
Volume 67, issue 11 November 2009
ISSN 1071-5819
ELSEVIER
International Journal of
Human-Computer Studies
Editors-in-Chief
E. Motta
S. Wiedenbeck
Special Issue
Sonic Interaction Design

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Int. J. Human-Computer Studies 67 (2009) 976–993
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron

Guillaume Lemaitre[a,*], Olivier Houix[a], Yon Visell[b,c], Karmen Franinović[b], Nicolas Misdariis[a], Patrick Susini[a]

[a]STMS-IRCAM-CNRS UMR9912, F-75004 Paris, France
[b]Zurich University of the Arts, Zurich, Switzerland
[c]McGill University, CIM and CIRMMT, Montreal, Canada

Received 24 September 2008; received in revised form 30 June 2009; accepted 21 July 2009
Available online 28 July 2009

Abstract

This paper reports on an approach to the design of continuous sonic feedback in tangible interfaces, and on quantitative evaluation methods intended to guide such design tasks. The issues it addresses may be of central relevance to areas of the emerging discipline of sonic interaction design that have begun to address the unique problems of designing sound for highly interactive contexts. Three experiments were conducted to assess two key aspects of the sound design developed for an abstract object designed for these experiments, which we refer to as the Spinotron. First, a comparison of sound source identification was made between three cases: passive listening to temporally static sounds; passive listening to dynamically evolving sounds; and listening to sounds generated through active manipulation of the artifact. The results show that control over the sound production process influences the material of the objects in interaction identified as the source of the sounds. Second, in a learning experiment, users' performance with the Spinotron device was compared between a group of participants that were provided only with passive proprioceptive information, and for another group who were also presented with synthetic sound produced by the artifact. The results indicated that the sound, when present, aided users in learning to control the device, whereas without the sound no learning was observed. Together, these results hold promise toward creating a foundation for the design of continuous sound that is intended to accompany control actions on the part of users, and toward establishing a basis for experimental/quantitative evaluation methods and gathering basic knowledge about sensory-motor activity engaged in tangible sonic interactions.

Keywords: Sonic interaction design; Auditory perception; Sound identification

0. Introduction

The use of sounds in human–computer interfaces, whether within the graphical user interface of a desktop computer, or the browser of a mobile phone, is widespread in our everyday lives. Moreover, new technologies for sensing and embedded computation, and related economies of scale, have made it possible for designers to consider sonic augmentations of a much wider array of everyday objects that incorporate electronic sensing and computational capabilities, either for aesthetic purposes (such as enhancing the roar of a car engine) or for functional reasons (such as improving the usability of a browser on a music player with a small visual display).

To date, the use of such sounds has been primarily limited to signals indicating discrete changes of state in the system involved, or discrete actions of a user; for example, the sound of crumpled paper is played when a computer file is deleted. Where continuous auditory feedback is concerned, however, far less is known to guide the designer, who at the same time faces a more complex design task, as the sound is no longer produced in a static or isolated way, but is rather coupled to human action in real time.

*Corresponding author. Fax: +33 1 44 78 15 40.
E-mail address: Guillaume.Lemaitre@ircam.fr (G. Lemaitre).

1071-5819/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2009.07.002

The most refined approaches to such design come from music, and musical instrument design has been at the forefront of interaction design with sound for many years. To produce a good tone, a violinist bows a string, and (particularly during training) adjusts his or her bowing action continuously, as required, by listening to the sound that is produced. This sonic feedback informs the player about the state of the violin, but also guides the player's control, modifying bow speed, pressure, angle, and so forth. Feedback of this type can therefore be regarded as part of a continuous and dynamical loop: a user is continuously controlling an instrument (here, a violin); this manipulation of the instrument produces sounds that vary in a coherent way with the actions of the user or the player; in turn the sounds affect how a user is performing. Transposing this complex aspect of traditional musical instruments to the design of digitally augmented devices raises a number of technical and aesthetic questions that have been explored in research communities surrounding new digital instruments for musical expression.[1]

However, the design of musical artifacts is guided by different aims than those that are relevant to product design. The effective application of auditory displays to HCI contexts that are not primarily concerned with musical performance is arguably limited by a lack of suitable design methodologies and evaluation methods. Comparatively few examples within HCI exist, and even a smaller number have contributed knowledge toward the design and evaluation of continuous auditory feedback in functional contexts.

The goal of the design and evaluation activities that are reported upon in this paper is to develop further knowledge toward a basis for the design of continuous auditory feedback in sonically augmented interfaces. Our studies are based on an interface that we refer to as the *Spinotron*. It is a tangible, one degree-of-freedom controller that is endowed with both sensing and synthesized sonic feedback, design of which was based on the metaphor of the rotation of a virtual ratcheted wheel, driven in a manner analogous to a child's toy consisting of a spinning top.

From the standpoint of perception, the level of dynamical interactivity embodied by such artifacts is very different from the situation of passive listening in which most perceptual studies are carried out. If experiments are carried out in such a setting, participants are not listening to sequences of static sounds selected by an experimenter, but are instead dynamically exploring the sounds of an interactive object. This context may be thought to be more closely allied with enactive views of perception (e.g. Bruner, 1966) than with experimental auditory psychology. A second goal of the study reported on in this article was to investigate the manner in which listeners that must manipulate a system to generate sound, rather than just passively listening to them, perceive the cause of those sounds differently. More precisely, the questions that were experimentally addressed in this study are:

- How does manipulation modulate the perception of the cause of the sounds?
- Does the sound guide users in learning how to control the Spinotron so as to drive the speed of the ratcheted wheel?

This paper is divided into four sections. In Section 1, we report on related work in sonic interaction. Section 2 details the design of the Spinotron. Sections 3 and 4 report on three perceptual studies: the experiments in Section 3 assess whether the speed of the wheel can be used to convey information to the users, and allow to select the model parameters that generate sounds that are the most coherent with the metaphor of a ratcheted wheel. The experiment reported on in Section 4 first studies the interpretation of the sounds, as the Spinotron is being manipulated, and as the users listen to the sounds passively. Second, it investigates influence of sonic feedback on how users learn to control the speed of the virtual ratcheted wheel, using two different control modes.

1. Related work

1.1. Sound in HCI: from iconic to dynamic

Human computer interaction has evolved tremendously in the past 30 years. Many methods for interaction have been developed, making use of a large variety of devices and techniques, and sound is no exception.

Historically, HCI has focused on sounds in the form of short abstract static signals, typically warning or feedback sounds. The use of these sounds is now relatively common in applications such as hospital or car equipment, or high performance aircraft (Patterson et al., 1986; Edworthy et al., 1991; Stanton and Edworthy, 1999). The use of iconic (or metaphorical) sound notifications, in which virtual objects are augmented with an acoustical behavior similar to those of related counterparts in everyday physical environments, was proposed by Gaver (1986, 1989); see, for example, the Sonic Finder, ARKola and EAR projects (Gaver, 1994). The identified sound event consequently carries an associated meaning: for example the crush of paper is associated to the removal of a file.

More recently, this proposal has been extended so as to be applied to the design of sound for interactive devices (Müller-Tomfelde and Steiner, 2001), in some cases using real time sound modeling of a virtual mechanical phenomenon causing the sounds (Müller-Tomfelde and Münche, 2001; Rocchesso et al., 2003). The application of such a causal representation, rather than an abstract one, is based on the hypothesis that sonic interactions with virtual objects should not require excessive cognitive effort on the part of users in order that they may understand and decode the information contained in the feedback they supply. For example, in Müller-Tomfelde and Münche

[1] See for instance the NIME community: http://www.nime.org.

(2001), interactions between a pen and a white board are sonified by synthesizing different qualities of the surface and of the pen (chalk on slate, boardmarker on a flipchart, etc.) and the type of gesture produced when handling the pen on the surface (pen-down, pen-up, etc.).

This approach has also been adopted in other devices that have used a causal approach to coupling sound and action as a core of the interactive experience that they afford. For example, the Shoogle (Williamson et al., 2007) is a mobile device which, when shaken, provides information about the state of its battery or about the contents of the SMS inbox. It does so via sound produced from a model of virtual balls of different sizes and materials bouncing inside the phone box in a viscous material. In such cases, continuous sonic feedback may present itself as a subtle sonic response to a naturally occurring human motion (for example, walking with a device in a pocket) and may allow users to access data without looking at a visual display. The widely distributed Nintendo Wii game controller has enabled engaging game experiences that benefit from couplings between sound and action. In the Wii Tennis game, the impact of a ball coincides with sonic and vibrational feedback emanating from the hand-held controller, creating an embodied sensation that one is physically hitting the ball.

As further illustrated by these examples, multisensoriality can also be an important component of sonic interactions (Rocchesso et al., 2008; Rocchesso and Polotti, 2008), as can the coupling of action and perception.

1.2. Approaches to performance assessment for the design of sound coupled to action

The study of human–computer interaction entails an understanding of perceptual-motor behavior, because these processes underly any form of human interaction. However, many of the methodologies that have been developed for this study, generally in the framework of information-processing theory, consider human–computer interaction from what may be regarded as a reductionist perspective (Proctor and Vu, 2007). For examples, reaction times, movement times or other chronometrical measurements have been widely used. Classical assessments of input devices (beginning, in HCI, with the computer mouse) have been based on Fitts' (1954) law, which predicts the time to reach a target as a function of the distance to the target and its width. These studies focused on time efficiency, and have not necessarily taken into account other factors that can influence user experience (for investigations of this latter issue, see Welsh et al., 2007). However, in the design of continuous auditory feedback to users' actions, sensory-motor experience "may not be best indexed merely by chronometrical methods" (Welsh et al., 2007, p. 29) and new methodologies may be required.

The methodology developed in Rath and Rocchesso (2005), Rath (2006, 2007), and Rath and Schleicher (2008) provides an interesting example of how sonic interactions might be experimentally investigated. They describe the Ballancer, a tangible interface consisting of a wooden plank that may be tilted by its user in order to drive a virtual ball rolling along the plank. A user tilting the latter hears the rolling sound produced by the virtual ball. This rolling sound is produced by a synthesis model that allows to vary the size, mass and shape of the ball. The authors used this interface to study subjects' abilities to use this auditory feedback in a task involving guiding the ball to a target region along the length of the plank. They found that the auditory feedback, in combination with a visual display, allowed users to guide the ball to the target area more rapidly compared to a case in which they were provided with visual feedback alone. Moreover, in a comparison using the same task with the rolling ball sound and with a "synthetic" sound (i.e. one that does not mimic any physical system) that preserves the same information, subjects were found to perform better, early in training, with the realistic sound than with the "synthetic" one, whereas the latter provided better performance after training. Participants reported having preferred the realistic sounds, despite performing better with the "synthetic" feedback.

As in the case of the Ballancer, we investigate the Spinotron by means of a dynamic task in which participants are asked to control the state of a virtual ratcheted wheel using synthesized auditory feedback generated by the device.

2. Interface design: the Spinotron

Similarly to the Ballancer, the Spinotron (see Fig. 1) was conceived as an abstracted artifact capable of generating digitally synthesized sound through interaction in an intuitive way, via a metaphor based on a virtual physical mechanism (see Fig. 2). Specifically, it affords a simple one-dimensional mode of input, based on an act of manual vertical pumping of its central shaft. The artifact was designed with the aim of supporting the experiments of Sections 3 and 4.

A few criteria guided the basic design of this interface. First, it was desired that the mode of interaction with the artifact be simple and familiar, so that users would not have difficulty using it or understanding how to do so correctly. Furthermore, the mode of gestural control over the sounds should be continuous and *effective*, in the sense of supplying (virtual) energy to excite the sound, as described by Cadoz (1999), rather than merely modulating an ongoing sound process. Finally, we selected a mode of interaction that did not seem to have an especially strong association to a particular sound class, in order to avoid excessive bias on users' interpretation of the synthesized sounds during the experiments.

2.1. Mechanism

The physical interface consists of a cylindrically symmetric object, shown in Fig. 1, with a maximum height of

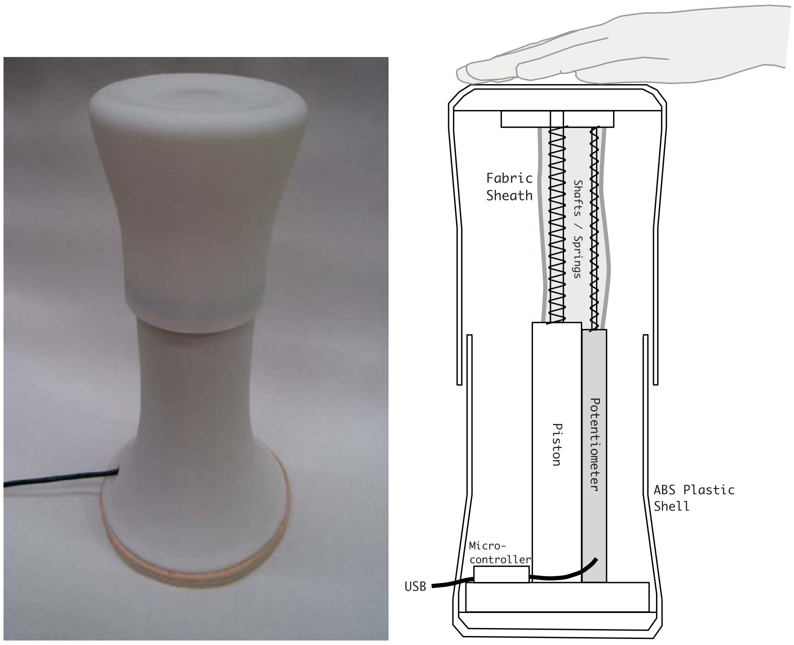

Fig. 1. Left: the Spinotron affords vertical pumping and generates digitally synthesized sounds. Right: the configuration of the Spinotron device.

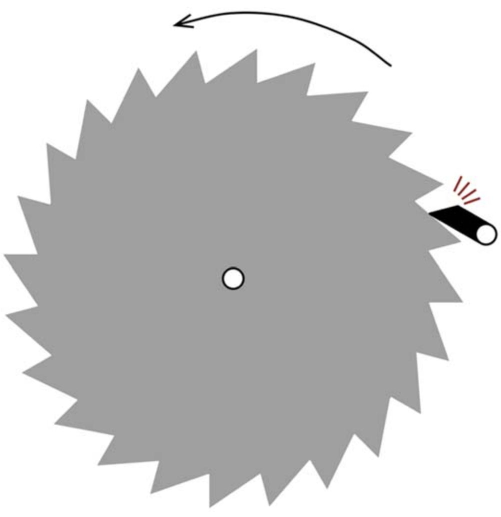

Fig. 2. A sound synthesis model was designed based on a ratcheted wheel, whose rotation is driven by the pumping motion of the Spinotron.

35 cm. Its shell was modeled in 3D software and was extruded in ABS plastic using a rapid prototyping printer (Dimension model BST 768). A linear mechanism, based on a spring-loaded lightweight industrial piston, allows the two halves of the device to telescope over the stroke of the piston (approximately 15 cm), one half into the other. The spring and piston were selected to make the device easily compressible by a wide range of potential users, without inducing excessive fatigue. A fabric sleeve in the interior prevents the mechanical components from producing significant noise. The position of the piston is sensed using a long-stroke linear potentiometer (ESI systems model LCP12S), and digitized using a small electronic board and microcontroller (Cypress systems PSOC). The digitized position is transmitted over a USB serial link to a computer running the sound synthesis and experiment software. As described below, depending on the experiment, sound was played back over loudspeakers or headphones.

2.2. Sound design

The design of the interactive sound synthesis model controlled by the Spinotron was guided by a few key objectives. For reasons described above, it was intended that the model be capable of providing continuous sonic feedback to the control actions of a user of the artifact. Moreover, because this study investigates the perception of the cause of the sound under different control conditions, the sound should be reasonably identifiable, while affording some room for interpretation. Moreover, this study is concerned with everyday sounds, analogous to auditory icons in a human–computer interface. Finally, as noted above, a simple control relation between the gestures performed with the interface and the sound produced was desired, since users were to be asked to perform a control task using the auditory feedback they generate.

Among the simplest everyday sounds are discrete impact events, which can be thought of as analogous to a short musical note. While a single impact lacks the temporal extent necessary for continuous feedback, a sequence of impacts can provide approximately continuous information in time—for example, through the instantaneous rate of impacts in the sequence, or their velocity of impact. More complex sound synthesis models were also considered (including a model of a ball moving in a rotating bowl) but were discarded during pre-testing because they proved to be too difficult for users to understand how to control or too ambiguous (or both).

The family of sounds ultimately selected was, as a result, based on a simple ratcheted wheel mechanism, capable of producing a regular sequence of impact events, similar to the freewheel of a bicycle (Fig. 2). Pumping activity supplied by its user causes the ratcheted wheel to rotate, in the manner of a child's spinning top. The rate of rotation (and, therefore, impact frequency) increases with the energy of pumping, in the way described below.

2.3. Sound synthesis

The sound synthesis model consists of two components: one is responsible for synthesizing each click of the ratchet, while the other determines the sequence of impacts and their amplitudes.

Impacts between the two elements of the ratchet (called the *wheel* and *pawl*) are synthesized by a physically motivated signal processing model of an inertial object (or "hammer") impacting a resonant object. The implementation was provided by the University of Verona SDT software library (Monache et al., 2008). The nonlinear impact force $f(t)$ is determined by a simplified phenomenological equation known as the Hunt and Crossley (1975) model:

$$f(t) = kx(t)^{\alpha} - \lambda x(t)^{\alpha}\dot{x}(t). \tag{1}$$

Here, $x(t)$ is the compression displacement and $\dot{x}(t)$ is the compression velocity. The impact force has parameters governing stiffness k, dissipation λ, and contact shape α. This force is coupled to a resonant object, modeled as a bank of 10 modal oscillators with impulse response $y(t) = \sum_i a_i e^{-b_i t}\sin(2\pi f_i t)$, determined by a set of amplitudes a_i, decay rates b_i, and resonant frequencies f_i. An impact event is synthesized by initializing Eq. (1) with the desired velocity v_I of impact and subsequently integrating the composite system in time. See Rocchesso et al. (2003) for a more detailed discussion. The synthesis model used in the experiments described below possess 10 resonant modes. Overall scale factors A, B, and F, for amplitude, decay, and frequency, are also introduced, allowing to multiply a nominal set of modal parameters.

The rate of impact events in time is determined by the rotation of a virtual ratcheted wheel, spinning at angular velocity ω, and characterized by a certain tooth height h, and number of teeth n_t per revolution. The impact rate $r(t) = \omega(t)n_t$. The velocity v_I of an impact depends on the ratchet's stiffness k_r, and on a modulation factor depending on the phase $\theta(t)$ of the rotation at the time of impact,

$$v_I = \sqrt{k_r}[1 + m\sin(\theta(t))]. \tag{2}$$

The modulation is a heuristic factor, controlled by a depth parameter m, intended to distinguish the ratcheted wheel sound from what otherwise seemed too similar to a linear ratchet mechanism. It can be thought of as arising from non-uniformity of the wheel's teeth.

2.4. Continuous control model

A user of the Spinotron is able to control the sound synthesis model by driving the virtual ratcheted wheel into rotation. The control relation between the wheel's angular velocity ω and the Spinotron's pumping displacement z is given by

$$\dot{\omega}(t) = \frac{\dot{z}(t)}{J}\,\Theta(\dot{z}(t)) - \gamma\omega(t). \tag{3}$$

Here, $\dot{\omega}(t)$ is the angular acceleration of the wheel, J is the rotational inertia of the wheel, γ is its rotational damping, and $\dot{z}(t)$ is the velocity of the pumping displacement. $\Theta(\cdot)$ is the Heaviside step function, equal to 1 if its argument is positive, 0 otherwise. The sign of z is defined so that the wheel's rotation accelerates in proportion to the velocity of pumping, but only when the interface is being compressed. The equation is implemented in discrete time using the Euler method.

2.5. Quantized control mode

The experiments described in Section 4 study sound source identification and task performance in a setting in which users are controlling the sound synthesis, nominally using the control model described above, which we call the *continuous control mode*. In addition, we investigated the same issues using a variation of that model with reduced fidelity. This *quantized control mode* was implemented by discretizing the impact rate obtained from the continuous control mode. In it, the relation between the angular velocity $\omega(t)$ of the wheel and the quantized rate $r_q(t)$ at which ratchet impacts are produced is made discontinuous as a function of the angular frequency:

$$r_q(\omega(t)) = k\omega_0\, n_t \quad \text{if } k\omega_0 < \omega(t) < (k+1)\omega_0. \tag{4}$$

Here ω_0 is a free parameter describing the frequency quantum, n_t is the number of ratchet teeth, and k is any integer such that $k \geq 0$.

Summarizing, in the quantized control mode, the system dynamics governing the angular frequency are identical to those in the continuous mode, but the resulting impact rate is quantized, and therefore only discretely coupled to the continuous rotation of the wheel. Thus, unlike the continuous mode case, the auditory feedback in the quantized mode possesses a piecewise-constant impact event rate.

The quantized model is therefore far less sensitive to the user's gesture. In the quantized mode, the user cannot have a fine control of the speed of the ratchet: he or she can keep the ratchet at predefined constant speeds.

3. Perception of the sounds and selection of sound parameters

The metaphor of the ratcheted wheel is communicated to users pumping the Spinotron through the ratchet sound. The latter consists of a series of impacts, the rhythm of which is driven by the speed of the wheel: the faster the wheel turns, the greater the density of impacts. Furthermore, the parameters of the model used to synthesize the sound of the ratchet may be selected to convey different impressions of materials of the wheel and the pawl. The experiments of this section study whether critical features of this metaphor can be understood by listeners sufficiently well for these sounds to be used in the experiments that follow. First, we investigated experimentally whether users can perceive and estimate the speed of the wheel by listening to the sounds produced, independently of the perceived material of the impacts (Experiment 1). However, not all materials can be expected to provide sounds that are consistent with a ratcheted wheel. It is very unlikely, for example, that the sounds perceived as made by hollow wooden objects will be associated with a wheel. The second part (Experiment 2) of this section provides the results of an experiment aiming at selecting sounds coherent with a wheel, among different parameter settings of the model.

3.1. *Perception of the speed of the ratchet (Experiment 1)*

The speed of the virtual ratcheted wheel is communicated to users by three redundant acoustic cues: the instantaneous rate of impacts, the wheel position-dependent modulation, and the loudness of the sound (when the speed of the wheel increases, the rate of impacts becomes denser, thus increasing the short-term loudness). The speed of the wheel will be used in Experiment 3 to assess the influence of the auditory feedback on how well a user can control the Spinotron. In order to verify that this information can be reliably used, the ability of a listener to estimate the speed of the wheel by listening to the sound was assessed, as we now describe.

3.1.1. *Method*

Participants: Nineteen participants (12 women and 7 men) volunteered as listeners and were paid for their participation. They were aged from 19 to 42 years old (median: 23.5 years old). All reported normal hearing.

Stimuli: Three parameter settings of the model of the ratcheted wheel were used (“B”, “G”, “F”), chosen to give the impression of three different kinds of wood (see Table 3). For each parameter setting, 13 sounds were created, corresponding to 13 different speeds of the ratchet (2.8, 5.6, 8.5, 11.3, 14.1, 16.9, 19.8, 22.6, 25.4, 28.2, 31.1, 33.9, 36.7 RPM). These different speeds of the ratchet correspond to sounds with a density of impacts varying from 1 to 13 impacts/s (there are 21 teeth on the wheel). The sounds were all between 3 and 4 s long. Their maximum levels varied from 47 and 65 dB(A). They had 16-bit resolution, with a sampling rate of 44.1 kHz.

Apparatus: The stimuli were amplified diotically over a pair of Sennheiser HD250 linear II headphones. Participants were seated in a double-walled IAC sound-isolation booth. The experiment was run using the PsiExp v3.4 experimentation environment including stimulus control, data recording, and graphical user interface (Smith, 1995). The sounds were played with Cycling’74’s Max/MSP version 4.6.

Procedure: The experiment had two main steps. In the first step (Step 1: free description of the cause), the participants were provided, for each parameter setting, with an interface allowing them to listen to the 13 sounds. They could listen to the sounds as many times as they wished. For each of the three parameter settings, they had to write down what they thought to be the physical cause common to all the 13 sounds.

In the second part (Step 2: estimation of the speed), the participants were told that the sounds they had heard had been produced by ratchets turning at different speeds. Then, in three sessions corresponding to the three parameter settings (“B”, “F”, “G”), they had to estimate the speed with a slider on a scale from 0 to 1. Each of the 13 sounds in each of the three parameter settings was presented twice (test/retest). At the beginning of each session, the participants were allowed to listen to all the sounds of the parameter setting. The sounds were randomly ordered within each session, and the order of the sessions was randomly assigned to each participant.

3.1.2. *Analysis*

Step 1: *Free description of the cause*. Because of their free form, the descriptions cannot be submitted to a simple quantitative analysis. They were primarily used to prepare the questions in Experiments 2 and 3. However, several trends in the answers have to be highlighted:

- A few descriptions were coherent with the metaphor of a ratcheted wheel (e.g. “the different rhythms make me think of a wheel turning at different speeds”).
- Many descriptions described wood or plastic.
- Many descriptions mentioned the resonance of some hollow objects (e.g. “resonating wood”, “a saucepan”, “a wooden bowl”, “a skin tied to a cylinder”).
- Some descriptions were of one object bouncing on another (“a ball is bouncing on a wooden object”).

Step 2: *Estimation of the speed*. The difference between the evaluated speed in the tests and the retests was, on average, 0.1 on a scale of 1.0, which is a fair consistency. Therefore the test and retest scores are averaged.

Fig. 3 represents the boxplots of the estimated speed as a function of the actual speed (in RPM), for the three parameter settings. The estimations range from 0 to 1 (i.e. full scale). Overall, the estimated speed increases when the speed parameter increases. The variation is larger at the

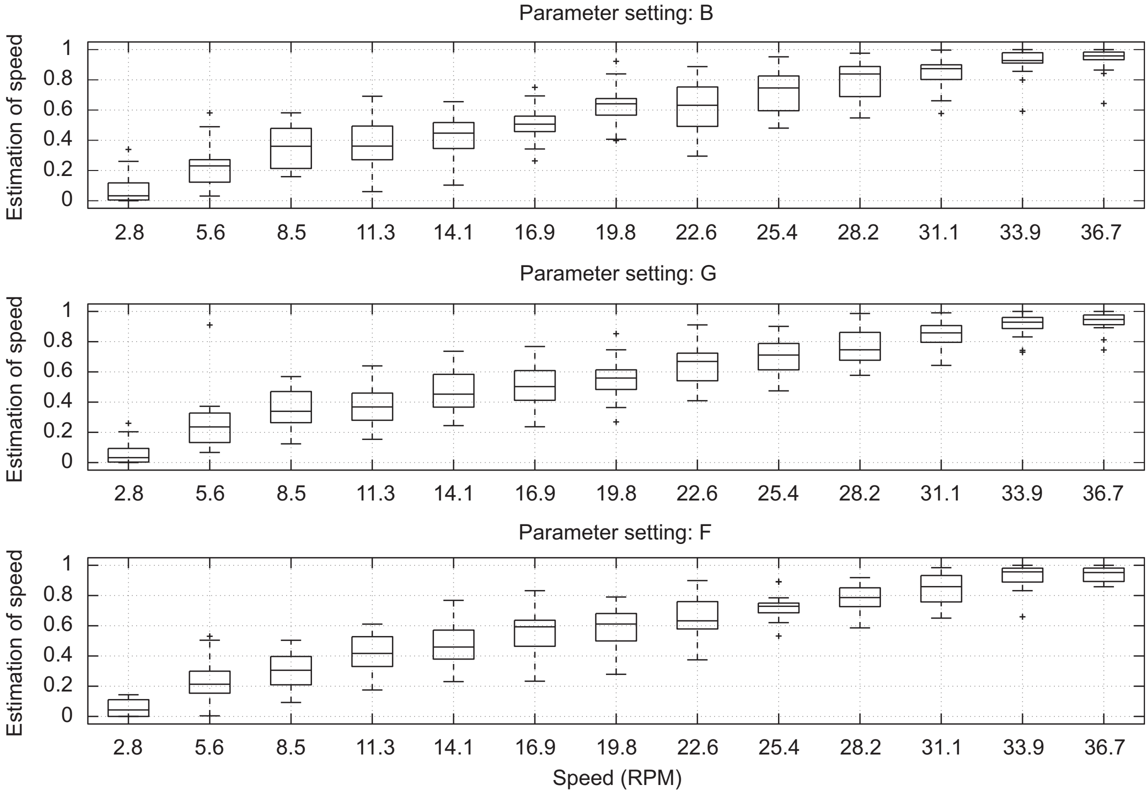

Fig. 3. In Experiment 1, the participants had to estimate the speed of a ratcheted wheel, by listening to the sounds. Boxplots of the estimation of the speed of the ratcheted wheel for the three parameter settings of the synthesis model are shown.

Table 1
Experiment 1: coefficients of the linear regression analysis between estimated speed (averaged over the participants) and actual speed (estimated speed $= b_0 + b_1 \times$ actual speed).

Parameter setting	b_0				b_1			
	Est.	SD	df	p	Est.	SD	df	p
B	0.089	0.0228	11	0.00**	0.025	0.0010	11	0.00**
F	0.093	0.0225	11	0.00**	0.024	0.0010	11	0.00**
G	0.089	0.0240	11	0.00**	0.025	0.0011	11	0.00**

The linear regression is fitted with the least square method. Est.: estimation of the coefficient, SD: standard deviation, df: degree of freedom, p: probability of the null hypothesis ($b = 0$); $^{**}p < 0.01$.

center of scale than at the extremes of the scale, because of a floor effect for lower values and a ceiling effect for higher values. The variation is homogeneous across the three parameter settings, and rather small, indicating again a good agreement between the participants.

Fig. 3 also shows that the relationship between the estimated speed, averaged between the participants, and the actual speed is linear for the three parameter settings (for the three parameter settings, $r(13) > 0.99$). In order to compare the estimated speed for the three parameter settings, the results are submitted to linear regression analysis (least square fitting method). The results of the analysis are summarized in Table 1. All the estimated parameters of the three regressions are significantly different from zero, and very close. To further test this latter assumption, the coefficients of the regressions for each pair of parameter settings are submitted to a Student t test. The results of the test indicate that all the coefficients of the three regressions are not significantly different (df $= 22$; b_0: $t(B \text{ vs. } F) = 0.298$, $p > 0.05$, $t(B \text{ vs. } G) = 0.149$, $p > 0.05$, $t(F \text{ vs. } G) = 0.444$, $p > 0.05$; b_1: $t(B \text{ vs. } F) = 0.0392$, $p > 0.05$, $t(B \text{ vs. } G) = 0.0054$, $p > 0.05$, $t(F \text{ vs. } G) = 0.441$, $p > 0.05$).

3.1.3. Discussion

The results of this experiment show that the listeners were able to readily estimate the speed of the ratchet, and indicate the range over which this estimation is linear, with respect to the speed parameter. Furthermore, the results show that the estimation is not influenced by the parameter setting used. Together, these results indicate that it is possible to use the ratchet model in the Spinotron if one wants to use the perceived speed of the ratchet as information conveyed to the user: he or she would perceive variations in this parameter almost perfectly linear.

However, the data collected in the informal interviews suggest that the material perceived by the listeners (driven by the parameter settings) has an influence on their comprehension of the metaphor of a ratcheted wheel set into rotation by the user pumping the Spinotron. Indeed, many descriptions mentioned hollow wooden objects hitten, or bouncing objects, which is not really coherent with the picture of a tiny mechanism within the Spinotron. It was therefore important to choose appropriate parameter settings.

3.2. Perception of the cause of the sounds: selection of model parameters (Experiment 2)

The latter discussion has emphasized the importance of synthesizing sounds that provide the users with sounds coherent with the intended metaphor. To reach this goal, three new parameter settings were introduced, as described in the following paragraph. They are designed so as to convey the impression of a tiny metallic mechanism (thought to be more coherent with the metaphor of the ratcheted wheel). To select among these parameter settings, an experiment testing the perception of the material conveyed by these parameter settings is also reported.

The experiment reported here uses two sorts of patterns of sounds: those made of series of impacts with a constant speed (corresponding to a constant speed of the wheel), and sounds made of impacts with an evolving speed (corresponding to accelerations and decelerations of the wheel). Data gathered here for two groups of participants (listening either to constant speed sounds or to evolving speed sounds) will be compared in Section 4 to a group of participants listening to the sounds while manipulating the device. The difference between the two groups will therefore not be analyzed here, but are reported in Section 4.

3.2.1. Method

Participants: Thirty-six participants (20 women and 16 men) volunteered as listeners and were paid for their participation. They were aged from 22 to 51 years old (median: 29 years old). All reported normal hearing. None of them had participated to Experiment 1.

Stimuli: Five parameter settings of the ratchet patch were used (“B”, “G”, “1”, “2”, “3”). Parameter settings “B” and “G” were used in the previous experiment (“F” was omitted for it sounds quite similar to “B” and “G”). Settings “1”, “2” and “3” were specifically created to give the impression of a tiny metallic mechanism. These sounds were obtained by increasing the modal frequencies (F), and increasing (parameter setting 1) or decreasing (settings 2 and 3) the decay factor (B) of the model. The parameters of the impact (surface α, mass and force stiffness k) were also changed (see Table 3). There were two groups of sounds. For each parameter setting there were four sounds corresponding to a constant speed of the ratchet (15.6, 31.1, 37.4, 49.2 RPM), and four sounds corresponding to four different patterns of acceleration and deceleration. The patterns are the following:

- Pattern 1: acceleration ($11.7 \rightarrow 42.6\,\text{RPM}/1.85\,\text{s}$)—deceleration ($42.6 \rightarrow 14.6\,\text{RPM}/1.85\,\text{s}$).
- Pattern 2: acceleration ($38.4 \rightarrow 48.3\,\text{RPM}/1.24\,\text{s}$)—deceleration ($48.3 \rightarrow 5.1\,\text{RPM}/2.65\,\text{s}$).
- Pattern 3: deceleration ($38.4 \rightarrow 3.7\,\text{RPM}/3.5\,\text{s}$).
- Pattern 4: acceleration ($9.0 \rightarrow 51.9\,\text{RPM}/3.4\,\text{s}$).

The sounds were all between 3 and 4 s long. Their maximum levels varied from 50 and 66 dB(A) in accordance with their speeds. They had 16-bit resolution, with a sampling rate of 44.1 kHz.

Apparatus: The apparatus is the same as in Experiment 1.

Procedure: The participants were split into two groups: one group (18 participants) listened only to the sounds corresponding to a ratchet turning at a constant speed. The other group (18 participants) listened only to sounds from a ratchet with a dynamically evolving speed.

The experiment had four steps. In the first step (Step 1: free description of the cause), the participants were provided, for each parameter setting, with an interface allowing them to listen to the four sounds. They could listen to the sounds as many times as they wished. For each of the five parameter settings, they had to write down what they thought to be the physical cause common to all the four sounds. Then (in Step 2: free description of actions and objects), they were provided with the same interface for each set of four sounds, but this time they had to freely describe what they thought to be the actions and the objects causing the sounds in each parameter setting. In the next stage (Step 3: choice of actions and materials), they had to choose among different actions (vibrating, bouncing, banging together, hitting, falling, going clickety-clack, turning, shaking, rolling) and materials (metal, glass, wood, plastic). These categories were created from the results of the free verbalization in the previous experiment.[2] In particular, the expression “going clickety-clack” (in French: “cliqueter”) was thought to describe precisely the sound of a small metallic ratchet. They were allowed to choose several materials or actions among the proposed categories. This was motivated by preliminary experiments in which participants wished to be able to select at least two materials, because they had heard two different objects (made out of two materials) interacting. In the last stage (Step 4: forced-choice verbal portrait selection), the participants were provided, for each of the five parameter settings, with a series of eight written verbal “portraits”, and had to choose among these verbal portraits the one they thought best corresponds with the sounds for the given parameter setting. These portraits were:

- A saucepan is being hit with a spoon (in French: “On frappe avec une cuillère sur une casserole”).
- A ball is bouncing (“Une bille rebondit”).
- Water is dripping onto a vessel (“Des gouttes d'eau tombe dans un récipient”).
- A percussion is being struck by sticks (“On frappe avec des baguettes sur une percussion”).

[2] The literature on everyday sound perception shows that listeners largely describe the objects and the actions causing the sounds, when they have to freely describe a sound. See Houix et al. (2007) for an overview.

- A ratchet is going clickety-clack ("Une roue dentée cliquette").
- Finger tapping ("On tapote des doigts").
- A casino roulette is turning ("Une roulette de casino tourne").
- A gear is turning ("Un engrenage tourne").

These verbal portraits were created by the authors on the basis of the free verbalization in Experiment 1. The participants could choose only one portrait.

In each step the order of the parameter settings was randomized, and the order of the display of the sounds within each parameter setting was also randomized.

Analysis: Fig. 4 represents the bar plots of the materials selected by the two groups of participants. On average, each participant selected 1.3 materials (the different cases of single vs. combination of selected materials are not analyzed here, see next section for more details). Post-experimental interviews revealed that they selected two different materials to indicate the interaction of two objects made out of two different materials (e.g. metal on wood). The results of the two groups are not very different, indicating that the temporal patterns of the sounds had only a little influence on the perception of the material of these sounds (see Experiment 3 in next section for a systematic investigation of this question). Among the three new parameter settings, number 1 is perceived as the sound of something made out of metal or glass, 2 is mainly perceived as the sound of an object made out of metal or plastic, and 3 as being made out of wood. The answers are more widely distributed over the categories for parameter settings B and G.

Fig. 5 represents the bar plots of the actions. On average, the participants selected for each parameter setting 2.3 categories of actions for the constant speed sounds, and 1.7 actions for the evolving speed sounds. The distributions of the two groups show only little difference (the difference between the distributions of the two groups is analyzed in Experiment 3). For parameter settings B and G, the most cited actions are "bouncing", "hitting" and "turning". For parameter settings 1, 2 and 3, the most cited actions are "going clickety-clack", "hitting" and "turning" (even though the answers are more widely distributed over the categories for parameter setting 1).

Fig. 6 represents the bar plots of the verbal portraits. Overall, the differences between the two groups are small (see next section for a systematic analysis). For parameter settings B and G, the answers are more widely distributed over the available categories. For parameter setting 1, "a percussion is being struck" is by far the most cited portrait. For parameter setting 2, the most cited portrait is "a ratchet is going clickety-clack". For parameter setting 3, the most cited portraits are "a casino roulette is turning" and "a ratchet is going clickety-clack".

3.2.2. Discussion

There are few differences between the two groups (one listened to sounds with a constant speed, the other group listened to sounds with an evolving speed). An interesting

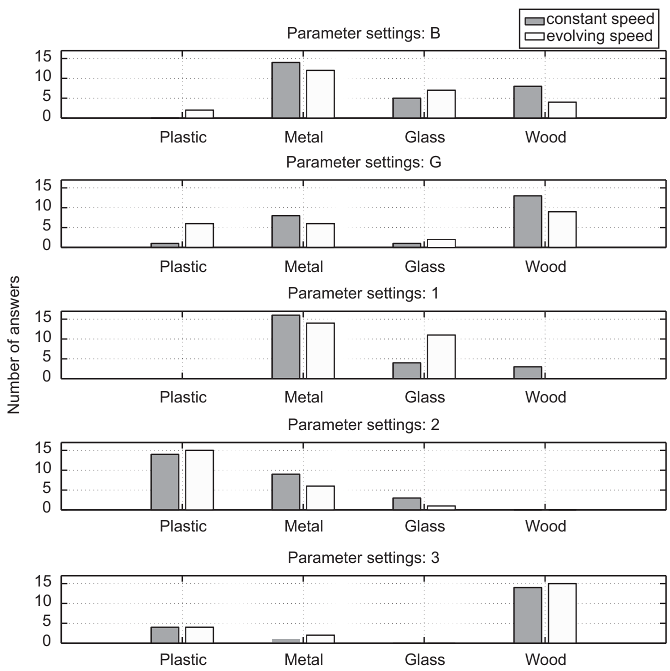

Fig. 4. Bar plots of the material selected by the participants, for five parameter settings of the synthesis model (Exp. 2). The colors of the bars correspond to two groups of participants: the first group listened only to sounds corresponding to a constant speed of the ratchet whereas the second group listened to sounds corresponding to an evolving speed.

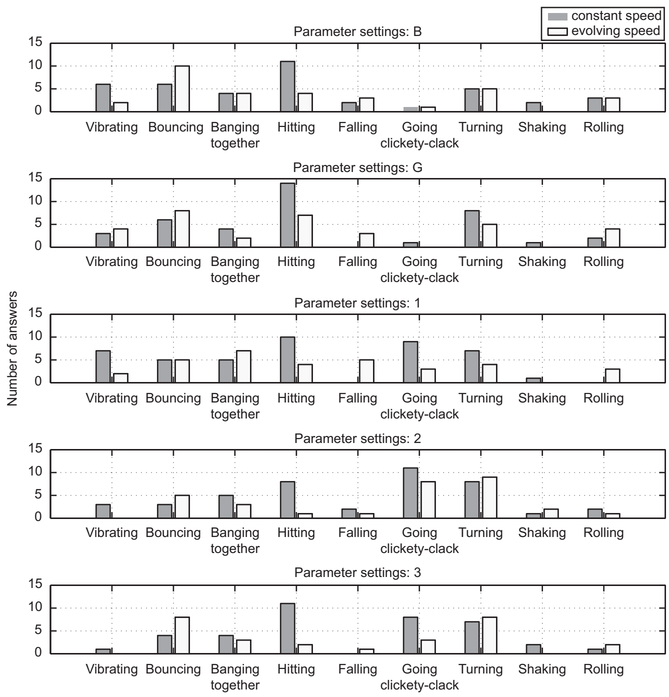

Fig. 5. Bar plots of the actions selected by the participants, for five parameter settings of the synthesis model (Exp. 2). The colors of the bars correspond to two groups of participants. The first group listened only to sounds corresponding to a constant speed of the ratchet whereas the second group listened to sounds corresponding to an evolving speed.

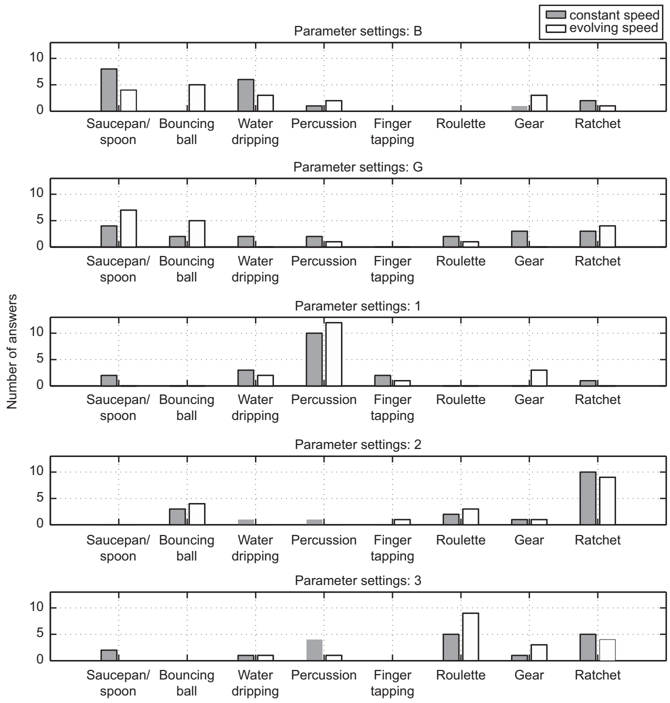

Fig. 6. Bar plots of the portraits (abbreviated here) selected by the participants, for five parameter settings of the synthesis model (Exp. 2). The colors of the bars correspond to two groups of participants: the first group listened only to sounds corresponding to a constant speed of the ratchet whereas the second group listened to sounds corresponding to an evolving speed.

distinction, however, is that participants who listened to constant speed sounds tended to describe them as sequences of impacts (''hitting''), while participants listening to the temporally varying speeds were more likely to describe the pattern of impacts: for example, they selected more often ''bouncing''. This question will be investigated in more details in the next section.

These results allow to select the parameter setting (parameter setting 2) used in the next experiment. These settings introduce a lighter mass of the impacting object (hammer mass), and produce sounds with higher modal frequencies and lower decay factors (see Table 3). Listeners described the action causing these sounds predominantly as ''going clickety-clack'', which is probably the best description of a ratchet sound, and they also selected the verbal portrait corresponding to the ratchet.

4. The manipulation of the Spinotron (Experiment 3)

With the parameter setting selected in the previous section, it is now possible to study how users learn how to manipulate the Spinotron. The goal of the experiment reported on in this section is twofold. First, it aims at comparing a group of participants interacting with the Spinotron without auditory feedback, and a group manipulating the Spinotron with feedback from the ratchet sounds selected in the previous experiment. In the latter group, the participants receive information about the interaction through force feedback (via passive mechanical feedback of the device), proprioception (the sense of their own body movement), visual and acoustic channels. In the former group, the participants do not receive information via the acoustic channel. The Spinotron was designed to provide continuous auditory feedback to accompany users' interaction, in such a way that sounds guide control of the speed of the virtual ratchet. It is therefore expected that users will perform worse when they are not provided with the sounds. Both control models described in Section 2 were tested: the continuous control mode, corresponding to the driven, rotating ratchet-wheel, and the quantized mode, in which the ratchet impact rate is discretized.

Second this experiment aims at studying what users report when required to describe the cause of the ratchet sounds while they themselves are allowed to manipulate the sound model, as compared to the case of passive listening. The assumption is that manipulating the Spinotron enables the listeners to understand the dynamics of the virtual mechanism, and therefore provides more cues to aid in the identification of the sound as originating from a ratcheted wheel. The second goal of this experiment is therefore to test this assumption.

4.1. Task design

The task that the users were required to perform with the Spinotron is simple: that of pumping the interface so as to keep the speed of the ratchet constant. The advantage of this task is that it can be performed with or without sound. Without sound, it amounts to simply pumping the device at a constant pace.

The effect of sound on users' performance is examined for both of the control modes of the ratchet model (continuous and quantized). In the quantized case, the auditory feedback is not as informative, because it communicates only the piecewise constant speed of the ratchet, while the continuous mode provides a continuously varying indication of the same. It happens that the task difficulty for the two control modes is not identical. Driving the ratchet within the prescribed target range is easier in the quantized mode, due to quantization effects. However, this does not matter for the present study, as our aim in the performance part of the experiment is not to compare performance between these two control modes, but rather to assess users' performance with and without auditory feedback, keeping the control mode constant.

As in Rath and Rocchesso (2005), the experiment reported in this section is a learning experiment. Across trials, the participants learn how to adjust their gesture so as to keep the speed of the ratchet at a constant target speed, as defined by the number of impacts per second. The experiment is therefore made of sequences of training and test phases. The target speed is defined as the range from 8.1 to 9.4 impacts per second (corresponding to 22.75–26.25 RPM of the virtual wheel for the continuous case). This target range was chosen, and the dynamical parameters selected, in such a way that, independently of the mode of interaction, the target area can be reached by pumping the Spinotron at a constant pace of 3 pumps per second.

4.2. Description of the experiment

Participants: Thirty participants (19 women and 11 men) volunteered as listeners and were paid for their participation. They were aged from 19 to 57 years old (median: 29 years old). All reported normal hearing. None of them had participated to Experiment 1 or 2.

Apparatus: The stimuli were played at the same time through a Yamaha P2075 amplifier to a pair a Tannoy Reveal loudspeakers and a pair of Beyerdynamic DT770 headphones. The headphones were used to mask the natural sound of the Spinotron (in addition to the fabric sleeve used to decrease the noise), and the loudspeakers were used during the demonstrations phases, when the experimenter had to demonstrate the Spinotron to the participant. Participants were seated in a double-walled IAC sound-isolation booth. The participants interacted with the interface through an Elo Touchsystems Intuitive touch screen (they did not use the keyboard nor the mouse). The software environment was the same as in previous experiments.

Stimuli and dynamics of the system: The stimuli were generated in real-time (parameter setting 2, see Table 3) by the synthesis model when participants interacted with the Spinotron.

Procedure: First, the participants were divided into two groups. The first group ($N = 18$) did the experiment with the sounds turned on, whereas the other group did the experiment with the sounds turned off ($N = 12$).

The procedure had two main parts. The first part (phase 1: description) was done by all the participants in the first group, using the continuous control mode. The participants had to freely manipulate the Spinotron, and to describe the sounds (materials, actions, portraits), following the same interface and procedure as in Experiment 2. The second part was the manipulation part, and was done by 12 participants of the first groups, and all the 12 participants in the second group. In each of the two main groups (sounds on/off), there were two subgroups of six participants, corresponding to each of the two control models. Beforehand, the procedure was demonstrated by the experimenter (with different sounds, and a different target). The participants had to manipulate the Spinotron so as to maintain a target constant speed of the ratchet. This phase was made of 12 trials. Each trial was made of a training step, and test step. In each training step, a visual indicator with three colors indicated to the participants whether the speed of the ratchet was below the target speed, within the target, or above the target. During each test step, the participants did not receive any visual feedback. After each training or test step, the participants were provided by a performance measure indicating how long they had maintained the speed of the wheel within the target. The training and test steps were 6 s long, and were initiated by a countdown, to allow the participants to get their hands on the Spinotron (Fig. 7).

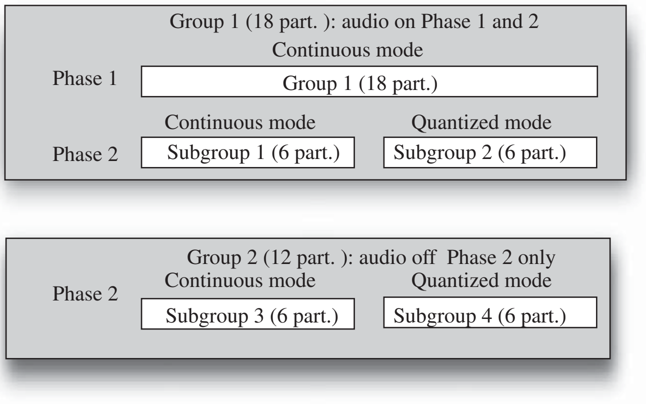

Fig. 7. Repartition of the participants in Experiment 3.

4.3. Analyses

4.3.1. Identification of the cause of the sounds

Materials: In this step of the experiment, the 18 participants could select one, or several material from four proposed categories ("plastic", "wood", "metal", "glass"). This can be analyzed as a forced-choice experiment, in which the participants are actually provided with 15 choices ("plastic", "wood", "metal", "glass", "plastic/wood", "plastic/metal", "plastic/wood/metal","plastic/wood/metal/glass", etc.).

In the group of participants listening to constant speed sounds in Experiment 2, 12 chose one single material, 4 chose a combination of two materials, and 2 chose a combination of three materials. In the evolving speed group, 14 chose a single material and 4 chose two materials. In Experiment 3, 15 chose one material and 3 chose two materials. The upper panel of Fig. 8 reports the distributions of materials selected by the participants, across all the categories of answers that have actually been made. In this figure, the bar plots from the previous experiment (Experiment 2) are also represented. Because some categories have received very few answers, some of these categories have to be collapsed to perform χ^2 analyses (to compare the groups of participants). Based on the results of Giordano and McAdams (2006), showing that plastic and wood, as well as metal and glass, are often confused, the categories plastic and wood, and metal and glass are collapsed, as well as the categories of answers corresponding to plastic and another material, resulting in the distribution reported in the lower panel of Fig. 8. A Pearson χ^2 test then reveals that the distributions of answers between the two groups of participants in Experiment 2 were not significantly different ($\chi^2(2, N = 18) = 1.02,\ p = 0.60$). However, the results of

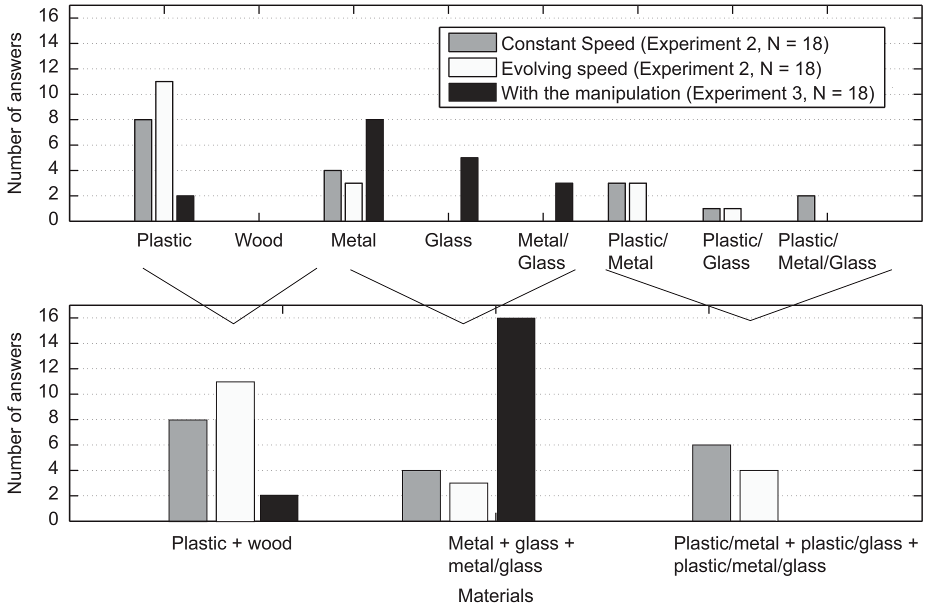

Fig. 8. Upper panel: bar plots of the material selected by the participants (Exp. 2 and 3) to describe the sounds of the ratchet (parameter setting 2). The colors of the bars correspond to three groups of participants. Lower panel: the categories that have received only few answers, and which are similar, are collapsed.

these two group are significantly different from the results of Experiment 3 (constant speed vs. manipulation: $\chi^2(2, N = 18) = 16.8$, $p<0.01$; evolving speed vs. manipulation: $\chi^2(2, N = 18) = 19.1$, $p<0.01$). Indeed, whereas the participants in Experiment 2 described the sounds as caused by different combinations of metal and glass or plastic and wood, there is a clear tendency for participants in Experiment 3 to describe the sounds mainly as metal or glass. Manipulating the device has therefore increased the impression of a metallic or glassy mechanism.

Actions: The participants in Experiment 3 could select any combination of answers among the nine proposed actions. This can be analyzed as a forced-choice experiment in which the participants are provided with 511 choices (all the combinations among the nine proposed actions). In Experiment 2 (constant speed), 14 participants among the 18 selected more than one actions. In Experiment 2 (evolving speed), nine selected more than one action. In Experiment 3, 11 participants selected more than one action. These combinations were almost all different. Because of the great number of different combinations, they had to be collapsed to be represented and analyzed. "Going clickety-clack" is set apart, because it corresponds to what is expected to be the description of the ratchet sound. "Bouncing" is also set apart because it implies a specific pattern of the impact sounds (Warren and Verbrugge, 1984). Fig. 9 reports the distributions of these categories of actions selected by the participants to describe the cause of the Spinotron. The three distributions are not significantly different (constant speed vs. evolving speed: $\chi^2(3, N = 18) = 3.36$, $p = 0.34$, constant speed vs. manipulation: $\chi^2(3, N = 18) = 6.46$, $p = 0.09$, evolving speed vs. manipulation: $\chi^2(3, N = 18) = 0.84$, $p = 0.84$).

Portraits: The participants in the three experiments could select only one portrait among the eight proposed verbal portraits. This experiment can therefore be analyzed as a forced-choice experiment with eight choices. The upper panel of Fig. 10 reports the distributions of portraits. To perform a χ^2 test, the categories with a too few number of answers are collapsed. Two main categories of portraits are build: the portraits coherent with a ratchet mechanism ("roulette", "gear", "ratchet"), and those not compatible ("saucepan", "bouncing ball", "water dripping", "percussion", "finger tapping"). The distributions of answers between the two groups of participants in Experiment 2 are identical. Then a Pearson χ^2 test reveals that the distribution in Experiment 3 is not different from the distributions in Experiments 2 (constant or evolving speed vs. manipulation: $\chi^2(1, N = 18) = 1.9$, $p = 0.17$). Together with the results of the selection of actions, this indicates that manipulating the device has not increased the perception of the metaphor of the ratchet mechanism.

4.3.2. Performances

Performance measure: The measure used to evaluate the performance of the users was exactly the measure provided to the users after each training or test step: the time during which the speed of the ratchet stayed within the target area. An example can be seen in Fig. 11, representing the speed maintained by two participant in the two modes.

Analysis of variance: Twelve participants performed the experiment with sounds (audio turned on), and 12 without sound (audio turned off). Formally, the experiment has a two between-subjects one within-subject repeated measure design, with the audio (on/off) and the control models

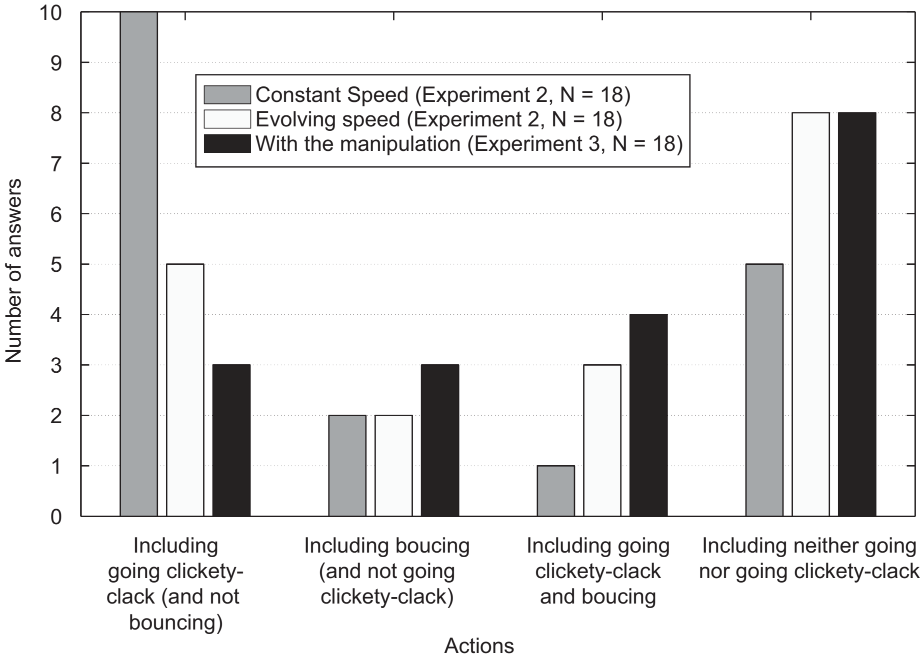

Fig. 9. Bar plots of the different categories of actions selected by the participants (Exp. 2 and 3) to describe the sounds of the ratchet (parameter setting 2). The colors of the bars correspond to three groups of participants. Each category bins together difference combinations of single actions.

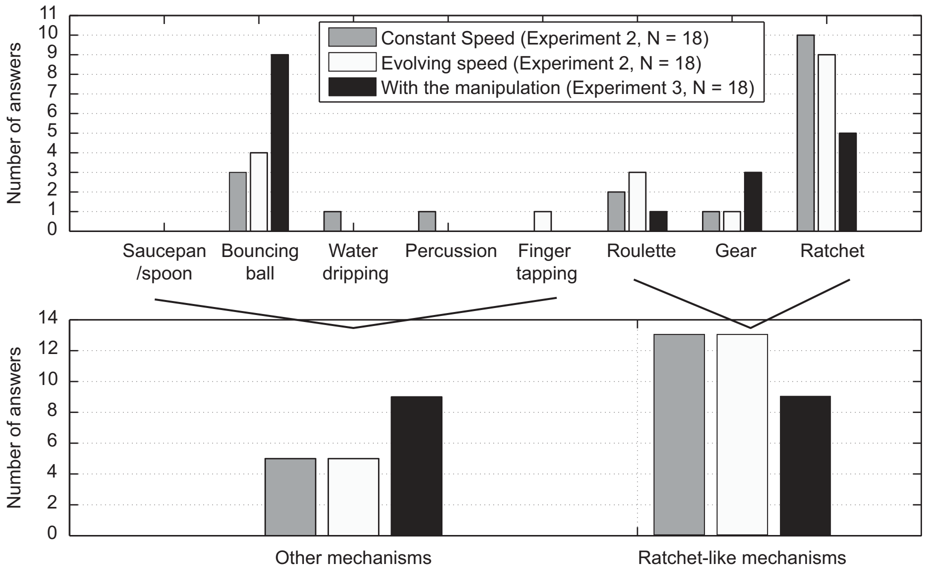

Fig. 10. Upper panel: bar plots of the portraits selected by the participants (Exp. 2 and 3) to describe the sounds of the ratchet (parameter setting 2). The colors of the bars correspond to two groups of participants. Lower panel: the categories that have received only few answers, and which are similar, are collapsed.

(continuous, quantized) as the between subjects factors, and the number of trials (1–12) as the within-subject factor. The dependant variable is the performance score.

The data are submitted to a repeated measure ANOVA. In the analysis the number of degree of freedom is corrected with the Geisser–Greenhouse correction ($\varepsilon = 0.606$). Table 2 reports the statistics of the ANOVA. With an alpha value of 0.05, the principal effects of control model and number of trials are significant. The interaction between the audio and the number of trials is also significant.

Fig. 12 represents the average values of the performances of the participants as a function of the number of trials, the control model, and the use of sounds.

The effect of the control mode is significant ($F(1,23) = 23,492,\ p<0.01$) and accounts for most of the variance in the data (partial $\eta^2 = 0.54$), indicating that the performance is better with the quantized mode than with

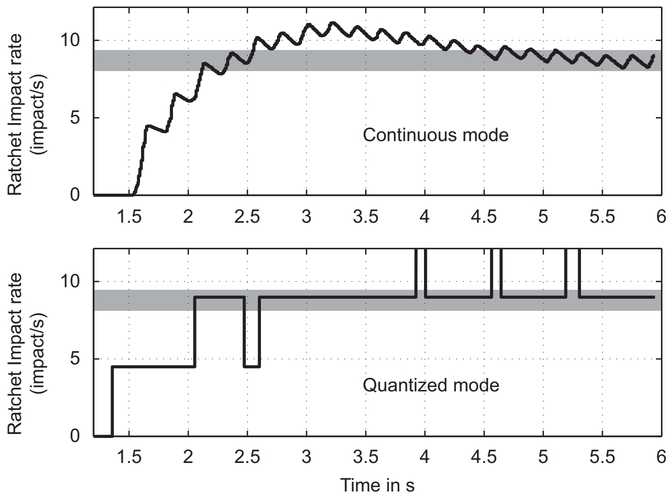

Fig. 11. Experiment 3. Example of two trials of two different participants trying to maintain the speed of the ratchet in the target area (in gray). The top panels correspond to the continuous control mode, and the bottom to the quantized control mode.

Table 2
Analysis of variance for Experiment 3.

Source	df	F	η^2	p	GG
Between subjects					
Audio (A)	1	1.966	0.090	0.176	
Control model (M)	1	23.492	0.540	0.000**	
A × M	1	0.965	0.018	0.552	
S within-group error	20	(0.190)			
Within subjects					
Trials (T)	11	2.930	0.128	0.001**	0.011*
T × A	11	2.747	0.121	0.002**	0.017*
T × M	11	0.899	0.043	0.542	0.496
T × A × M	11	0.699	0.034	0.739	0.646
T × S within-group error	220	(0.018)			

Note: Values enclosed in parentheses represent mean square errors. S, subjects. $^{**}p<0.01$. $^{*}p<0.05$. GG, probability after Geisser–Greenhouse correction of the degree of freedom.

continuous mode. As noted in the task design description given in Section 4.1, this may be explained by the lower task difficulty in the quantization control mode condition. In any case, the purpose of this study was not to compare performance between the control modes.

The effect of the number of trials is significant ($F(11,253)=2.930$, $p<0.05$), indicating that performance overall increases with the number of trials: the participants learn how to use the Spinotron across the trials. The effect of audio is not significant ($F(1,23)=1.966$, $p=0.176$), indicating that, overall, the performances are not significantly better when users listen to the sounds of the Spinotron. But, the interaction between audio and the number of trials is significant ($F(11,253)=2.747$, $p<0.05$), and accounts for almost as much variance in the data as the effect of trials ($\eta^2=0.128$ and 0.121). This indicates that the increase of performance along the trials depends on whether the users could listen to the sounds or not. This can be visualized in Fig. 12: while the increase of performance is barely noticeable on the left panel of the figure (without sounds), it appears clearly on the right panel of the figure. It can be concluded that hearing the sounds has allowed the users to learn how to use the Spinotron (performance improved across trials), whereas manipulating the Spinotron without auditory feedback has not allowed learning (performance did not improve across trials).

The right panel of Fig. 12, and the absence of significant interaction between mode and number of trials indicate that the speed of learning across the trials is the same for the two control models.

It is also worth noticing that the data presented here show large inter-individual differences, as demonstrated by the large confidence interval in Fig. 12. As an example, Fig. 13 represents the individual performances of two participants belonging to the group of participants using the continuous mode with sounds.

4.4. Discussion

The first part of the experiment studied how the participants described the cause of the sounds, when they had to freely manipulate the Spinotron. Particularly, the results were compared to the results of Experiment 2, in which the participants could hear the sounds only passively (the sounds were not caused by their manipulation of the device). The comparisons show that manipulating the device has changed the material reported by the participants: whereas in Experiment 2, the participants mainly described the sounds as caused by different combinations of objects made out of plastic, wood, metal or glass they report in Experiment 3 objects made out of metal or glass. Manipulating the device has, however, not changed the perception of the action causing the sound, nor the verbal portraits selected by the participants to describe the cause of the sounds. Contrary to what was assumed in the introduction of this section, manipulating the device has not enabled the listeners to better understand the dynamics of the of the virtual mechanism. The only effect which is significantly affected by the manipulation of the Spinotron is that the perception of metal or glass is clearly improved: the assumption by the user of a tiny metallic mechanism is probably the more plausible. One hypothesis might be that the solely passive nature of the haptic feedback (missing any variable mechanical resistance due to the virtual spinning wheel) has impeded the acceptance of the metaphor.

The second part of Experiment 3 consisted of a performance task. Participants had to pump the Spinotron so as to keep the speed of the ratchet constant. Participants alternated learning phases, in which they were provided with a visual feedback, and test phases in which there was no visual feedback. There were 12 trials. Half of the participants received no auditory feedback. Therefore, they

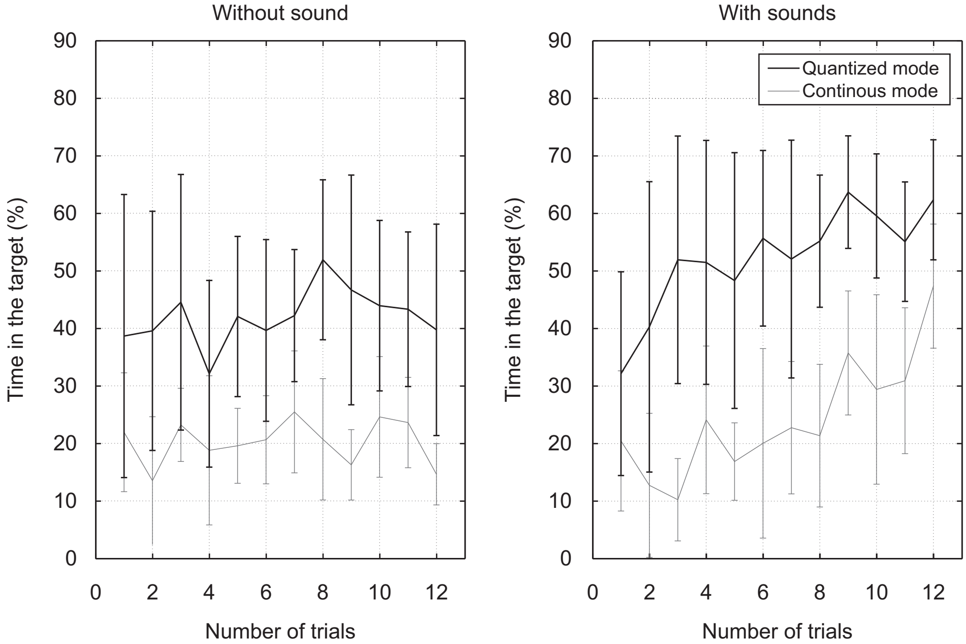

Fig. 12. Experiment 3. Average values of the performances of the participants as a function of the number of trials, the control models, and the use of sounds. The panel on the left corresponds to the group of participants manipulating without the sounds, and the panel on the right to a group of participants manipulating the Spinotron and listening to the sounds. The vertical lines represent the 95% confidence interval.

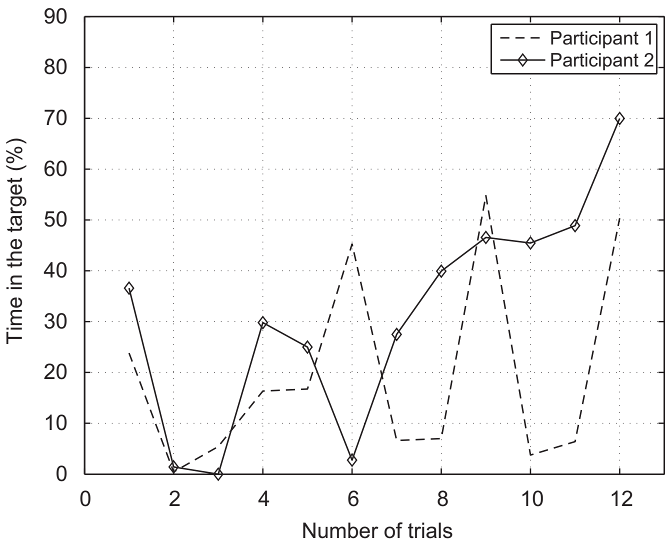

Fig. 13. Experiment 3. Performance of two participants using the continuous mode and hearing the sounds of the Spinotron.

could only perform the task by focusing on their gesture. The other group of participants could hear the sounds of the ratchet model. Comparing the performance of the two groups shows that listening to the sound of the ratchet led to an improvement of performance along trials, indicating that the sounds guided users in learning how to adjust their control gestures. This was not the case when the users did not receive any auditory feedback: without sound, they could not improve their performance across trials, which indicates that they have not succeeded in learning how to control the Spinotron more finely.

Participants were split into two groups and provided with one of two different control models. In the continuous mode, the auditory feedback varied continuously with the control, while in the quantized case the auditory feedback was piecewise constant in nature. Although the task difficulty was not the same between the two modes (as reflected in the performance data), the qualitative conclusions were identical for both. Furthermore, the improvement of performance across trials did not exhibit any difference between the two groups, indicating that participants did not learn the continuous mode faster than the quantized. However, since the task difficulty was not identical for the two control modes, this comparison is not entirely revealing. Moreover, while the results obtained revealed no significant qualitative differences between the two control modes that were presented to participants in the experiment, the same might not be true of other choices of control model (for example, other choices of quantization mapping). The results obtained here regarding the quantized control mapping are preliminary, but perhaps worthy of further investigation.

5. Discussion and conclusion

This paper has reported on an approach to the design of continuous sonic feedback in tangible artifacts, and on quantitative evaluation methods intended to guide such design. An interactive artifact was designed: the Spinotron. It is a physical object enabling pumping, and driving the real-time synthesis of a ratcheted wheel. A set of experiments was conducted to assess two key aspects of the sonic interactions involved: how manipulation modulates the perception of the sounds; how sound guides manipulation.

5.1. Summary of the experimental procedures

In Experiment 1 participants had to estimate the speed of the ratcheted wheel rotating. The results showed that the listeners were able to estimate the speed of the ratchet with a fair accuracy. The estimation of speed did not depend on the choice of the model parameters controlling the timbre (and therefore the perceived material) of the sounds.

In Experiment 2, the participants were required to select among different materials, interactions and verbal portraits describing the cause of the sounds. There were two groups of participants: one group listened to sounds corresponding to steady speeds of the wheel, and the other group listened to sounds corresponding to dynamically evolving speeds of the wheel. Experiment 2 allowed to select the parameter settings that best gave the impression of a ratcheted wheel.

Experiment 3 studied the manipulation of the Spinotron. The core of this experiment was a learning experiment. Across trials, participants had to learn how to manipulate the Spinotron so as to maintain a constant speed of the ratchet. One group of participant did the experiment without the sounds of the ratchet model in order to verify if the sounds really guide the user in performing the task.

5.2. Discussion

The results of Experiments 2 and 3 allow a comparison of sound source identification between three cases: passive listening of temporally static sounds; passive listening of dynamically evolving sounds; and listening to sounds generated through active manipulation of the artifact. Note that, even if the speeds of the sounds used in Experiments 2 and 3 are not exactly the same, their ranges are, however, comparable. The results showed that control over the sound production process modulated only slightly how listeners perceive the cause of the sounds. Indeed, only the perception of the material of the objects in interaction was modulated by the manipulation: in the passive listening case, the participants described the sounds as made by plastic, wood, metal of glass objects. In the case of the manipulation, the participants described the sounds mainly as caused by objects made out of metal or glass. However, manipulating the device did not change the portraits selected by the participants to describe the cause of the sounds.

Contrary to what was assumed, exploring the dynamics of the sound model, as users did in Experiment 3, did not reduce the ambiguity of virtual mechanism embedded in the Spinotron. A conclusion might be that manipulation has no influence at all on the perception of the virtual mechanism. Another possible explanation is that a mechanism transforming an up and down pumping movement into a rotation movement is difficult for participants not familiar with ratchets and gears to conceive of (although similar examples do exist in child spinning tops). Envisioning metallic balls bouncing inside the Spinotron when it is pressed may be easier to imagine.The possible influence on the verbal portraits of the fact that only passive tactile feedback was available to users (omitting any mechanical resistance due to the virtual system) might prove interesting for further investigation.

In Experiment 3, users' performance with the Spinotron device was compared between a group of participants that were provided only with passive tactile feedback and proprioceptive information, and for another group who were also presented with the synthetic sound produced by the artifact. The results of this learning experiment showed that improvement of performance was only possible if the users were provided with the sounds of the ratchet: without auditory feedback, the participants could not improve their performance across trials. This is not a trivial result. First, the task could be easily done without the sounds, solely by learning the correct control gesture. Second, many participants had spontaneously reported to have focused only on their gesture. The results also showed that, regardless of which control model was being used, performance with sounds increased across the trials. Even with the small number of trials allowed, control learning was improved by the sonic feedback. As discussed above, participants who were not provided with the sounds were not able to increase performance across trials.

5.3. Perspectives

The Spinotron, with its dynamic and continuous sonic interaction, offers an interesting experimental framework to study the perception of sounds. Particularly, the experimental results have demonstrated that what the listeners report as the material of the objects in interaction causing the sounds is slightly modulated by the manipulation of the device causing the sound. These differences, though small, do not seem to be attributable to the difference between sounds that occurred in passive listening and those heard during manipulation, because care was taken to include in the passive listening trials a range of different kinds of sounds potentially encountered in interaction: steady sounds with different speeds, and combinations of accelerations and decelerations. Similarly, the two control modes used by participants included ratchet sounds with dynamically varying and piecewise-constant impact rates. The primary difference between the two listening cases was the manipulation. What precisely is responsible for this modulation in causal identification has yet to be explored. The role of the specific shape, size, mechanism, material of the Spinotron, and what these visual and haptic qualities afford, remain to be

Table 3
Parameters used in Experiments 1–3.

Parameter settings	B	G	F	1	2	3
Hammer mass	1.22	1.22	1.24	1.08	0.07	1.23
Force stiffness (k)	163	163	146	171	199	185
Contact surface shape (α)	43	43	106	43	47	46
Dissipation coefficient (λ)	0	0	0	0	0	0
Frequency factor (F)	1.7	0.55	0.47	4.9	10.0	1.7
Decay factor (B)	10.82	6.44	13	12.68	0.26	0.66
Gain factor (A)	5	3	1	12	30	1

These parameters do not have a physical unit. See Section 2.3.

investigated, as do aspects of what the user can imagine as a mechanism. It is also very likely that auditory perception cannot be isolated from other sensory modalities in this setting.

Together, these results hold promise toward creating a framework for investigating the design and selection of families of continuous sounds that are intended to accompany control actions on the part of users. Indeed, studying and abstracting existing mechanical interactions (here, a spinning top) has led us to design an artifact that implements just such a continuous dynamical interaction. The experimental results demonstrate that the sonic interaction allowed users in the study to learn how to better control the input device relative to the task posed, while no improvement was possible without the auditory feedback, despite the fact that most users reported that they did not consciously attend to the sounds.

Acknowledgments

This work was funded by the European Project *CLOSED*: *Closing the Loop of Sound Evaluation and Design*, FP6-NEST-PATH no. 29085. The authors would like to thank Hans Hansen and Bruno Giordano for assistance with the statistics.

Appendix A. Ratchet model parameters

Parameters used in Experiments 1–3 are shown in Table 3.

References

Bruner, J., 1966. Toward a Theory of Instruction. Harvard University Press, Cambridge, MA.

Cadoz, C., 1999. Musique, geste technologie. In: Genevoix, H., Vivo, R.D. (Eds.), Cultures Musicales: Les Nouveaux Gestes de la Musique. Parentheses.

Edworthy, J., Loxley, S., Dennis, I., 1991. Improving auditory warning design: relationship between warning sound parameters and perceived urgency. Human Factors 33 (2), 205–231.

Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47 (6), 381–391.

Gaver, W.W., 1986. Auditory icons: using sound in computer inter-faces. Human Computer Interactions 2, 167–177.

Gaver, W.W., 1989. The sonic finder: an interface that use auditory icons. Human Computer Interactions 4, 67–94.

Gaver, W.W., 1994. Using and creating auditory icons. In: Kramer, G. (Ed.), Auditory Display: Sonification, Audification and Auditory Interfaces. Westview Press, Boulder, CO.

Giordano, B.L., McAdams, S., 2006. Material identification of real impact sounds: effect of size variation in steel, glass, wood and plexiglass plates. Journal of the Acoustical Society of America 119 (2), 1171–1881.

Houix, O., Lemaitre, G., Misdariis, N., Susini, P., Franinovic, K., Hug, D., Franinovic, K., Otten, J., Scott, J., Visell, Y., Devallez, D., Fontana, F., Papetti, S., Polotti, P., Rocchesso, D., 2007. Everyday sound classification: sound perception, interaction and synthesis. Part 1: state of the art. Deliverable 4.1 part 1 of the CLOSED project. Available at: ⟨http://closed.ircam.fr/deliverables.html⟩. URL: ⟨http://closed.ircam.fr/deliverables.html⟩.

Hunt, K., Crossley, F., 1975. Coefficient of restitution interpreted as damping in vibroimpact. ASME Journal of Applied Mechanics (42), 440–445.

Monache, S.D., Devallez, D., Drioli, C., Fontana, F., Papetti, S., Polotti, P., Rocchesso, D., 2008. Sound synthesis tools for sound design. Deliverable 2.2 of the CLOSED project. Available at: ⟨http://closed.ircam.fr/deliverables.html⟩. URL: ⟨http://closed.ircam.fr/deliverables.html⟩.

Müller-Tomfelde, C., Münche, T., 2001. Modeling and sonifying pen strokes on surfaces. In: Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland, December.

Müller-Tomfelde, C., Steiner, S., 2001. Audio-enhanced collaboration at an interactive electronic whiteboard. In: Proceedings of the 2001 International Conference on Auditory Display, Sydney, Australia.

Patterson, R.D., Edworthy, J., Shailer, M.J., Lower, M., Wheeler, P., 1986. Alarm sounds for medical equipment in intensive care areas and operating theatres. Institute of Sound and Vibration Research Report No. AC598, Southampton, UK.

Proctor, R.W., Vu, K.-P.L., 2007. Human information processing: an overview for human–computer interaction. In: Sears, A., Jacko, J.A. (Eds.), The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. CRC Press, Boca Raton, FL, pp. 43–62 (Chapter 2).

Rath, M., 2006. On the relevance of auditory feedback for subjective and objective quality of control in a balancing task. In: Proceedings of the 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems, Berlin, Germany, pp. 85–88.

Rath, M., 2007. Auditory velocity information in a balancing task. In: Proceedings of the International Conference on Auditory Displays ICAD'07, Montréal, Canada, pp. 372–379.

Rath, M., Rocchesso, D., 2005. Continuous sonic feedback from a rolling ball. IEEE MultiMedia 12 (2), 60–69.

Rath, M., Schleicher, D.R., 2008. On the relevance of auditory feedback for quality of control in a balancing task. Acta Acustica united with Acustica 94, 12–20.

Rocchesso, D., Bresin, R., Fernström, M., 2003. Sounding objects. IEEE Multimedia 10 (2), 42–52.

Rocchesso, D., Polotti, P., 2008. Designing continuous multisensory interaction. In: Contribution to the SID Workshop, CHI, Florence, Italy.

Rocchesso, D., Serafin, S., Behrendt, F., Bernardini, N., Bresin, R., Eckel, G., Franinovic, K., Hermann, T., Pauletto, S., Susini, P., Visell, Y., 2008. Sonic interaction design: sound, information and experience. In: CHI '08: CHI '08 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 3969–3972.

Smith, B.K., 1995. PsiExp: an environment for psychoacoustic experimentation using the IRCAM musical workstation. In: Society for Music Perception and Cognition Conference '95. University of Berkeley, CA.

Stanton, N.A., Edworthy, J. (Eds.), 1999. Human Factors in Auditory Warnings. Ashgate Publishing Ltd.

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 (5), 704–712.

Welsh, T.N., Weeks, D.J., Chua, R., Goodman, D., 2007. Perceptual-motor interaction: some implications for HCI. In: Sears, A., Jacko, J.A. (Eds.), The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, second ed. CRC Press, Boca Raton, FL, pp. 27–41 (Chapter 1).

Williamson, J., Murray-Smith, R., Hughes, S., 2007. Shoogle: excitatory multimodal interaction on mobile devices. In: CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 121–124.