a formal listening test to assess the consistency between sounds and trajectories. The agreement was not expected to be one-to-one, and

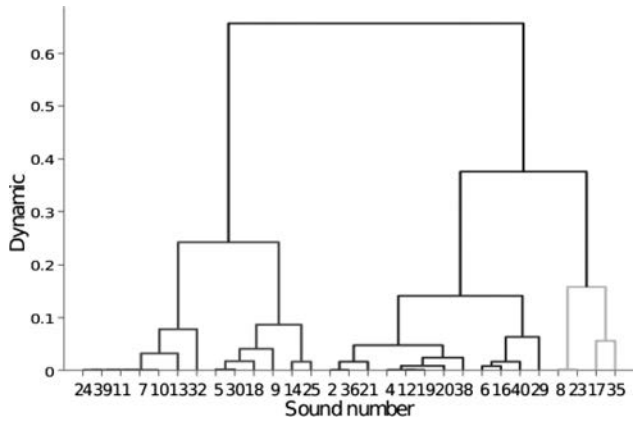

Fig. 9. Dendrogram for the mean *Dy* values (defined from the *Initial speed* and *Final speed*). At a threshold value of 0.3, three groups were distinguished, corresponding to the three dynamic levels: constant speed (left, sounds 24-25), and somewhat variant (middle, sounds 2-29), and greatly variant (right, sounds 8-35) speed.

we hypothesized that a given sound might be perceived as being coherent with several trajectories (and vice versa). We expected that at least the main perceptual categories previously identified would be reliably retrieved across subjects in the case of each variable. More specifically, with the *Shape* variable, we expected the subjects to be roughly consensual about the linearity of the trajectories and the presence or absence of regular or circular oscillations in the trajectory. In terms of the *Direction*, we expected at least the North, South, and Horizontal directions to be perceived. And we expected to find greater agreement between subjects at low *Oscillation frequency* values than at higher ones.

7.1 Control Strategy for Sound Synthesis

A sonification strategy based on physical considerations was used here so that the sound generated would correspond to the actual displacement of the sound source along a trajectory. This strategy is just one of many possible ones. Four main transformations are known to be required when simulating moving sound sources (i.e., pitch variations due to the Doppler effect, level variations depending on the source-listener distance, lowpass filtering to account for air absorption, and reverberation (see for instance Chowning [1971] on the implementation of these transformations)). We therefore applied these transformations to an initially stationary signal generated by a "sound texture" module based on several synthesis methods (additive, subtractive, and granular). Since the moving area of the point source was limited by the GUI window's size, the sound effects were too subtle to be clearly perceived under normal "physical" conditions. For instance, the Doppler effect is not perceptible at speeds below 9km/h (a frequency deviation of 8 Hz occurs in the case of a moving source emitting a 1 kHz sound signal perceived by a stationary listener) and a synchronization process with the sound file corresponding to normal physical conditions would require a much larger display window. To rescale the sound effects to the size of the display window, each transformation was therefore weighted by a coefficient C_{transf} with $transf = level, Doppler, absorption, reverb, elevation.$ The calibration of these coefficients is described in detail below. The level variation was simulated with a gain factor equal to

$$G_l = \frac{1}{1 + (SLD - 1) \times C_{level}},$$

where SLD is the source-listener distance. The pitch variations were due to the Doppler effect in the horizontal plane. This was simulated using a delay line and the delay time is defined by

$$D_t = \frac{C_{doppler} \times SLD}{c} + D_{t0},$$

where c is the sound speed and D_{t0} is the minimum delay time (the signal vector length). To simulate the air absorption corresponding to the SLD, a "shelf" low-pass filter with a fixed 500 Hz cut-off frequency and a gain factor of

$$G_f = \frac{1}{C_{absorption} \times SLD}$$

was used. Reverberation was simulated by a feedback delay network [Jot and Chaigne 1991] at a fixed reverberation time of 2 sec. The dry/wet ratio R_{wd} was controlled by

$$R_{wd} = \frac{1}{C_{reverb} \times SLD - 1}.$$

We added a sound effect to simulate displacements in the vertical plane by applying a linear pitch modification directly to the sound texture $P = C_{elevation} \times Z + P_0$, where P is the modified pitch, Z is the elevation (ranging from -1.5 to 1.5), and P_0 the reference pitch (in Hz). A source moving upwards or downwards was therefore simulated by an increasing or decreasing pitch, respectively. Note that the size of the trajectory was implicitly controlled via the SLD (under normal conditions) and the range of SLD variations depends on the Size variations.

Calibration. The coefficients C_{transf} were calibrated by five experts belonging to the research team. By "experts," we mean listeners who can anticipate the sound transformations liable to result from a given change in any of the synthesis parameters. The experts determined the magnitude of the transformation required to achieve, at best, sound effects coherent with a set of 10 trajectories representative of the interface possibilities: 8 shapes and 2 oscillating shapes with *Oscillation frequency* values of 2 and 8 cycles. The experts reported that many possible settings could be used to produce similar effects for a given trajectory, and the data revealed high variability between the experts' parameter settings (standard deviation on shapes and transformations was 1.29). This finding is most interesting, since it shows that the choice of synthesis strategy, which depends on the magnitude of each transformation, is not unique. Coefficient values were averaged across trajectories and experts to determine the final C_{transf} value for each transformation: $C_{level} = 0.7$ (attenuation equal to 16 dB for D_{max}); $C_{Doppler} = 1.5$ (which corresponds to a maximum delay time of 50 ms at the maximum distance allowed by the interface $D_{max} = 10m$); $C_{elevation} = 1.8$ (the range of pitch variations was 1 octave from $f_0 \simeq 150\ Hz$, where f_0 is the fundamental frequency of the sound (or the center frequency of a band-pass filter in the case of subtractive synthesis)); $C_{absorption} = 1$ (attenuation equal to 20 dB); $C_{reverb} = 2$ ($R_{wd} = \infty$ at $0.9D_{max}$). Note that the overall control strategy calibrated with these mean values was further tested and validated by the experimenter (cf., article Web page[5] for sound examples).

7.2 Stimuli

We constructed stimuli, based on the hypotheses put forward about the *Shape*, *Direction*, and *Oscillation frequency* variables (see introduction to Section 7). We used the same control interface as in the listening test (Figure 2) to generate trajectories, and recorded the visual animation displayed in the drawing window. We then generated the sound associated with each trajectory, using the previous

[5]http://lma.cnrs-mrs.fr/∼kronland/Motion.

control strategy (Section 7.1) with the same initial sound texture based on the granular looping of a recorded sound showing a rich spectrum. Note that this "texture" was not perfectly stationary, since the granular looping involved the superposition of 16 randomly selected portions (from 512 samples and Hanning windows) of an initial non-stationary sound. We obtained sounds directly from the synthesis tool and didn't apply further transformations (no loudness equalization in particular, since gain factors are fundamental parameters for simulating moving sound sources). The duration of the sound was synchronized with that of the source trajectory with a 40 ms linear start and end ramp. With all the stimuli, the average source-listener distance was set at 1 m and was estimated on the basis of the dimensions of the listener's head (around 15 cm in reality) displayed in the window. Note that due to the $1/SLD$ gain variations, a 1 m distance was the minimum allowed by the interface to avoid having diverging values. The variables that were not tested in the experiment were set to the following values to minimize their influence on the subjects' evaluation: *Randomness* was set at 0, *Size* at 0.5 (trajectories occupied around half of the screen), *Initial speed* and *Final speed* at 0.5 (i.e., constant speed), and *Back & forth* at 0.5. The sound duration was set at 4 sec. We tested separately the hypotheses about *Shape*, *Direction*, and *Oscillation frequency*. As regards the *Shape*, videos and the associated sounds were designed for the 8 shapes previously defined: *Wave*, *Line*, *Spring*, *Circle*, *Spiral*, *Hollow*, *Dome*, and *Bounce*. We set the *Number of cycles* at 4 for oscillating shapes and the *Angle* at -180°. As regards the *Direction*, videos and the associated sounds were designed for linear trajectories (*Line* shape, arbitrary choice) with 5 different directions: 1 Horizontal, 2 vertical ("Upward" and "Downward") and 2 slanted at angles of $\pm 25^\circ$ with respect to the horizontal line ("Slanting up" and "Slanting down"). As regards the *Oscillation frequency*, videos and the associated sounds were designed for wave-like trajectories (*Wave* shape, arbitrary choice) with 4 different *Number of cycles* values: 3, 4, 6, and 8 cycles. The corresponding *Oscillation frequency* values were 0.75, 1, 1.5, and 2 Hz, respectively. The *Angle* was set at -180°. We generated all the possible sound/video combinations corresponding to each variable (i.e., N^2 possible combinations with N elements). Sixty-four sound/video combinations were designed for the *Shape*, 25 combinations for the *Direction* and 16 combinations for the *Oscillation frequency*.

7.3 Subjects

Twenty subjects participated in the experiment (8 females and 12 males) and 13 were musicians (they had been playing an instrument for at least 10 years). Twelve of the subjects had participated in the previous experiment.

7.4 Procedure

Since the test did not require any specific calibration, subjects performed the test using their own computers and audio systems. A Max/MSP interface specifically designed for this study was distributed to the subjects, who were asked to take the test in a quiet room using headphones and to keep a constant sound level during the test. All the sound/video combinations (i.e., 105 combinations) were presented randomly at a single session. In each trial, subjects launched the sound by clicking on a "Play sound" button on the screen and the video was played synchronously with the sound. We asked the subjects to rate the degree of coherence between sound and video using a linear slider scaled from 0 ("they don't match") to 20 ("they match perfectly"). They could play the sound/video combinations several times and return to their previous responses at any time during the test.

7.5 Results

The listening test lasted for 15.3 min on average (std: 2.6 min), and each pair of stimuli was played 1.25 times on average. The mean degree of coherence of each sound/video combination was calculated from subjects' ratings. Results are presented in the three coherence matrices shown in

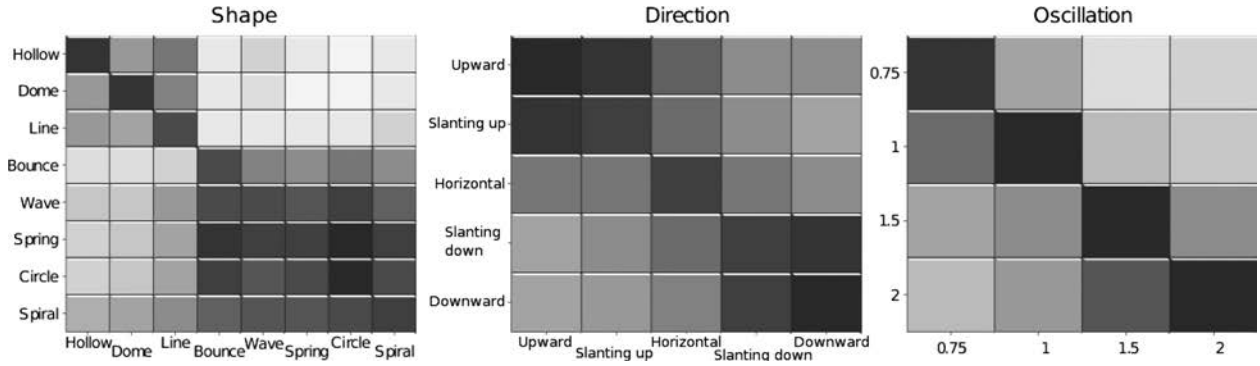

Fig. 10. Mean coherence matrices computed from the individual subjects' coherence matrices for the *Shape* (left), *Direction* (middle), and *Oscillation frequency* (right) variables. Rows correspond to videos and columns to sounds. The degree of coherence is given by the grey scale: the greater the coherence, the darker the color becomes.

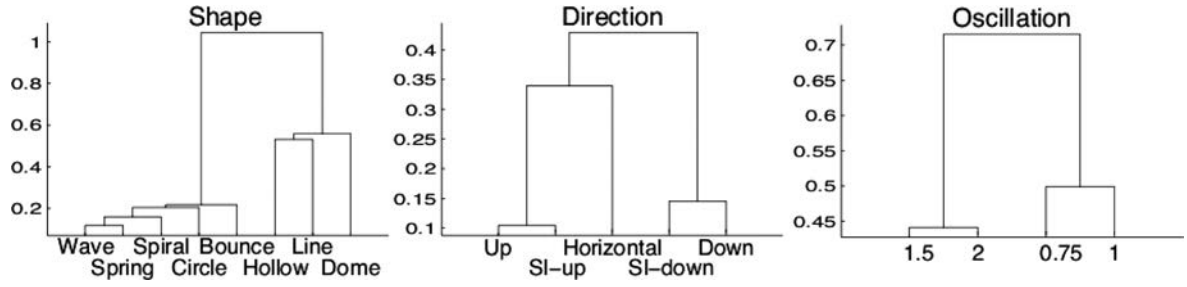

Fig. 11. Dendrograms for the *Shape* (left), *Direction* (middle), and *Oscillation frequency* (right) variables.

Figure 10. Note that all the cells were filled, since we tested all the sound/video combinations. At first sight, the diagonal part of the coherence matrices seems to be "darker" than either the upper or lower parts. With each matrix, we performed an analysis of variance to compare the mean rates obtained on the diagonal with those obtained in the non-diagonal cells: the diagonal was associated with significantly higher rates of coherence than the rest of the matrix for all three variables (*Shape*: $\chi^2(63, 1216) = 826, p < 0.01$; *Direction*: $\chi^2(24, 475) = 173.44, p < 0.01$; and *Oscillation frequency*: $\chi^2(15, 304) = 182.39, p < 0.01$). This result indicated that the matching sound/video combinations were judged to be the most coherent, which confirmed the validity of the present synthesis strategy. We then transformed the coherence matrices into dissimilarity matrices computed as distance matrices. For this purpose, each sound was represented by a point in an n-D space where the "coordinates" were given by the coherence rates in a column of the matrix. We therefore computed the distances between sounds using the Euclidean norm in an 8-D space for *Shape*, a 5-D space for *Direction*, and a 4-D space for *Oscillation frequency*. We performed a hierarchical clustering using the single linkage method (nearest neighbor distance) on each variable. The results presented in the dendrograms in Figure 11 are in line with our hypotheses. In the case of *Shape*, the linear category (*Line*, *Hollow*, *Dome*) was clearly discriminated from the other shapes at a threshold value of 0.7. By contrast, circular and regular oscillations were confused, which may be partly attributable to the fact that all the oscillating trajectories were produced with exactly the same frequency, phase, duration, and size. In the case of the *Oscillation frequency*, two clusters corresponding to high values (1.5 and 2 Hz) and low values (0.75 and 1 Hz) were detected at a threshold value of 0.6. Within these clusters, the

distance between sounds was smaller at high frequency values (between 1.5 and 2) than at low values (between 0.75 and 1). In the case of the *Direction*, we detected three clusters at a threshold value of 0.2: ascending ("Slanting up" and "Upward"), descending ("Slanting down" and "Downward") and Horizontal trajectories. Interestingly, slanting directions were confused with perfectly vertical directions, even when the angle of incidence was quite small. Results of the validation test were in line with our hypotheses, and confirmed that the method described in Section 4 yields general meaningful information about perceptually relevant aspects of motion evoked by sounds. Further inspection of the coherence matrices indicated a nonsymmetric tendency, which suggested the existence of a perceptual asymmetry in the assessment of sound/video combinations. To examine this asymmetry more closely, we compared the sound and video dissimilarity matrices, using a one-way multivariate analysis of variance. For this purpose, we computed dissimilarity matrices for videos using the same procedure as for sounds by taking the rates of coherence in one row of each matrix as the "coordinates" of the videos. Multivariate analysis of variance showed the existence of significant differences between sounds and videos in the case of *Shape*, whereas the differences were not significant with the other two variables (*Shape*: Wilk's $\lambda = 0.1335$, $\chi^2(28) = 48.33$, $p < 0.01$, *Direction*: Wilk's $\lambda = 0.8253$, $\chi^2(10) = 4.3$, $p = 0.786$, *Oscillation frequency*: Wilk's $\lambda = 0.478$, $\chi^2(6) = 10.163$, $p = 0.118$). With *Shape*, the fact that the distances between videos were significantly longer than those between sounds means that videos were more easily discriminated than sounds. Further studies are now required to determine whether these differences were due to the synthesis strategy adopted or to a fundamental difference between the perception of sounds and that of videos.

8. TOWARD A GENERIC CONTROL OF MOTION EVOKED BY SOUNDS

Designing an intuitive control strategy for motion evoked by sounds is still a great challenge facing both composers of music and sound designers. Our previous findings seemed to suggest some useful ways of addressing this issue. Having identified perceptually relevant variables, it was possible to draw up a typology for evoked motion based on the main categories of shapes defined (linear, circular, and regular), oscillating frequencies (low and high), randomness (none, low amplitude irregularities, high amplitude irregularities), direction (south, north, and horizontal), size (small, medium, and large) and dynamics (constant speed, somewhat variant, greatly variant). Some of these variables will presumably turn out to have similar perceptual correlates, since the present results showed, for instance, that a sound characterized by a high oscillation frequency can also be perceived as highly random. It should be mentioned that this typology is the result of one particular experiment with a limited number of stimuli and subjects and with an experimental protocol that has its limits, in particular due to the compromise between learning issues, and drawing accuracy that had to be done when constructing the GUI. This means that additional aspects of evoked motion might be revealed in alternative studies. Based on this motion typology, we proposed a generic control strategy for sound synthesis. The procedure was then designed on the basis of these considerations:

—This study provides evidence that drawings are a useful and intuitive means of describing motion evoked by sounds. It therefore seemed logical to develop a control system whereby evoked motion is described in the form of drawing, or more specifically, in the form of temporally sampled graphic trajectory. Actually, any input data generating continuous trajectories could be used, such as data recorded by motion capture systems (e.g., for orchestral conductors' gestures), pen tablets, and so on.

—Complex trajectories can be decomposed into elementary patterns, based on the categories defined in the present typology. These patterns can be subsequently associated with symbolic representations.

—The sonification strategy is far from being unique, and it is up to the users (e.g., sound designers) to define what sound effects should be associated with each variable.

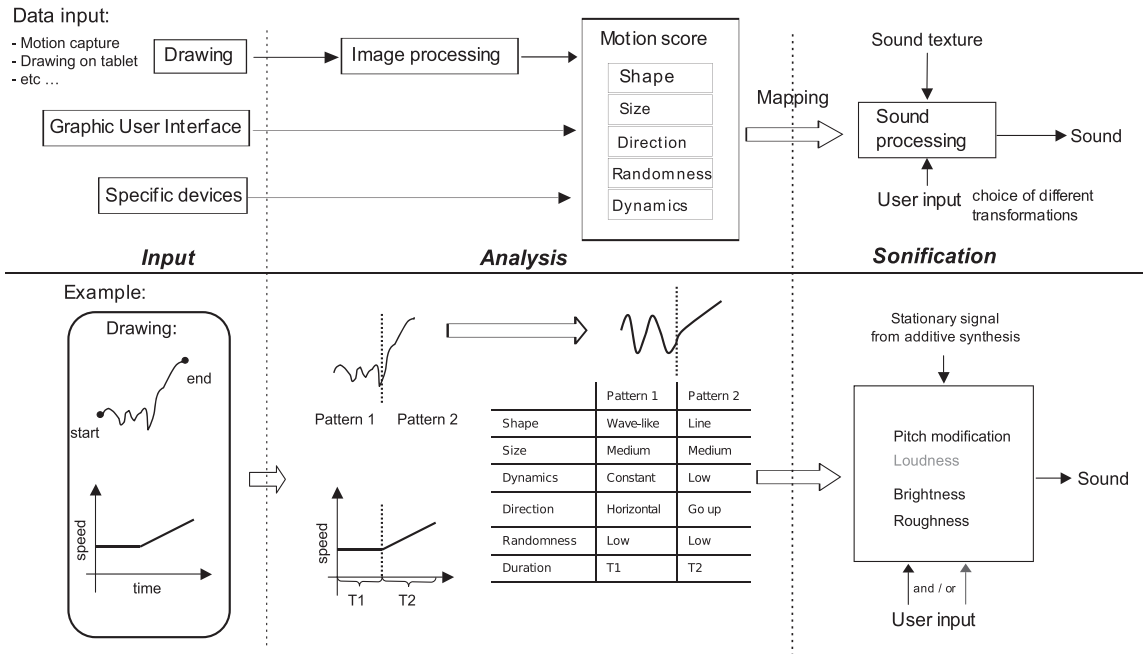

Fig. 12. Generic process for the control of motion evoked by sounds, consisting of three steps: analyzing drawings generated by data inputs, characterizing the basic patterns underlying the main perceptual variables, and the sonification process. The lower part of the figure shows an example of the data processing resulting from a drawing on a tablet.

The generic control strategy presented in Figure 12 involved three steps: a) generating continuous trajectories from any input device; b) computing a "motion score" giving a discrete (temporal sampling), symbolic representation of the continuous trajectory; and c) defining the sonification process in which the mapping strategy, that is, the relationships between the variables and the sound effects, is defined by the user. Input data is broken up into the basic patterns defined in the present typology of motion. Retrieving these patterns is similar to some extent to handwritten letter recognition, where the segmentation process consists in comparing parts of a written letter with previously defined letter shapes. Each segment is defined in terms of the previously defined typology (size, direction, oscillating frequency, etc.). Most of these characteristics could be estimated using automatic analysis methods (see Bishop [2006] for a review). In our case, this automatic process would require adjusting the parameter ranges based on the obtained categories for each variable. This step leads to a symbolic representation of motion, which is fairly similar in some respects to a musical score. Based on our findings, at least three different levels should be relevant for most of the parameters (e.g., small, medium, and large for Size; south, north, and horizontal for Direction, etc.). If a randomness amounting to less than 1% of the overall size is defined from the drawing analysis, we can state that it will not be necessary to include the randomness in the corresponding sound. The "motion score" could also be directly given by the GUI used for this current study. Due to the low satisfaction expressed by subjects about the dynamic control possibilities of the GUI, suitable means of exactly defining the dynamic profile of sound trajectories (for example, by drawing the profile "by hand," as suggested in the subjects' comments, see Section 5.1) should be further considered. For the Randomness, the data obtained enabled us to calibrate the corresponding slider more exactly, especially its range, since more than 95% of all the ratings were below half of the maximum value (cf., Section 6.5). Although three levels of control can

be defined for Size, synthesis experiments revealed that individual calibrations should also be taken into account, by allowing synthesizer users to specify their own range of size variations. Otherwise, if specific controls are needed for some trajectory parameters, their control can be mapped on separate devices. Both global control strategies defining for instance the Shape and Size of a trajectory, and more specific controls related to one particular GUI parameter such as the degree of Randomness could be envisaged. As far as the sonification process is concerned, it is up to the users to choose the mapping strategy. In the case of the physically-based control strategy presented in Section 7.1, the mapping between the trajectory parameters and the acoustic parameters can be defined as follows. For instance, the source-listener distance can be readily mapped to the intensity level or the reverberation rate. The amount of oscillation can be mapped to either periodic variations in the amplitude envelope or to the brightness or the parameters of a distortion system (with periodic variations in the distortion rate). In the case of more aesthetic applications such as musical compositions, less physically-based strategies would be worth envisaging. Depending on the control strategies adopted, a generic tool of this kind can be applied in the fields of sound design, virtual/augmented reality, sonification, music, and a wide range of other fields. Further studies are now required to improve the control possibilities available in the most widely used applications (for example, the control of *Size* should normally be independent of the reverberation rate). Our aim is now to define the relationships between drawing variations and signal variations, which will require developing a calibration process, to produce changes that are perceived linearly in response to any linear input.

9. CONCLUSION

The aim of this study was to characterize the general concept of motion evoked by sounds for synthesis and control purposes. The original experimental task, in which we asked the listeners to assess the motion evoked by sounds by answering a questionnaire and producing drawings, used a specially developed graphic user interface. The results obtained showed that drawings accurately described the motion evoked by sounds without any use of verbal descriptions. We developed this experimental procedure with synthesis perspectives in mind, and designed the user interface so that it could also be used for control purposes. To explore the various aspects of motion at both physical and more metaphoric levels, we used a specific class of sounds called “abstract sounds” as stimuli. It was assumed that these sounds would reduce the effects of sound source recognition and focus listeners’ attention on the intrinsic aspects of sounds evoking motion. Based on the listeners’ responses to the questionnaire and the analysis of their drawings, we identified some perceptually relevant variables accounting satisfactorily for the motion perceived in the sounds: Shape, Oscillating frequency, Direction, Size, Randomness, and Dynamics. Three main variables identified were then validated by performing a coherence test with synthetic sounds generated by a physically-based strategy for synthesizing moving sounds. These results constitute a further step towards defining an overall typology of motion evoked by sound and designing a generic control strategy which can be used with input data provided by motion capture systems of various kinds, giving a symbolic representation of the continuous trajectory. This process could be improved and refined by investigating other sounds and various synthesis strategies for sound transformation. Additionally, different GUIs with alternative drawing possibilities might also reveal new perceptually relevant aspects linked to perceived motion, and should therefore be investigated in future studies. In this study, we have also taken a step towards developing a general methodology for designing synthesis/transformation tools with intuitive control strategy. In particular, the use of abstract sounds and experimental protocols based on interactive tests (where subjects can adjust the control parameters directly) yielded some essential information about listeners’ strategies that could not have been obtained using other approaches. One of the main advances to which this study may lead is the identification of signal morphology mediating information about

motion. Due to the complexity of the nonstationary stimuli used in this study, defining perceptually relevant acoustic descriptors in the time-frequency domain is an issue which still remains to be solved from the signal processing point of view. The acoustic attributes used in the present physically-based control strategy constitute useful starting points. Another promising approach in which mathematical tools (such as Gabor multipliers) are combined with perceptual data is currently being developed [Olivero et al. 2010]. These approaches are currently investigated in the multi-disciplinary project MetaSon (http://metason.cnrs-mrs.fr/; ANR-10-CORD-0003) associating researchers in mathematics, signal processing, cognitive neurosciences, and acoustics.

REFERENCES

ALAIS, D. AND BURR, D. 2004. No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Res. 19*, 2, 185–194.

ANDERSEN, T. H. AND ZHAI, S. 2008. "Writing with music": Exploring the use of auditory feedback in gesture interfaces. *ACM Trans. Appl. Percept. 7,* 17:1–17:24.

ARAMAKI, M., BESSON, M., KRONLAND-MARTINET, R., AND YSTAD, S. 2011. Controlling the perceived material in an impact sound synthesizer. *IEEE Trans. Audio, Speech, Lang. Process. 19*, 2, 301–314.

ARRIGHI, R., ALAIS, D., AND BURR, D. 2006. Perceptual synchrony of audiovisual streams for natural and artificial motion sequences *J. Vision 6*, 3.

BEVILACQUA, F., MÜLLER, R., AND SCHNELL, N. 2005. Mnm: A max/msp mapping toolbox. In *Proceedings of the Conference on New Interfaces for Musical Expression (NIME '05)*. 85–88.

BISHOP, C. M. 2006. *Pattern Recognition and Machine Learning. Information Science and Statistics*. Springer, Berlin.

CARLILE, S. AND BEST, V. 2002. Discrimination of sound source velocity in human listeners. *J. Acoustical Soc. Amer. 111*, 2, 1026–1035.

CHEN, Y., SUNDHARAM, H., RIKAKIS, T., OLSON, L., INGALLS, T., AND HE, J. 2008. *Experiential Media Systems - The Biofeedback Project. in Multimedia Content Analysis: Theory and Applications*. Springer, Berlin.

CHOWNING, J. 1971. The simulation of moving sound sources. *J. Audio Eng. Soc. 19*, 1, 2–6.

DACK, J. 1999. Systematizing the unsystematic. In *Proceedings of the Arts Symposium of the International Conference for Advanced Studies in Systems Research and Cybernetics*.

DIXON, N. AND SPITZ, L. 1980. The detection of auditory visual disynchrony. *Perception 9*, 6, 719–721.

EITAN, Z. AND GRANOT, R. Y. 2006. How music moves: Musical parameters and listeners images of motion. *Music Percept. 23*, 3, 221–247.

FREMIOT, M., MANDELBROJT, J., FORMOSA, M., DELALANDE, G., PEDLER, E., MALBOSC, P., AND GOBIN, P. 1996. Les Unites Semiotiques Temporelles: Elements nouveaux d'analyse musicale Documents Musurgia Ed. Diusion ESKA. MIM Laboratoire Musique et Informatique de Marseille.

FRIBERG, A. AND SUNDBERG, J. 1999. Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *J. Acoustical Soc. Amer. 105*, 3, 1469–1484.

GAVER, W. W. 1993. What in the world do we hear? An ecological approach to auditory source perception. *Ecological Psychol. 5*, 1, 1–29.

GLASBERG, B. AND MOORE, B. 2002. A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc. 50*, 5, 331–342.

GOUNAROPOULOS, A. AND JOHNSON, C. 2006. Synthesising timbres and timbre-changes from adjectives/adverbs. In *Applications of Evolutionary Computing*, 664–675.

HOFFMAN, M. AND COOK, P. R. 2006. Feature-based synthesis: Mapping acoustic and perceptual features onto synthesis parameters. In *Proceedings of the International Computer Music Conference (ICMC)*.

HONING, H. 2003. The final ritard: On music, motion, and kinematic models. *Comput. Music J. 27*, 3, 66–72.

JOHNSON, M. L. AND LARSON, S. 2003. Something in the way she moves"–Metaphors of musical motion. *Metaphor and Symbol 18*, 2, 63–84.

JOT, J. M. AND CHAIGNE, A. 1991. Digital delay networks for designing artificial reverberators. In *Proceedings of the 90th Convention of the Audio Engineering Society*.

KACZMAREK, T. 2005. Auditory perception of sound source velocity. *J. Acoustical Soc. Amer. 117*, 5, 3149–3156.

KRONMAN, U. AND SUNDBERG, J. 1987. Is the musical ritard an allusion to physical motion. *Action Percept. Rhythm Music 55*, 57–68.

LARSSON, P. 2010. Tools for Designing emotional auditory driver-vehicle interfacesauditory displa*y*. Lecture Notes in Computer Science, vol. 5954, Springer, Berlin, 1–11.

LE GROUX, S. AND VERSCHURE, P. F. M. J. 2008. Perceptsynth: Mapping perceptual musical features to sound synthesis parameters. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

LIBERMAN, A. M. AND MATTINGLY, I. G. 1985. The motor theory of speech perception revised. *Cognition 21*, 1, 1–36.

LUTFI, R. A. AND WANG, W. 1999. Correlational analysis of acoustic cues for the discriminationn of auditory motion. *J. Acoustical Soc. Amer. 106*, 2, 919–928.

MALLOCH, J., SINCLAIR, S., AND WANDERLEY, M. 2008. A network-based framework for collaborative development and performance of digital musical instruments. In *Computer Music Modeling and Retrieval. Sense of Sounds*, 401–425.

MCDERMOTT, J., GRIFFITH, N. J. L., AND O'NEILL, M. 2008. Evolutionary computation applied to sound synthesis. In *The Art of Artificial Evolution*, Springer, Berlin, 81–101.

MERER, A., ARAMAKI, M., KRONLAND-MARTINET, R., AND YSTAD, S. 2010. On the potentiality of abstract sounds in perception research. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR'10)*. 207–219.

MERER, A., YSTAD, S., KRONLAND-MARTINET, R., AND ARAMAKI, M. 2008. Semiotics of sounds evoking motions: Categorization and acoustic features. Lecture Notes in Computer Science, vol. 4969, Springer, Berlin,139–158.

MIRANDA, E. R. 1998. Machine learning and sound design. *Leonardo Music J. 7*, 49–55.

OLIVERO, A., TORRÉSANI, B., AND KRONLAND-MARTINET, R. 2010. A new method for Gabor multipliers estimation: Application to sound morphing. In *Proceedings of the European Signal Processing Conference (EUSIPCO '10)*. 507–511.

PELLEGRINO, G., FADIGA, L., FOGASSI, L., GALLESE, V., AND RIZZOLATTI, G. 1992. Understanding motor events: A neurophysiological study. *Exper. Brain Res. 91*, 176–180.

RIECKE, B. E., FEUEREISSEN, D., AND RIESER, J. J. 2009. Auditory self-motion simulation is facilitated by haptic and vibrational cues suggesting the possibility of actual motion. *ACM Trans. Appl. Percept. 6*, 20,1, 20–22.

SCHAEFFER, P. 1966. *Traité des objets musicaux*. Editions du seuil.

SCHAFFERT, N., MATTES, K., AND EFFENBERG, A. 2010. A sound design for acoustic feedback in elite sports auditory display, Lecture Notes in Computer Science, vol. 5954, Springer Berlin,143–165.

SCHÖN, D., YSTAD, S., KRONLAND-MARTINET, R., AND BESSON, M. 2010. The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *J. Cognitive Neurosci. 22*, 5, 1026–1035.

SHOVE, P. AND REPP, B. 1995. Musical motion and performance: Theoretical and empirical perspectives. In *The Practice of Performance*, J. Rink, Ed., Cambridge University Press, 55–83.

STEINER, H.-C. 2006. Towards a catalog and software library of mapping methods. In *Proceedings of the Conference on New Interfaces for Musical Expression (NIME '06)*. IRCAM | Centre Pompidou, Paris, 106–109.

SUMMERFIELD, Q. AND MCGRATH, M. 1984. Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quwt. J. Exper. Psychol. 36*, 1, 51–74.

THIEBAUT, J.-B., HEALEY, P. G. T., AND BRYAN KINNS, N. 2008. Drawing electroacoustic music. In *Proceedings of the International Computer Music Conference (ICMC)*.

VOGT, K., PIRRÒ, D., KOBENZ, I., HÖLDRICH, R., AND ECKEL, G. 2010. PhysioSonic - Evaluated movement sonification as auditory feedback in physiotherapy auditory display. Lecture Notes in Computer Science, vol. 5954, Springer, Berlin,103–120.

WARREN, W. AND VERBRUGGE, R. 1984. Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *J. Exper. Psychol. 10,* 5, 704–712.

ZWICKER, E. AND FASTL, H. 1990. *Psychoacoustics, Facts and Models.* Springer, Berlin.

Received July 2011; revised August 2012; accepted August 2012