From Sound to Shape: Auditory Perception of Drawing Movements

Etienne Thoret, Mitsuko Aramaki,
and Richard Kronland-Martinet
Laboratoire de Mécanique et d'Acoustique UPR7051 CNRS,
Marseille, France

Jean-Luc Velay
Laboratoire de Neurosciences Cognitives UMR7291 CNRS,
Marseille, France

Sølvi Ystad
Laboratoire de Mécanique et d'Acoustique UPR7051 CNRS, Marseille, France

This study investigates the human ability to perceive biological movements through friction sounds produced by drawings and, furthermore, the ability to recover drawn shapes from the friction sounds generated. In a first experiment, friction sounds, real-time synthesized and modulated by the velocity profile of the drawing gesture, revealed that subjects associated a biological movement to those sounds whose timbre variations were generated by velocity profiles following the 1/3 power law. This finding demonstrates that sounds can adequately inform about human movements if their acoustic characteristics are in accordance with the kinematic rule governing actual movements. Further investigations of our ability to recognize drawn shapes were carried out in 2 association tasks in which both recorded and synthesized sounds had to be associated to both distinct and similar visual shapes. Results revealed that, for both synthesized and recorded sounds, subjects made correct associations for distinct shapes, although some confusion was observed for similar shapes. The comparisons made between recorded and synthesized sounds lead to conclude that the timbre variations induced by the velocity profile enabled the shape recognition. The results are discussed in the context of the ecological and ideomotor frameworks.

Keywords: biological motion, action-perception, friction sound synthesis

Supplemental materials: http://dx.doi.org/10.1037/a0035441.supp

The perception of movement induced by sounds is a widely investigated research topic involving researchers from diverse areas, ranging from physics to cognitive neuroscience. However, compared with other aspects of movement investigation, one aspect seems to have been accorded little attention, namely the auditory perception of biological movement. This is the subject of

the present investigation. In the following, we give an overview of former studies and theoretical frameworks that have dealt with the auditory perception of acoustical events and the perception and production of biological movements.

From an acoustical point of view, the auditory perception of movement induced by one or several sound sources was initially addressed by investigating the auditory consequence of an actual displacement of a physical source. It is well known for instance, that the perception of a passing source is related to the physical phenomena occurring during sound propagation. Hence, a sound emanating from a source at a distance from the listener is less intense, more band-limited and more reverberant than a sound from a nearby source. As the source approaches the listener, its intensity and bandwidth increase in parallel, and its level of reverberation decreases. In addition, frequency shifts, due to the Doppler effect, occur when there is a change in the speed of the moving source relative to the listener. Therefore, the overall variation of the generated sound is determined by combining variations in intensity, frequency, bandwidth, and reverberation. Such sound can be reproduced satisfactorily under monophonic playback conditions, which implies that specific sound morphologies, mainly related to the timbre, can evoke movement.

An audio researcher, a sound engineer, or a musician is usually required to reproduce such types of effects by manipulating the intrinsic characteristics of the sound. These characteristics were widely exploited for sound modeling purposes and music composition. Therefore, it is reasonable to consider that evoked movements not only concern the physical displacement of a source but

may also refer to more metaphoric notions of movement, like musical movements, for instance. Indeed, music analyses have led to semiotic descriptions of perceived movements in musical pieces (Frémiot et al., 1996). Other studies have focused on the nature the relationships between music and motion in general (Honing, 2003; Johnson & Larson, 2003). To investigate sound attributes related to the general concept of motion evoked by sounds, Merer, Ystad, Kronland-Martinet, and Aramaki (2008) conducted a free categorization task that used monophonic abstract sounds, that is to say, sounds whose source was not readily identifiable. On the basis of this, they identified the main movement categories such as "Rotate," "Pass by," or "Fall down," which were associated to different acoustic descriptors, such as the ratio of frequency modulation or the amplitude modulation for instance. Merer et al. further addressed the perceptual characterization of these evoked motions by studying the drawings produced by a group of subjects using a purpose-made graphical user interface while listening to sounds (Merer, Aramaki, Ystad, & Kronland-Martinet, 2013). Based on an analysis of the drawings, some perceptually relevant variables accounting satisfactorily for the motion perceived in the sounds were identified.

From a theoretical perspective, according to the ecological theory of perception, the acoustic properties that carry the information for sound source identification are known as invariants. Originally introduced for vision, Gibson (1966) defined the invariants as the properties of the environment that don't vary, and thus, reveal a structure in a sensorial flow enabling perception and action. Following this ecological perspective, Gaver (1993b) proposed that perceiving a sound event is more than just a pattern matching with memorized representations and that sound inherently conveys consistent information about the physical world through invariant acoustic features, such as a temporal pattern or a spectral relationship, contained in the acoustic flow. More precisely, the information that specifies the nature of the sound source is known as a *structural invariant* and the information specifying the type of change or the action involved is referred as the *transformational invariant* (McAdams, 1993).

Concerning structural invariants, many studies have investigated the relations between physical characteristics of objects and the perception of the sound produced when they are impacted (e.g., Giordano, Rocchesso, & McAdams, 2010; Grassi, 2005; Grassi, Pastore, & Lemaitre, 2013; Kunkler-Peck & Turvey, 2000; McAdams, Chaigne, & Roussarie, 2004). Some of these studies revealed certain physical characteristics linked to dispersion and dissipation that are important for the recognition of a sound source. It has been shown, for instance, that sounds contain sufficient information to enable one to discriminate the material of impacted objects (Aramaki, Besson, Kronland-Martinet, & Ystad, 2011; Klatzky, Pai, & Krotkov, 2000; Wildes & Richards, 1988), and, to a certain extent, to recognize their shapes (Carello, Anderson, & Kunkler-Peck, 1998; Lakatos, McAdams, & Caussé, 1997). In particular, Giordano and McAdams (2006) evaluated the effect of the size of an object on the perception of the material for an impact sound and identified robust acoustical descriptors that explain material identification.

Concerning transformational invariants (Gaver, 1993a), they are related to the actions carried out on a given object. For instance, Warren and Verbrugge (1984) showed that based on the rhythm of a series of impacts contained in a sound, it is possible to predict if a glass had broken or bounced. Li, Logan, and Pastore (1991) revealed that one is able to recognize the gender of a person walking merely by listening to the footstep sounds produced, and more precisely that this was due to the differences in spectral peaks and the contribution of high frequency components for the feminine gender. Repp (1987) concluded that the sound of two hands clapping was sufficient to imagine the spatial conformation of the two hands.

For its part, the structural invariant of an object can be recognized even if it is associated with different transformational invariants. Hence, a ball is still recognized as such even if it is submitted to different actions like bouncing or rolling. Similarly, several studies demonstrated that the sound produced by a rolling ball could reveal its perceived size or velocity (Houben, Kohlrausch, & Hermes, 2004, 2005). Note that a study by Lemaitre and Heller (2012) highlighted that our auditory system is better tuned to recognize an object's action than its material based on the sounds produced by different interactions: rubbing, rolling, bouncing, or impacting.

In the ecological perspective, the concept of invariant is formalized with the notion of affordances, which can be defined as the ability of an object, here a sound, to evoke its use. In the case of sounds, the extracted invariants afford the potential actions that enable recognition and categorization of an action. Therefore, links between perceptual abilities and actions, here concerning auditory perception, are primordial (Castiello, Giordano, Begliomini, Ansuini, & Grassi, 2010).

Although fully compatible with the ecological approach described above, these abilities were also regarded as arising from cognitive processes involving the concept of representation. The case of speech perception is an appropriate example of such a sensorimotor coupling between the auditory perception of an event, and the motor representation that is inferred. Indeed, motor theory of speech perception suggests that we do not perceive sounds exclusively as auditory information, but that we perceive it as potential intended phonetic gestures (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattilngly, 1985). It has been shown that we have learned to bind the sounds produced by objects and modulated by actions with all their other perceptual properties in order to create a unified percept (Hommel, 2004). Recent neuroscientific research has shed light on brain mechanisms underpinning such percepts that might arise from multimodal neurons coding of both movements and their sensory consequences. Especially, it has been shown that mirror neurons in the monkey ventral premotor cortex discharge when the animal performs a specific action, but also when it hears the corresponding action-related sound without seeing the action in question (Kohler et al., 2002). However, the movements have to respect the relevant rules of production, otherwise the generated sound can lead to a misinterpretation of the action. More recently, Young, Rodger, and Craig (2013) discussed auditory-motor relations involved in the real time reproduction of walking sounds in the ideomotor framework. This perspective considers cognitive representations as a structural coupling between perceptions and actions (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1997). In Young et al. (2013), they proposed that the ability to synchronize our walking with a walking sound is made possible thanks to a common neural representation of the perceived and the generated action with regards to recent models (Cisek & Kalaska, 2010). As such, this novel approach tends to reconcile direct and indirect perspectives to perception and action (Norman, 2002).

In the present work, we focus on the auditory perception of a particular type of movement belonging to a category of biological movement, that is, drawing movements, which are, essentially, specific to humans. Because drawings generate a visual trace, but no dominant sound, these movements are far less connected to audition than they are to vision. Indeed, the quality of drawing is judged, not by its sound, but by its visual trace. However, if we listen carefully to the sound produced during drawing, we can hear the pen friction against the paper, especially when the surface is rough. However, given that it is unusual to pay attention to such sounds, it seems, a priory, very difficult, to say the least, to try to infer from these sounds what is drawn or written. Yet, these sounds are more than just noise. Due to the friction between the pen tip and the paper asperities, the kinematics of the drawing movement produce timbre variations in the sound that may, to a certain extent, enable the recognition of a specific movement produced by the writer. In particular, we wanted to find out if it would be possible to infer by ear what is drawn based, only, on the friction sounds generated. Our aim was thus to identify acoustic cues that reflect the movements underlying the drawing action and to ascertain the extent to which this information allows us to infer the characteristics of the drawn shapes.

These issues were investigated in three experiments. As a first step, we verified whether subjects were able to associate timbre variations in a friction sound to kinematic variations produced by the gesture during the drawing process. To this end, we began by consulting a number of seminal studies from Viviani and colleagues who carried out extensive investigations of the production and perception of biological movements in both the visual and sensorimotor modalities and, particularly, the relationship between graphical movements and geometric shapes. They highlighted the link between the velocity of drawing movements and the curvature of the drawn shape (Viviani & Terzuolo, 1982) and proposed a power law relation between the angular velocity of the pen and the curvature of the drawn trajectory (Lacquaniti, Terzuolo, & Viviani, 1983; Viviani & Flash, 1995; Viviani & McCollum, 1983).

In our first experiment (Experiment 1), we adapted a protocol of Viviani and Stucchi (1992) to the auditory perception of friction sounds. The subjects were asked to manipulate the power law modulating the friction sound (simulating the sounds produced by the pen) so as to imitate as closely as possible a sound evoking a natural and fluid drawing movement. This experiment required the use of real-time generated, synthesized stimuli. Our first hypothesis was, therefore, that timbre variations of a natural and fluid gesture should be recognizable through the exponent value of the power law. And thus, such timbre variations could be considered as an auditory transformational invariant enabling the recognition of some human gestures. As a second step, we speculated that, if a friction sound can indeed evoke a gesture, one might be able to visualize, to a certain extent, the geometric shape that has been drawn on the basis of the friction sound only. We thus asked subjects to associate friction sounds to given visual shapes in a series of experiments (Experiments 2 and 3); both recorded and synthesized sounds were evaluated. Using synthesized stimuli made it possible to focus on a single gesture parameter, namely the velocity profile, and to determine the extent to which this parameter is relevant for a sound/shape association task. In Experiment 2, the shapes were assumed to be easily distinguishable, both from a perceptual point of view and from the kinematics of their

underlying drawing movements. In Experiment 3, we included sounds and shapes that were assumed to be perceptually more similar to each other to assess if similar velocity profiles imply lower recognition rates.

The results of the experiments will be discussed from an ecological perspective to determine whether the velocity profile can be considered as a relevant transformational invariant of drawing movement. In addition, the results will be discussed in light of the ideomotor framework proposed by Young et al. (2013), in particular in terms of representations as a structural coupling between the sensorial flow and its processing. Before presenting these three experiments, we will describe in the following section, the general principles of the synthesis of friction sounds.

Synthesis of Friction Sounds

Synthesis is an appropriate tool for investigating the perception of the underlying gesture evoked by friction sounds. Gaver (1993b) and Van den Doel, Kry, and Pai (2001), for example, proposed a simple, physically informed model for generating synthetic friction sounds from given velocity and pressure profiles. This model simulates the physical sound source resulting from successive impacts of a plectrum on the asperities of a surface, in the context of our study the movement of a pencil on paper. The surface roughness is modeled by a noise reflecting the heights of the surface asperities. A common model for such noise is the fractal model whose spectrum is: $S(f) = 1/f^b$, where b can range between 0 and 2 and f is the frequency (Van den Doel et al., 2001). If $b = 0$, the noise is white and corresponds to a rough surface, and as the value of b increases, the surface smoothens. In the following, we set $b = 0$. The sound was then generated either (a) by reading the noise (stored in a wavetable) which velocity is linked to the velocity profile of the pencil rubbing against the paper (and therefore to the velocity of the gesture); or (b) by lowpass filtering this noise with a cut-off frequency that varies according to the velocity profile of the pencil. This latter method is computationally more efficient. The mapping between the velocity of the pen (in $cm.s^{-1}$) and the synthesis model is arbitrary.

In addition to the velocity, other gesture parameters such as the pressure or the angle of the pencil could be considered in a more sophisticated model. For instance, the pressure of the pencil could be related to the intensity of the sound. However, because the focus of the current study is the velocity profile, the pressure is kept constant throughout the present studies. Thus, only the velocity profile of the gesture was used to control the production of the friction sound.

Experiment 1

As a first approach to investigating our capacity to recognize specific gestures through sounds, we based on a protocol of Viviani and Stucchi (1992) on the perception of visual biological motion. They asked subjects to adjust the velocity of a visual dot until they perceived its displacement as being uniform (i.e., a perceived constant speed). The movement of the dot was constrained by the power law expressed as:

$$v_t = KC^{-\beta} \tag{1}$$

where v_t is the tangential velocity, C the radius of the trajectory curvature and K, the velocity gain factor linked to the overall movement speed (Viviani & Terzuolo, 1982; Lacquaniti, Terzuolo,

& Viviani, 1983; Viviani & McCollum, 1983; Viviani & Flash, 1995). During the task, subjects were unaware that they were adjusting the exponent β of the power law. Results revealed that, to perceive a uniform visual displacement, the subjects adjusted the exponent by an average value of one third. This value corresponded to the actual velocity profile of the physical movement used during the drawing production. Conversely, when the exponent was null (i.e., constant speed regardless of the curvature, corresponding to a uniform physical movement), the movement of the dot was perceived as accelerating in the curved parts of the trajectory. In the following, we will refer to Equation (1) with β = 1/3 and call this relation the 1/3 power law. Note that this power law optimizes motion smoothness by minimizing the rate of change of acceleration (the jerk), as was previously shown by numerical simulation and mathematical analysis (Richardson & Flash, 2002; Viviani & Flash, 1995).

In the present study, we designed an experiment similar to that of Viviani and Stucchi (1992) in which subjects had to implicitly act on the exponent value of the power law within the auditory modality. In particular, the moving visual spot that described the trajectory, which was used in Viviani and Stucchi's study, was, here, replaced by a synthetic friction sound modulated according to the velocity profile. Although Viviani and colleagues aimed at investigating the perceptual relationship between the two visual variables, that is, the kinematics and the curvature, it should be noted that our goal was to investigate whether the manipulation of one variable related to the gesture's kinematics (determined by the velocity profile) allowed the evocation of a natural and fluid gesture. In particular, we assumed that the specific sound variations caused by the 1/3 exponent are recognizable through the auditory modality. If our hypotheses proved correct, this would imply that the timbre variations of the sound convey perceptual information about the physical movement, that is, the pen accelerating over the straight sections of the traced strokes and slowing down in the most curved sections.

Method

Subjects. Twenty participants took part in this experiment, 3 women and 17 men. Their average age was 29.42 years (*SD* = 12.54). Before participating in this experiment, none of the subjects were familiar with the topic being investigated.

Stimuli. Friction sounds were synthesized using the previously described friction model. To avoid evoking specific shapes, we considered friction sounds associated with pseudorandom trajectories generated from the trajectories of a moving point $(x(t), y(t))$, defined on the basis of the following parametric functions:

$$\begin{cases} x(t) = A_0 \sum_{k=0}^{3} a_k \sin(\omega_{x,k} t) \\ y(t) = B_0 \sum_{k=0}^{3} b_k \sin(\omega_{y,k} t) \end{cases} \quad (2)$$

A new set of parameter values for the above equations was randomly computed every 15 s (arbitrary choice), with the exception of the constant values $A_0 = 7$, $B_0 = 5$, $a_0 = 1$, and $b_0 = 1$. In particular, the values of $\omega_{x,k}$ and $\omega_{y,k}$ were randomized between 0 and 0.6 Hz and a_k and b_k were randomized between 0.5 and 0.9. Hence, the movement of the point was

predictable for only 15 s. The velocity profile corresponding to this pseudorandom trajectory was then computed from the power law expressed in Equation (1), where $K = 10$ m.s^{-1} and the curvature is defined by the following expression:

$$C(t) = \frac{|\ddot{x}\dot{y} - \dot{x}\ddot{y}|}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}}$$

where x and y were defined in (2), the dot and double dot represent the first and the second time derivative, respectively. The corresponding friction sound was then synthesized in real-time according to the exponent value of the power law, which was modified directly by the subjects during the experiment. Examples of stimuli for four exponent values (β = 0; β = 0.33; β = 0.7; β = 0.9) are available in supplemental material online.

Task and procedure. Participants were seated in front of a computer screen in a quiet room. Sounds were presented through Sennheiser HD-650 headphones. The experiment began with a 2-trial training session, followed by a 6-trial session constituting the formal test. For each trial, a pseudorandom trajectory was computed, and the corresponding friction sound was synthesized in real-time. The trajectory varied across trials and subjects. The trajectory was not displayed to the subjects and only the sound was presented continuously during the trial.

Subjects were asked to imagine that the sound they heard was produced by someone drawing a random shape (such as a scribble) on a rough surface. They were then asked to modify the sound using two assigned buttons ("<" and ">" presented on the computer screen) until they arrived at a sound that they judged the most "natural" and "fluid," according to a human gesture. The graphical interface was designed with the real time software Max/MSP.[1] The subjects were unaware of how their interaction with the buttons modified the sound and they also did not know that they were actually adjusting the exponent β of the power law ("<" and ">" buttons corresponded to decreasing and increasing the β values, respectively). Subjects were advised to take their time when listening to the sound and to explore the full range of values with the buttons during the adjustment process. The exponent values ranged between 0 and 1.0816 in steps of 0.0416. The initial exponent values were randomized at each trial. No time constraint was imposed. For each subject, one exponent value for each of the six trials was collected.

Results

The exponent values were averaged across trials, first for each subject, and then across subjects. An average value of 0.361 (*SD* = .084) was found. A comparison of this mean to one third was then performed by means of a one-sample two-tailed *t*-test. For statistical analyses, effects were considered significant if the *p* value was equal to or less than .05.

No significant differences between the mean values of the distributions were found, $t(19) = 1.53$; $p = .14$; $d = .329$. An analysis of the results according to the initial values of the exponent was performed to evaluate a possible ascending or descending effect on the subjects' performance. A classical effect of ascending

and descending threshold was observed and a significant correlation between the initial and the final values was found ($r = .34$; $p < .05$). As proposed in Carlyon et al. (2010), we performed a complementary analysis to confirm that exponent values had actually been adjusted to values close to one third. Two groups of final exponents were considered: trials in which the initial values were greater than one third and those in which they were lower than one third. Out of all the trials, 70 initial values were higher than one third and 50 were lower. For ascending thresholds, the mean value of the final exponent was .31 (95% CI [0.28; 0.34]); for the descending thresholds, the mean value of the exponent was .39 (95% CI [0.36; 0.43]). Moreover, we compared the standard deviations of the initial and final values with one-tailed two-sample t tests to check whether subjects adjusted the final exponents to a given value, so reducing their dispersion. The tests revealed that the final standard deviations were significantly lower than the initial ones, $t(19) = -7.86$; $p < .05$, confirming that the subjects actually carried out an adjustment.

Discussion

In this experiment, the subjects had to act on the sound they heard to render the evoked human gesture as "natural" and as "fluid" as possible. They were told that the sound was produced by someone drawing a shape with no specific geometry on a sheet of paper using a pen. The sound was generated by synthesis from the velocity profile expressed in Equation (1) and could be adjusted according to a hidden variable corresponding to the exponent β of the power law. Indeed, it is worth noting that the subjects were unaware of the acoustical characteristics of the sound they were adjusting with the control buttons.

The main finding of this first experiment is that, when asked to adjust a sound to evoke a "natural" and "fluid" graphical movement, subjects adjusted the timbre variations so that the velocity profile matches the 1/3 power law. In other words, the timbre of the sound of a moving pen appears to vary in accordance with the kinematic rule governing real graphical movements (Viviani & Terzuolo, 1982). This result reveals that we are able to imagine a human natural gesture, that is, biological movements, from the timbre variations of the sound in accordance with previous findings linked to the recognition of acoustic events through specific acoustical patterns, so-called transformational invariants. In other words, if one adopts a sensorimotor perspective, our perceptual processes are shaped by our motor competencies making us able to recognize such events and to interact with them (Viviani, 2002).

In this first experiment, subjects knew that the evoked movement corresponded to scribbles, and they were not required to make inferences regarding the actual shapes drawn. Because the 1/3 power law linking the velocity of the gesture and the curvature of the drawn trajectory was relevant for the evocation of natural human gestures through sounds, it may be possible, to a certain extent, to evoke a given trajectory from the friction sound and to make associations between the drawn shapes and friction sounds. Indeed, the association process may be based on the fact that higher velocities are due to smaller curvatures and conversely. Although only one variable was manipulated for the synthetic sound, that is, the velocity profile, the implicit knowledge of the 1/3 power law should enable an association between a visual depiction of the drawn shape and the friction sound. This was the focus of our second and third experiments.

Experiment 2

In drawing, a shape can be described both by its visual geometry and, in motor terms, by the movement required to trace it. Freyd (1983a; 1983b) first demonstrated that a motor representation is intrinsically linked to the perception of the visual shape. In particular, she found that readers use motor knowledge when decoding static, hand-written material. Furthermore, a close functional relation between the visual shape of a character and the corresponding graphic movement has been established: It was found that looking at graphic shapes activates cortical motor processes if the subjects already know how to draw these shapes (James & Gauthier, 2006; Longcamp, Anton, Roth, & Velay, 2003; Longcamp, Tanskanen, & Hari, 2006). Longcamp et al. (2008) assumed that the specific movements used to write a novel graphic shape are memorized and are, furthermore, involved in its subsequent visual recognition. In line with these results, Viviani and Stucchi (1989) established that a perceived shape, described by the displacement of a punctual spot, can be accurately deduced from the kinematics of the moving spot. Furthermore, they highlighted cases in which the perception of a moving spot could interfere with the perception of the resulting shape. For instance, when a point is moving along an elliptic trajectory with a small eccentricity at a constant speed, that is, without respecting the 1/3 power law, the perceived ellipse is modified and the resulting shape was assimilated as a circle.

By contrast, no sound is naturally associated with a visual shape because a shape is an abstract object and therefore cannot be readily associated with a physical sound source. In Experiment 1, we confirmed the relevance of the 1/3 power law in making the sound coincide naturally with a natural and fluid human gesture. Because this law links the velocity of the movement and the curvature of the drawn shape, we can assume that the pen trajectory was processed implicitly in the previous task.

In this second experiment, our aim was to investigate further whether friction sounds can also inform about the drawn shapes. We asked subjects to associate a given friction sound to a static visual shape they imagined was drawn during the sound production. Subjects had to choose between four visual shapes corresponding to the actual drawn ones. The rationale for this association task was to limit the number of possible shapes to be identified. Based on the results of Experiment 1, we hypothesized that the relevant information required by the subjects to associate the sounds with the shapes is contained in the velocity profile, and that, according to the 1/3 power law, we are able, to a certain extent, to recover geometric information about the drawn shapes from sounds. On this basis, we began by recording the natural sounds produced by a writer for different elementary drawn shapes. Indeed real friction sounds contain fine modulations that may vary, for example, according to the pen angle with respect to the paper. In addition, for downstrokes and upstrokes, variations of this angle cause variations in the pen pressure. To test our hypothesis and to investigate the relevance of the velocity profile alone, we also considered synthetic friction sounds that depended solely on the velocity profiles collected from a writer.

Method

Subjects. Twenty participants took part in the experiment: 9 women and 11 men. The average age was 30.65 years ($SD =$ 13.11). None of the subjects were familiar with the topic of the study before the test. Eight subjects had also participated in Experiment 1.

Stimuli.

Static shapes. Preliminary informal tests revealed that distinguishing sounds corresponding to shapes with cusps from those without cusps was perceptually easy because cusps gave rise to a discontinuity in the sound produced, that is, a stop, that was highly perceptible. Based on these observations, we assembled a corpus of four shapes, two of which had no cusps (circle, ellipse) and two of which had cusps (arches, line). Shapes without cusps included a circle and an ellipse that differed by their eccentricity. Those with cusps were arches and line that differed by their cusp positions: the arches cusps were located periodically along the paper, and the line cusps were located at the same positions, that is, at the extremities. Figure 1 presents the four shapes.

Recorded sounds. Recording sessions took place in a quiet recording studio. A member of the staff drew the four shapes as fluidly as possible on a Wacom graphic tablet–Intuos 3. Excepted for the arches, he was asked not to lift the pen from the tablet for 25 s to make sure that several periods were drawn for each shape; for the arches he was asked to start drawing from the left side of the sheet until he reached its end, and then to start again at the initial position. The movements of the writer were recorded on the tablet at a sampling frequency of 200 Hz and with a spatial precision of 5.10^{-3} mm.

In addition, monophonic recordings of the sounds produced during the drawing sessions were made at a sampling frequency of 44100 Hz with a cardioid Neumann-KM84i microphone positioned about 30 cm above the tablet. Sequences of these recordings were selected for the experiment based on a geometrical and temporal examination of the writer's performances. In particular, we selected sequences of similar durations (about 5 seconds) during which the writer executed the shapes in a regular manner
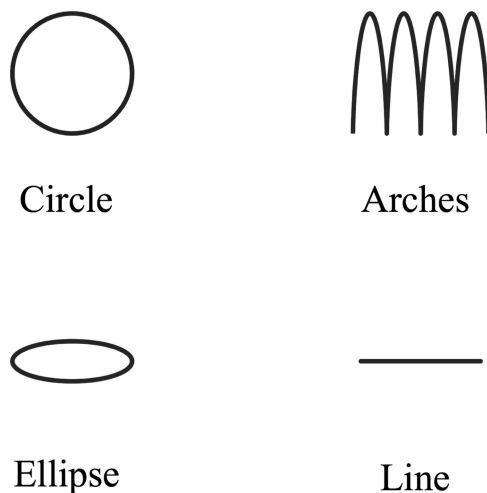
Table 1

Geometrical and Temporal Characteristics of the Performances Chosen for the Stimuli of the Experiment 2

Shape	Length (cm)	Duration (s)
Circle	62.5	5.2
Ellipse	89.32	5.8
Arches	87.11	5.1
Line	88.7	5.2

Note. The circle length, which corresponds to the drawn length used to synthesize stimuli, appears to be shorter than the three other shapes. However, the duration of the associated sound is almost the same as the three other ones because the mean velocity is lower for this shape. The recorded and synthesized stimuli of Experiment 2 are available in supplementary material online.

for a given number of periods. For ellipses, arches, and lines, the selected sequences corresponded to four periods. For circles, only two periods were considered because the mean duration was about twice that of the three other shapes. Table 1 summarizes the characteristics of the performances chosen as stimuli for the experiment. The segmentation was performed with a windowing function of 10 ms to avoid clicks at the beginning (fade in) and at the end (fade out) of each sound sequence. All the sounds were normalized at -3 dB. The velocity profiles corresponding to the selected sequences were computed from the tablet data for the four shapes. Figure 2 presents one period of the velocity profile of each shape. They are normalized in amplitude and low-pass filtered at 10 Hz. The zero-crossings corresponding to the cusps are marked

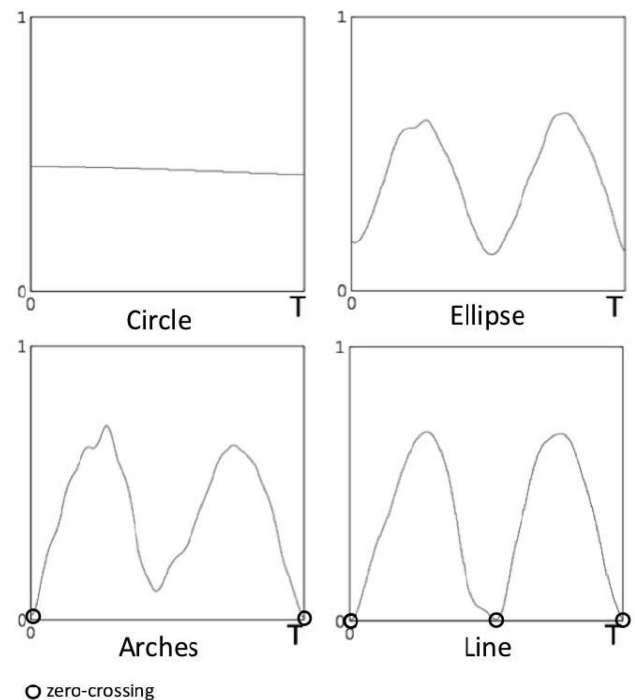*Figure 1.* The four shapes used in Experiment 2. For arches, four periods are presented.

Figure 2. One period of the velocity profiles of the four shapes used in Experiment 2. The zero-crossings are marked with black circles. The velocity profiles are low pass filtered at 10 Hz and normalized in amplitude.

with black circles. It is noticeable that the shapes have very different velocity profiles, which led us to hypothesize that they should be distinguishable from a perceptual point of view.

Synthesized sounds. The friction sounds were synthesized using the same friction model as in Experiment 1. They were generated from the velocity profiles collected from the writer during the recording sessions for each of the four shapes.

Task. Participants were seated in front of a computer screen in a quiet room. They listened to the sounds through Sennheiser HD-650 headphones. The graphical interface was designed with the software Max/MSP. The experiment comprised a session of eight trials, that is, four trials with the recorded sounds and four trials with the synthesized sounds. Trials were randomized across participants. In each trial, four shapes and four sound icons were displayed on the computer screen. Subjects were asked to associate each shape to the sound they believed was produced when the shape was drawn. Each sound could be associated to one shape only and vice versa. In practice, the four shapes were displayed on the right-hand side of the screen and the four icons representing the sounds on the left hand side. The shapes were always displayed in the same order. The sounds, which were represented by identical icons, were randomly permuted for each trial. Subjects carried out the association task by moving the sound icon next to the corresponding shape with the computer mouse. They were informed that the sounds had been recorded from a writer drawing each of the four shapes fluently, without lifting the pen. No time limitation was imposed and the subjects could listen to the sounds as often as they wished. For each trial, four sound/shape associations were collected.

Data Analysis

The associations were rated as 1 if the sound was associated to the correct shape and as 0 otherwise. Then, for each subject and each type of sound (synthesized and recorded), the values were averaged across trials to compute an association matrix containing the scores of the association task. In the following, we define the score of success, that is, the score of a correct sound/shape association, as the diagonal values of an association matrix.

To examine the subjects' errors, we compared the scores of the associations between a given sound and the four shapes, and between a given shape and the four sounds. Paired two-tailed *t*-tests were performed on the score of success and the six other scores of association. In addition, for each score, a test of conformity to a standard was carried out with threshold corresponding to 25% of success (equal to the chance threshold).

To evaluate if the task was executed similarly for recorded and synthesized sounds, we carried out a global comparison of the association matrices of the two sound types. To do this, we computed the cophenetic distances between shapes for each matrix and we performed a Pearson's correlation test on these distances. In practice, the analysis was carried out as follows: For each type of sound, we determined, first, a so-called "dissociation" matrix, D^*, defined by: $D^* = 1 - S^*$, where S^* is the symmetrized version of the average association matrix S^* A pairwise "distance" matrix, D, was determined from D^* by choosing the euclidean metric. A hierarchical clustering analysis of D (complete linkage) was then carried out. The cophenetic distances were computed from the resulting dendrogram and were assembled into a vector. The cophenetic distances

corresponded to the distances between the shapes estimated at all nodes of the dendrogram.[2] Then, to compare the matrices, the two vectors of cophenetic distances were submitted to a Pearson's correlation test.

Finally, a more precise comparison of the scores of recorded and synthesized sounds was carried out by performing two-sided Wilcoxon signed-ranks test on the rates of success for each shape with the type of sound (recorded vs. synthesized) as a factor.

For all statistical analyses, effects were considered significant if the *p* value was equal to or less than .05. All *p* values were adjusted (Bonferroni correction) for multiple testing.

Results

Table 2 presents the association matrix averaged across subjects for each type of sound. For all sounds, the score of success was significantly above chance ($p < .001$ for each shape, $d_{circle} = 13.19$; $d_{ellipse} = 2.01$; $d_{arches} = 1.91$; $d_{line} = 9.42$ for recorded sounds, $d_{circle} = 13.19$; $d_{ellipse} = 2.81$; $d_{arches} = 2.23$; $d_{line} = 4.9$ for synthesized sounds) and was higher than 80% (highest scores for the line and the circle with almost 100%). Moreover, the scores of success differed significantly from the three other association scores ($p < .001$ for all comparisons). These results revealed that the four sounds had been associated correctly to the corresponding shapes.

Results also showed that, based on the cophenetic distances, the matrices for recorded and synthesized sounds were strongly correlated, $r(4) = 0.89$; $p < .05$. Moreover, the Wilcoxon's tests showed that success rates did not differ between recorded and synthesized sounds for each shape (Circle: $z = 0$; $p = 1$; Ellipse: $z = -1.265$; $p = .21$; Arches: $z = -.632$; $p = .52$; Line: $z = -1.13$; $p = .25$). This revealed that the two types of sound provided similar association scores.

Discussion

The results of this experiment showed that the subjects were able to associate a given friction sound (selected among four) to the correct shape. The scores of success were high for all shapes. Furthermore, the scores obtained for synthesized and recorded sounds did not differ significantly, although the synthesized sounds were modulated by the velocity profiles only. In fact, the two types of sounds differed by the fact that other variables, such as the pressure, the orientation of the mine of the pen and the irregularity of the roughness of the rubbed surface, were implicitly contained in recorded sounds. However, despite of these additional features, the scores were not higher for recorded sounds. This consideration is in line with the study by Schomaker and Plamondon (1990), revealing that no general biological relation exists between these additional features and the kinematic characteristics of a drawn shape. In any case, as the subjects did not draw the shapes themselves, they could not have established any relation between potential acoustical cues linked to pen pressure or angle and the geometry to improve their discrimination of the stimuli.

The result obtained supports the assumption that the velocity profile is perceptually relevant and seems to convey pertinent

---

[2] This type of analysis is used in phylogenetics to evaluate similarities between matrices and to make a comparison between, for instance, an "empirical" classification computed from macroscopic observations and an "objective" one computed from DNA sequences (Sokal & Rohlf, 1962).

Table 2

Association Matrices for the Experiment 2

Shape	Circle	Ellipse	Arches	Line
		Recorded sounds		
Circle	**98.75*****	1.25	—	—
	1.25	1.25	—	—
Ellipse	1.25	**81.25*****	17.5	—
	1.25	6.25	6.31	—
Arches	—	17.5	**80*****	2.5
	—	6.31	6.44	1.72
Line	—	—	2.5	**97.5*****
	—	—	3.51	1.72
		Synthesized sounds		
Circle	**98.75*****	1.25	—	—
	1.25	1.25	—	—
Ellipse	—	**87.5*****	12.5	—
	—	4.97	4.97	—
Arches	—	11.25	**82.50*****	6.25
	—	4.62	5.76	3.08
Line	1.25	—	5	**93.75*****
	1.25	—	2.92	3.08

Note. Mean scores and SE in percentage for each shape with recorded (top) and synthesized sounds (bottom). Null values are noted by "—." Significance of the comparison to chance test: $^*\, p < .05.$ $^{**}\, p < .01.$ $^{***}\, p < .001.$

information for shape identification. In particular, distinct events such as the silences (corresponding to the zero-crossings in the velocity profile) might be one of the causes of the high recognition scores. From a cognitive point of view, we assume that the association between sounds and visual shapes is enabled based on an internalized model of the gesture evoked by the visual depiction and the perception of sound variations according to the velocity profile of the gesture.

As a matter of fact, in this experiment, the high scores of success were obtained on a set of shapes that were quite distinct as regards the presence (or absence) of cusps and consequently, as regards the underlying movement involved in the drawing of them. In the following experiment, our aim was to examine whether the sound/shape association could also be successfully achieved with shapes that are similar to each other, geometrically, and with acoustic cues based mainly on continuous variations in sound timbre, hence we excluded acoustical cues like silences. More similar shapes should imply more similar velocity profiles that may, in turn, produce friction sounds that might be more difficult to differentiate.

Experiment 3

Method

The procedure and data analysis were the same as in Experiment 2.

Subjects. Eighteen participants took part in the experiment, 8 women and 10 men. Their average age was 31.56 years ($SD = 13.73$). None of the subjects were familiar with the topic of the study prior to the test. Seven of these subjects participated in Experiment 1, and 17 subjects in Experiment 2.

Stimuli

Static shapes. As for Experiment 2, preliminary informal tests were carried out to choose the shape corpus and the associated

sounds based on geometrical and perceptual criteria. In particular, we considered a set of shapes without cusps on the supposition that the corresponding friction sounds would be less distinguishable from a perceptual point of view. Therefore, we kept the circle and the ellipse shapes from the corpus of Experiment 2 and replaced the shapes presenting cusps, that is, arches and lines by loops and lemniscates that do not contain cusps. The four selected shapes are presented in Figure 3.

Recorded sounds. The stimuli corresponding to the circles and ellipses were conserved from Experiment 2. For the loops and lemniscates, the recording sessions took place in the same conditions and with the same writer as in Experiment 2. Sequences corresponding to four periods of sound recordings were selected on the basis of geometrical and temporal characteristics of the writer's performances, described in Table 3. Figure 4 presents the velocity profiles of the four shapes. It is noticeable that these profiles are more similar to each other than those in Experiment 2, which should imply more perceptual confusion between the associated sounds. Only the circle seems to have a velocity profile easily distinguishable from the others.

Synthesized sounds. Friction sounds were synthesized using the same friction model as in Experiments 1 and 2, and by using the velocity profiles collected from the writer during the recording sessions.

Results

The results, presented in Table 4, revealed that, for both recorded and synthesized sounds, all shapes were associated to the correct sound with scores higher than statistical chance; loops with recorded sounds, however, proved an exception to this with a success rate of 29.17% ($p < .001$ for each shape excepted the recorded loops, $d_{circle} = 6.12$; $d_{ellipse} = 0.53$; $d_{lemniscate} = 1.26$; $d_{loops} = 0.13$ for recorded sounds, $d_{circle} > 100$; $d_{ellipse} = 1.30$;
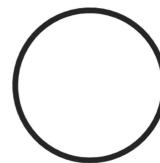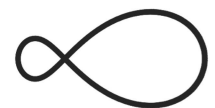
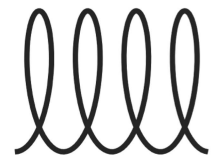*Figure 3.* The four shapes used in Experiment 3. For loops, four periods are presented.

Table 3
Geometrical and Temporal Characteristics of the Performances Chosen for the Stimuli of Experiment 3

Shape	Length (cm)	Duration (s)
Circle	62.5	5.2
Ellipse	89.32	5.8
Lemniscate	145.46	5.6
Loops	92.1	5.4

Note. The lemniscate appears to be longer than the three other shapes, but the duration of the associated sound is almost the same as the three other ones because the mean velocity is higher for this shape. The recorded and synthesized stimuli of Experiment 2 are available in supplementary material online.

$d_{lemniscate} = 2.23$; $d_{loops} = 0.7$ for synthesized sounds). The circle and the lemniscate presented the highest scores of success and no confusion with other shapes was observed.

In contrast, results revealed some confusion between ellipses and loops; their scores of association were above statistical chance and did not differ significantly from each other in both directions of association: (a) loop sound with ellipse shape, and (b) ellipse sound with loop shape. Confusion was observed for both recorded (51.39% and 45.83%) and synthesized (45.83% and 43.06%) sounds. These association scores did not differ significantly from the score of success for the ellipses and loops for both recorded sounds, $t(17) = 0.73$; $p = .47$ and $t(17) = 1.89$; $p = .22$, and synthesized sounds, $t(17) = 0.77$; $p = .66$ and $t(17) = 0$; $p = 1$.

Results also revealed that the association matrices for recorded and synthesized sounds were significantly correlated, $r(4) = .94$,

Figure 4. One period of the velocity profiles of the four shapes used in Experiment 3. The velocity profiles are low pass filtered at 10 Hz and normalized in amplitude.

Table 4
Association Matrices of the Experiment 3

Shapes	Circle	Ellipse	Lemniscate	Loops
		Recorded sounds		
Circle	**97.22*****	2.78	—	—
	2.78	2.78	—	—
Ellipse	—	**41.67*****	6.94	**51.39*****
	—	7.29	2.72	6.22
Lemniscate	2.78	9.72	**68.06*****	19.44
	2.78	3.58	8.04	6.87
Loops	—	**45.83*****	25	29.17
	—	7.89	7.83	7.36
		Synthesized sounds		
Circle	**100*****	—	—	—
	0	—	—	—
Ellipse	—	**50*****	4.17	**45.83*****
	—	4.52	2.26	5.05
Lemniscate	—	6.94	**81.94*****	11.11
	—	2.72	6.00	7.46
Loops	—	**43.06*****	13.89	**43.06*****
	—	4.87	5.04	6.00

Note. Mean scores and SE in percentage for each shape with recorded (top) and synthesized sounds (bottom). Null values are noted by "—." Significance of the comparison to chance test: $^* p < .05$. $^{**} p < .01$. $^{***} p < .001$.

$p < .001$. Moreover, the Wilcoxon's tests showed that the scores of success did not differ between the recorded and synthesized sounds of each shape (Circle: $z = -1$; $p = 1$; Ellipse: $z = -.99$; $p = .32$; Lemniscate: $z = -1.348$; $p = .18$; Loops: $z = -1.29$; $p = .20$). This revealed, therefore, that the two types of sounds provided similar association scores.

Discussion

The results of this experiment confirmed some of those obtained in Experiment 2: the circle was still perfectly associated with the correct sound (for both recorded and synthesized sounds) even given the three other sounds that were less distinguishable from a perceptual point of view. However, the scores of success for the ellipse were lower than those obtained in Experiment 2. Furthermore, these results revealed some confusion between the ellipse and loops in the shape/sound and sound/shape associations. The lemniscate was well associated with high scores but was marginally confounded with loops and ellipse (scores of association were lower than chance threshold). These confusions can be explained by the proximity of their velocity profiles (see Figure 4). Finally, there was no difference between the scores and confusions elicited by the recorded sounds and those elicited by the synthesized ones. Although the score differences between the recorded and synthesized sounds were not significant, the synthesized sounds always yielded higher scores than the recorded ones, especially for the ellipse, the lemniscate and the loops, for which the score difference was greater than 10%.

These results reinforce the assumption, made in Experiment 2, that the velocity profile constitutes appropriate perceptual information for the association task and the processing of the underlying gesture. However, some confusion was observed in Experiment 3 between the ellipse and loops, for both recorded and synthesized sounds. This indicates that these sounds do not contain enough information on the
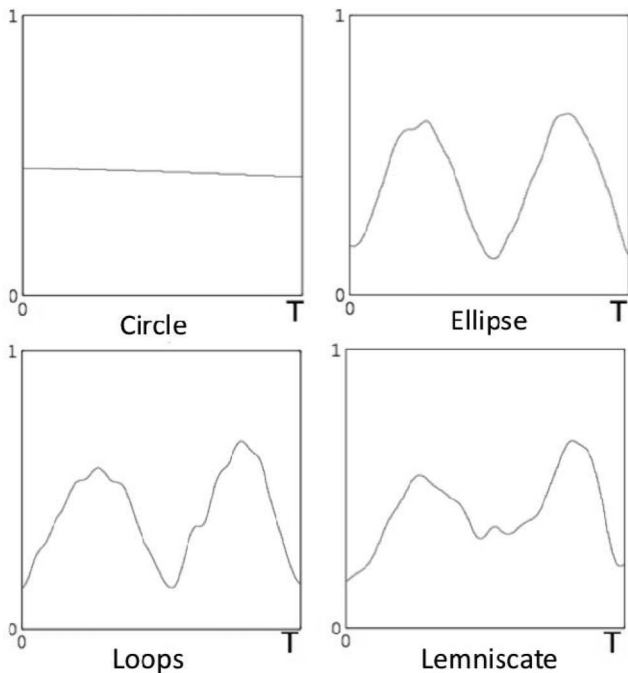
drawn shapes to distinguish those with similar geometries. By highlighting this confusion, this experiment revealed some limitations to the possibility of a direct relation between sound and shape. This result is also in accordance with the 1/3 power law: to imagine the underlying gesture from the friction sound, we extract kinematics information (velocity profile) and we associate accelerations and decelerations to the curvature characteristics of the drawn shape. Hence, if shapes have similar geometries, the velocity profiles are also similar and the associated sounds are less distinguishable from the auditory point of view.

General Discussion

In the series of experiments presented here, we investigated the human ability to perceive biological movements through sounds and furthermore, to retrieve the drawn shapes from the sounds in an association task. To our knowledge, this is the first time that this topic has been formally addressed. The Experiment 1 investigated whether the velocity profile is a relevant feature for auditory perception using a protocol close to one of Viviani and Stucchi (1992) in the visual modality. To this end, we used a real-time synthesis model to generate friction sounds in which timbre variations were modulated by the velocity profile. This enables to investigate the perceptual relevance of this parameter only. We concluded that these timbre variations should be generated by a velocity profile that obeys the 1/3 power law to evoke a natural and fluid biological movement. This study revealed that sounds can adequately inform about human movements if their dynamical acoustic characteristics are in accordance with the way the movements are performed.

The Experiments 2 and 3 further investigated whether participants were able to extract the spatial characteristics of visual shapes from the sounds and we opted for an association task. In Experiment 2, we compiled a sound corpus assumed to be perceptually easy to discriminate. In Experiment 3, the corpus comprised less discriminable stimuli. As expected, high scores of success for distinct shapes and some confusion for similar shapes were observed. Discriminating between visual shapes on the basis of their produced sounds is, therefore, possible if the acoustic characteristics of the sounds differ sufficiently. The lack of score differences between recorded and synthesized sounds confirmed that these characteristics are related mainly to the velocity profile of the underlying drawing movement, which complies with the 1/3 power law.

From an ecological perspective, the specific pattern of timbre variations induced by the velocity profile can be considered as a transformational invariant enabling the evocation of the underlying drawing movement. Such timbre variations afford the action of drawing fluidly and naturally, and this result is in line with many studies dealing with auditory perception of an acoustical event (Gaver, 1993b; Li, Logan and Pastore, 1991; Repp, 1987; Warren & Verbrugge, 1984). To complement the ecological framework we will also discuss the results according to the ideomotor framework (Hommel et al., 2001; Prinz, 1997). In a recent study by Young et al. (2013), subjects were asked to reproduce walking patterns from walking sounds in real time with different stride lengths, and to discriminate these stimuli in perceptual tasks. They observed that the characteristics of the reproduced walking patterns were similar to the target ones. Moreover, the subjects were able to perform the tasks accurately even when only the kinematic information was present in the synthesized target walking sounds. This suggests

that the auditory perception of an action seems to activate the same motor schemes as when we act, and finally that, listening to an action enables one to imagine it, and even, to reproduce it in real time. The authors proposed that common cognitive representations are involved both in the perception and in the planning of an action, as suggested in recent studies which argued in favor of a reunification of direct and indirect perception (Cisek & Kalaska, 2010; Norman, 2002). In the case of their experiments, the authors assumed that the tasks were well accomplished thanks to an audio-motor unified percept. We propose the same conceptual processing of sensory information in the perception of human drawing through sounds. In particular, inferring drawn shapes from sounds requires an internalized association between these two percepts. This relation is not obvious because graphical shapes are not commonly associated to sounds. However, based on results from Experiment 1 and from studies found in the literature (cf. Introduction), we suggest that both a visual shape and a sound can evoke a given human gesture and that the relevant information is contained in the velocity profile as conveyed by the movement of a visual dot or by timbre variations of the sound. We propose that the evoked human gesture serves as a medium to accomplish the task of associating a shape and a sound thanks to an amodal representation of biological gesture, as proposed in Viviani, Baud-Bovy, and Redolfi (1997). Indeed, we were unable to conclude whether subjects deduced the shape from the sound or the sound from the shape. The sensorimotor representation of biological movement constitutes the central point of such a process. This can be reinforced by the results from Viviani and Stucchi (1989), which showed that the link between the dynamical representation of the shape (a moving spot) and the drawn (static) shape is mediated in a similar way by a sensorimotor representation of the underlying movement.

Hence, we speculate that sensory information (visual, auditory, and sensorimotor) can be integrated to provide a unified percept of the "drawing event." This conception also suggests that common motor rules constrain the perception of both auditory and visual biological movements, thus confirming previous results on the existence of motor-perceptual relations. As suggested by Young et al. (2013), these results are also in line with motor theories of speech perception, especially with the version assuming a role of motor knowledge in perceptual processes (Viviani & Stucchi, 1992 for a review; Jeannerod, 1995; Zatorre, Chen, & Penhune, 2007; Bangert et al., 2006 in the context of music perception and production). More generally, such sensory integration may rely on a multimodal representation, such as the model proposed by Griffiths and Warren (2004). Indeed, substantial literature pertaining to such cognitive models does not exist at this time.

Conclusion and Perspectives

This study demonstrated that the friction sounds produced when someone is drawing are sufficiently informative to evoke the underlying gesture, and to a certain extent the drawn shape. We focused on the kinematic information (i.e., the velocity profile) using a synthesis process and showed that it is possible to calibrate the timbre of a sound such that the evoked motion corresponds to a biological movement with a velocity profile that matches the 1/3 power law. At last, two experiments pointed out that shapes can be retrieved from friction sounds, and that this discrimination was enabled when only the kinematic information was present in the sound. The main results of

this study are twofold: first, from an ecological perspective, the velocity profile can be considered as a transformational invariant as it affords the recognition of human movements from the generated friction sounds. Second, a relation is enabled between drawn shapes and produced sounds, which strongly suggests that common rules constrain perception and action of biological motions.

Several perspectives can be highlighted at this point. First, it would be of interest to investigate more globally and accurately, the relations between sound, visual shape, and evoked movement. In particular, the auditory neural correlates of the sensorimotor representation of biological movements have already been investigated and it has been shown that those related to the 1/3 power law of motion are much stronger than other types of motion (Dayan et al., 2007). As revealed by the experiments reported here, synthesis is an efficient way to investigate the brain correlates of the auditory modality because any velocity profile (respecting or not the 1/3 power law) can be easily generated by this procedure. Moreover, Lewis et al. (2004) showed that certain cortical areas (posterior portions of the middle temporal gyri) are involved in both visual biological motion and sound perception and might participate in the audio-visual integration process.

Second, in Experiments 2 and 3, participants associated a given sound to a given shape among a limited number of visually displayed shapes. To access the representation of the drawn visual shape directly from the sound, it would be interesting to conduct an identification task in which participants are asked to draw the shape evoked by a sound without any visual reference and to compare both the drawn shape and the kinematics of the drawing movement with the real ones. A parameterized graphical user interface based on an interactive synthesis tool could be used for this purpose (cf. Merer et al., 2013). However, the confusions observed between similar shapes in Experiment 3 support that retrieving the correct drawn shape without any visual model may prove somewhat difficult. Another interesting perspective would be to carry out the same experiments, but with the recordings of each subject, to investigate whether idiosyncratic knowledge of pressure and pen angle behaviors could have improved shape recognition for those shapes that were confounded. It might, therefore, be of interest to exaggerate the values of the synthesis parameters, particularly those related to the velocity profile, to try to boost shape identification. This possibility may be useful in many applications for sound design, sonification, or even for musical purposes.

References

Aramaki, M., Besson, M., Kronland-Martinet, R., & Ystad, S. (2011). Controlling the perceived material in an impact sound synthesizer. *Audio, Speech, and Language Processing, 19,* 301–314. doi:10.1109/TASL.2010.2047755

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., . . . Altenmüller, E. (2006). Shared network for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage, 30,* 917–926. doi:10.1016/j.neuroimage.2005.10.044

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science, 9,* 211–214. doi:10.1111/1467-9280.00040

Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., . . . Dauman, R. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *Journal of the Association for Research in Otolaryngology, 11,* 625–640. doi:10.1007/s10162-010-0222-7

Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS One, 5,* e12240. doi:10.1371/journal.pone.0012240

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33,* 269–298. doi:10.1146/annurev.neuro.051508.135409

Dayan, E., Casile, A., Levit-Binnun, N., Giese, M. A., Hendler, T., & Flash, T. (2007). Neural representations of kinematics laws of motion: Evidence for action-perception coupling. *Proceedings of the National Academy of Sciences of the United States of American, 104,* 20582–20587. doi:10.1073/pnas.0710033104

Frémiot, M., Mandelbrojt, J., Formosa, M., Delalande, G., Pedler, E., Malbosc, P., & Gobin, P. (1996). Les Unités Sémiotiques Temporelles: éléments nouveaux d'analyse musicale. *diffusion ESKA.* MIM Laboratoire Musique et Informatique de Marseille, documents musurgia edition [The temporal semiotic unities: new element of musical analyses].

Freyd, J. J. (1983a). Representing the dynamics of a static form. *Memory & Cognition, 11,* 342–346. doi:10.3758/BF03202447

Freyd, J. J. (1983b). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33,* 575–581. doi:10.3758/BF03202940

Gaver, W. W. (1993a). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1–29. doi:10.1207/s15326969eco0501_1

Gaver, W. W. (1993b). How do we hear the world? Explanations in ecological acoustics. *Ecological Psychology, 5,* 285–313. doi:10.1207/s15326969eco0504_2

Gibson, J. J. (1966). *The senses considered as perceptual systems,* Boston, MA: Houghton Mifflin

Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America, 119,* 1171–1181. doi:10.1121/1.2149839

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal Experimental Psychology: Human perception and Performance, 36,* 462–476. doi:10.1037/a0018388

Grassi, M. (2005). Do we hear size or sound: Balls dropped on plates. *Perception & Psychophysics, 67,* 274–284. doi:10.3758/BF03206491

Grassi, M., Pastore, M., & Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? *Acta Psychologica, 143,* 96–104. doi:10.1016/j.actpsy.2013.02.005

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5,* 887–892. doi:10.1038/nrn1538

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494–500. doi:10.1016/j.tics.2004.08.007

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning, *Behavioral and Brain Sciences, 24,* 849–878. doi:10.1017/S0140525X01000103

Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal, 27,* 66–72. doi:10.1162/014892603322482538

Houben, M. M. J., Kohlrausch, A., & Hermes, D. J. (2004). Perception of the size and speed of rolling balls by sound. *Speech Communication, 43,* 331–345. doi:10.1016/j.specom.2004.03.004

Houben, M. M. J., Kohlrausch, A., & Hermes, D. J. (2005). The contribution of spectral and temporal information to the auditory perception of the size and speed of rolling balls. *Acta acustica united with Acustica, 91,* 1007–1015.

James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory? Motor brain network. *Neuropsychologia, 44,* 2937–2949. doi:10.1016/j.neuropsychologia.2006.06.026

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33,* 1419–1432. doi:10.1016/0028-3932(95)00073-C

Johnson, M. L., & Larson, S. (2003). Something in the way she moves, metaphors of musical motion, *Metaphor and Symbol, 18,* 63–84. doi:10.1207/S15327868MS1802_1

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments, 9,* 399–410. doi:10.1162/105474600566907

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297,* 846–848. doi:10.1126/science.1070311

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 279–294. doi:10.1037/0096-1523.26.1.279

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta Psychologica, 54,* 115–130. doi:10.1016/0001-6918(83)90027-6

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics, 59,* 1180–1190. doi:10.3758/BF03214206

Lemaitre, G., & Heller, L. M. (2012). Auditory perception of material is fragile, while action is strikingly robust. *Journal of the Acoustical Society of America, 131,* 1337–1348. doi:10.1121/1.3675946

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14,* 1008–1021. doi:10.1093/cercor/bhh061

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America, 90,* 3036–3049. doi:10.1121/1.401778

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461. doi:10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36. doi:10.1016/0010-0277(85)90021-6

Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage, 19,* 1492–1500. doi:10.1016/S1053-8119(03)00088-0

Longcamp, M., Boucard, C., Gilhodes, J. C., Anton, J. L., Roth, M., Nazarian, B., & Velay, J. L. (2008). Learning through hand-or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience, 20,* 802–815. doi:10.1162/jocn.2008.20504

Longcamp, M., Tanskanen, T., & Hari, R. (2006). The imprint of action: Motor cortex involvement in visual perception of handwritten letters. *Neuroimage, 33,* 681–688. doi:10.1016/j.neuroimage.2006.06.042

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). New York, NY: Oxford University Press.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America, 115,* 1306–1320. doi:10.1121/1.1645855

Merer, A., Aramaki, M., Ystad, S., & Kronland-Martinet, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transaction on Applied Perception (TAP), 10,* 1. doi:10.1145/2422105.2422106

Merer, A., Ystad, S., Kronland-Martinet, R., & Aramaki, M. (2008). Semiotics of sounds evoking motions: Categorization and acoustic features. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *CMMR 2007. Sense of sounds* (pp. 139–158). Berlin, Germany: Springer, LNCS.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences, 25,* 73–96.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9,* 129–154. doi:10.1080/713752551

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America, 81,* 1100–1109. doi:10.1121/1.394630

Richardson, M. J. E., & Flash, T. (2002). Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *The Journal of Neuroscience, 22,* 8201–8211.

Schomaker, L. R. B., & Plamondon, R. (1990). The relation between pen force and pen-point kinematics in handwriting. *Biological Cybernetics, 63,* 277–289. doi:10.1007/BF00203451

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon, 11,* 33–40. doi:10.2307/1217208

Van den Doel, K., Kry, P. G., & Pai, D. K. (2001). FoleyAutomatic: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques ACM,* 537–544.

Viviani, P. (2002). Motor competence in the perception of dynamic events: A tutorial. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action* (pp. 406–442). New York, NY: Oxford University Press.

Viviani, P., Baud-Bovy, G., & Redolfi, M. (1997). Perceiving and tracking kinesthetic stimuli: Further evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1232–1252. doi:10.1037/0096-1523.23.4.1232

Viviani, P., & Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: Converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 32–53. doi:10.1037/0096-1523.21.1.32

Viviani, P., & McCollum, G. (1983). The relation between linear extent and velocity in drawing movements. *Neuroscience, 10,* 211–218. doi:10.1016/0306-4522(83)90094-5

Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics, 46,* 266–274. doi:10.3758/BF03208089

Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 603–623. doi:10.1037/0096-1523.18.3.603

Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience, 7,* 431–437. doi:10.1016/0306-4522(82)90277-9

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704–712. doi:10.1037/0096-1523.10.5.704

Wildes, R. P., & Richards, W. A. (1988). Recovering material properties from sound. In W. A. Richards (Ed.), *Natural computation* (pp. 356–363). Cambridge, MA: MIT Press.

Young, W., Rodger, M., & Craig, C. M. (2013). Perceiving and reenacting spatiotemporal characteristics of walking sounds. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 464–476. doi:10.1037/a0029402

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8,* 547–558. doi:10.1038/nrn2152