

A synthesis model with intuitive control capabilities for rolling sounds

Simon Conan, Olivier Derrien, Aramaki Mitsuko, Sølvi Ystad, Richard Kronland-Martinet

▶ **To cite this version:**

Simon Conan, Olivier Derrien, Aramaki Mitsuko, Sølvi Ystad, Richard Kronland-Martinet. A synthesis model with intuitive control capabilities for rolling sounds. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2014, 22 (8), pp.1260 - 1273. 10.1109/TASLP.2014.2327297 . hal-01050466

HAL Id: hal-01050466
https://hal.science/hal-01050466

Submitted on 27 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Synthesis Model with Intuitive Control Capabilities for Rolling Sounds

Simon Conan†*, Olivier Derrien‡, *Member, IEEE*, Mitsuko Aramaki†, *Member, IEEE*, Sølvi Ystad†, Richard Kronland-Martinet†, *Senior Member, IEEE*

Abstract

This paper presents a physically inspired source-filter model for rolling sound synthesis. The model, which is suitable for real-time implementation, is based on qualitative and quantitative observations obtained from a physics-based model described in the literature. In the first part of the paper, the physics-based model is presented, followed by a perceptual experiment, whose aim is to identify the perceptually relevant information characterizing the rolling interaction. On the basis of this experiment, we hypothesize that the particular pattern of the interaction force is responsible for the perception of a rolling object. A complete analysis-synthesis scheme of this interaction force is then provided, along with a description of the calibration of the proposed source-filter sound synthesis process. Finally, a mapping strategy for intuitive control of the proposed synthesis process (i.e. size and velocity of the rolling object and roughness of the surface) is proposed and validated by a listening test.

Index Terms

Sound analysis, sound synthesis, audio virtual reality, rolling sounds, source-filter model, physics-based model.

I. Introduction

In this article we present a synthesis model for rolling sound interaction that has been developed to enable intuitive controls based on semantic labels. The control of synthesized everyday sounds is still

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with: †LMA, CNRS, UPR 7051 / Aix-Marseille Univ, Centrale Marseille, F-13402 Marseille Cedex 20, France; ‡Université de Toulon / LMA, CNRS, UPR 7051, F-13402 Marseille Cedex 20, France. *Corresponding author e-mail: conan@lma.cnrs-mrs.fr. Co-authors e-mail: {derrien, aramaki, ystad, kronland}@lma.cnrs-mrs.fr.

of great interest because of an increasing desire from sound designers to create and manipulate sounds for various applications. Among these applications are for instance video game development [1], [2], sound design [3], sonification [4] and virtual/augmented reality for motor rehabilitation [5], [6]. So far, the most commonly used tools for these purposes have been large data bases with indexed recorded sounds, but this approach has substantial limitations. Firstly, designing sounds from samples can be a very long process and the sound designer then needs to be highly skilled to combine and customize sound elements to create the desired sound effects. Moreover, the desired result is not always achieved (e.g. sound samples may not initially have the desired duration or be entirely coherent with the visual scene). Secondly, all the sound scenarios have to be designed in advance, therefore giving no possibilities for unforeseen interactions between the user and his or her virtual environment. In addition to the use of data bases, other sonification strategies exist, based on transformations made with basic sound attributes (pitch, loudness, tempo, spatial cues, etc.) [4]. However, these approaches do not allow subtle manipulations of timbre according to the evocations induced by sounds.

Sound synthesizers are an interesting solution to these limitations, since they offer the potential for infinite sound generation and transformation and are not limited by physical constraints as real recorded sounds are. In addition, sounds can be generated in real time based on user actions (which means that not all the possible scenarios have to be provided beforehand). In addition, intuitive control strategies combined with sound synthesis processes could be used for the creation and manipulation of sounds according to the designer's perception. This aspect is still challenging, in particular when generating sounds in interactive contexts, since human auditory perception and cognition must be taken into account in synthesizer design [7].

For the design of intuitive synthesis controls, we adopted a conceptual framework based on the *action-object* paradigm, whereby the perceptual features linked to object recognition on one hand, and to the action affecting the object on the other hand are modeled and controlled separately. Relevant indications for the determination of these perceptual features can be identified by considering the ecological approach proposed by Gibson with respect to visual perception [8]. This approach would suggest the existence of specific acoustic patterns in the captured signal, or the so-called *invariant structures*, enabling the identification of objects and events. The ecological optics approach has been applied to ecological acoustics by Warren and Verbrugge [9] to explain sound event recognition, and later formalized by McAdams [10]. The so-called invariants are split in two categories: *structural invariants*, which allow the human auditory system to recognize physical properties of a sounding object (for example its material, shape, etc.), and *transformational invariants*, linked to the nature of the interaction between the object and

its environment (for instance breaking, rolling, bouncing, etc.). For instance, a vibrating string produces a particular spectrum and can therefore be identified as such, whether it is plucked (e.g. guitar), bowed (e.g. violin) or struck (e.g. piano). Similarly, it is possible to recognize a bottle by the sound it produces, whether it bounces or breaks [9], or whether a cylinder that bounces or rolls is made of glass, wood or metal [11].

From a synthesis point of view, we chose a source-filter approach, which is well adapted to the proposed *action-object* paradigm, since the source part models the action properties and the filter part the object properties. The source-filter approach has been widely used in various applications such as speech synthesis [12] or musical sound synthesis [13]. Source-filter models offer the advantage of low-complexity implementation, which is essential for real-time purposes.

More closely related to our concern, models simulating rolling sounds and more generally any solid interactions (rubbing, scratching, etc.) were proposed notably by [14], [15] and by [16]–[19], following the pioneering work of [20]. Among these approaches, several signal models can be found that use a source-filter approach based on physical considerations. However, these studies do not address the design of intuitive strategies for control of the synthesis processes. The aim of our current study is to propose a physically-inspired signal model that simulates sounds evoking rolling interactions and that can be controlled by intuitive parameters (such as the speed and the size of the rolling object). To obtain such a model, it is first necessary to identify the transformational invariant responsible for the evocation of rolling interactions through listening tests. Further, a real-time compatible synthesizer that reproduces the behavior of the transformational invariant should be developed. This model should be sufficiently generic to enable continuous transitions towards other types of interactions such as rubbing or scratching. Such an approach has been successfully implemented in previous studies by the authors [21], [22]. Finally, in order to offer intuitive controls, the perceptual relevance of the various synthesis parameters with respect to the control parameters of interest should be estimated through listening tests, in order to propose an appropriate mapping for intuitive control. Even if the proposed model is physically-inspired, it is worth noticing that our aim is not to perfectly model physical reality, but instead to propose a model that effectively evokes perception of a rolling sound. The model aims to create the action contribution within the adopted *action-object* paradigm. The design of intuitive control for the object part has already been addressed in our previous studies [23], [24]. In practice, this object part is currently implemented using a resonant filter bank as proposed by Mathews and Smith [25]. The corresponding intuitive control enables simulation of the perceived material, size and shape of a given object based directly on semantic labels. Continuous sound transformations between material categories (for instance from metal to wood) can be

controlled within a two-dimensional navigation space.

The current paper is organized as follows. Section II presents a state of the art on rolling sound synthesis models, with a particular emphasis on physics-based models, signal models and analysis/synthesis schemes. Section III formally investigates the determination of perceptual invariants related to the evocation of rolling, with reference to a physics-based model used to generate a calibrated force signal. A perceptual experiment is used to highlight the fact that the interaction force between the rolling ball and the surface over which it rolls carries perceptually relevant auditory cues that evoke a rolling interaction. In section IV, a signal model for the evocation of rolling interaction is proposed, together with a complete analysis/synthesis scheme. An intuitive control strategy is presented in Section V and evaluated with a formal test in Section VI. Finally, some possibilities for control strategies involving a generalization of the proposed model towards other continuous interactions are discussed in Section VII. Again drawing on the *action-object* paradigm, we also discuss the possibility of creating sound metaphors by combining invariant sound morphologies related to a given action with invariant sound morphologies related to a given object (independently of the action). Such tools might offer potential new approaches in sound design and could, for instance, be of great interest in perceptual studies to generate calibrated stimuli [24], [26]. All the sounds discussed in this paper, additional sound examples and a video demonstration of the real-time synthesizer are available at http://www.lma.cnrs-mrs.fr/~kronland/Rolling_Sounds/.

II. Rolling Sound Synthesis: State of the Art

In this section, we review the literature on rolling sound synthesis models. In the first part, physics-based models are presented. Synthesis approaches based on signal considerations are subsequently reviewed.

A. Physics-Based Models

Despite the fact that the underlying physical phenomena behind sound production by rolling objects have not been clearly elucidated [14], several physics-based approaches can be found in the literature.

For moving sources [27], Stoelinga and Chaigne propose the adaptation of a model for simulation of a single impact on a damped plate, as described by Chaigne and Lambourg in [28], [29]. This model considers flexural vibrations of a rectangular Kirchhoff-Love plate [30], coupled with an impactor via Hertz's law of contact [31]. The problem is solved using a numerical method based on a finite-difference scheme. The plate features micro-scale irregularities that the impactor (the ball) scans along its trajectory across the plate. This is simulated as a white spatial distribution of asperities, acting as a perturbation

in the compression term of Hertz's law of contact. In this case the interaction force F_c between the ball and the plate, obeying Hertz's law of contact, becomes:

$$F_c = \begin{cases} k \left[\underbrace{R - (\eta(t) - W_p(x(t), y(t), t) - W_s(x(t), y(t)))}_{\Upsilon} \right]^{3/2} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for} \quad \Upsilon > 0, \\ 0 \qquad\qquad \text{otherwise}, \end{cases} \tag{1}$$

where $(x(t), y(t))$ are the coordinates of the ball on the plate, parameterized by time t, R the radius of the ball, η the vertical displacement of its center of gravity, W_p the vertical displacement of the plate (i.e. its flexural vibrations), W_s the surface profile (simulated by a white spatial distribution) and k the Hertzian elastic constant.

Rath and Rochesso also investigated a physics-based model to synthesize rolling sounds in [15]. This method differs from the one proposed by Stoelinga and Chaigne by the use of a nonlinear contact model including a dissipative component (Hunt and Crossley's model [32]). This model was first used for synthesis purposes by Avanzini and Rochesso to generate impact sounds [33]. It should be noted that the Hunt and Crossley model assumes a non-infinitesimal contact surface and the use of such a nonlinear model for rolling interaction could be criticized from a physical point of view. Indeed, since the interactions between a ball and sharp micro asperities are modeled by micro-contacts, it might be reasonable to assume simple point-like contact, and then a linear Hooke's law to model the contact force. Nevertheless, the perceptual relevance of this model has been confirmed through several listening tests [15], [40]. Another difference with Stoelinga and Chaigne's model is that the resonant object is modeled as a lumped system, with a set of N parallel second-order linear oscillators (mass-spring-damper systems), each representing an eigenmode of the resonant object (i.e. its natural frequency and damping); where Stoelinga and Chaigne directly discretized the physical equation of the vibrating plate. The model is described as follows:

$$\begin{cases} x = x_e - \sum_{i=1}^{N} x_r^{(i)} \\ \ddot{x}_r^{(i)} + g_r^{(i)} \dot{x}_r^{(i)} + \left[\omega_r^{(i)}\right]^2 x_r^{(i)} = \frac{1}{m_r^{(i)}} f(x, \dot{x}), i \in [\![1, N]\!] \\ \ddot{x}_e = g - \frac{1}{m_e} f(x, \dot{x}) \end{cases} \tag{2}$$

with

$$f(x, \dot{x}) = \begin{cases} kx^{\alpha} + \lambda x^{\alpha} \dot{x} = kx^{\alpha} (1 + \mu \dot{x}) & , x > 0 \\ 0 & , x \leq 0 \end{cases} \tag{3}$$

The parameters labeled with an e stand for the exciter (e.g. a hammer), whereas parameters labeled with an r stand for the resonant object (e.g. a plate impacted by the hammer). x_e and m_e are respectively the

vertical displacement and mass of the exciter. $\omega_r^{(i)}$ and $g_r^{(i)}$ are respectively the natural frequency and damping of the i^{th} resonant object's mode, and $m_r^{(i)}$ is the "mass" of the i^{th} mode[1], which controls inertial properties of the oscillator. $x_r^{(i)}$ is the vertical displacement of the i^{th} oscillator. g is the gravitational constant (9.81 m/s^2). The impact interaction force is represented by $f(x, \dot{x})$. The contact stiffness k is defined as:

$$k = \frac{4}{3}\sqrt{R}\left(\frac{1-\nu_e^2}{E_e} + \frac{1-\nu_r^2}{E_r}\right)^{-1} \tag{4}$$

where E and ν are respectively the Young's modulus and the Poisson ratio and R is the ball radius. λ is the damping constant (we define $\mu{=}\lambda/k$), and the nonlinear exponent α takes into account the local geometry around the contact surface (according to Hertz theory, we will assume $\alpha{=}3/2$ throughout this paper).

In order to adapt this nonlinear contact model to the rolling interaction, a surface profile is introduced. Based on physical considerations, the surface $\mathfrak{S}_r$ is assumed to be imperfect (i.e. not perfectly smooth on a micro-scale, as is the case for real surfaces). Hence, the ball tracks the surface profile and hits some of the asperities, depending on their size. The ball's velocity parallel to the surface is a forced velocity and it is assumed in this model that it is not affected by the interaction. This concept is sketched in Fig. 1. As the contact point moves along this surface, a vertical displacement (labeled x_{offset} in Fig. 1) referred to as the offset-curve is added to the distance variable x. Thus, the model actually simulates a locally bouncing ball, whose energy is cumulated over time according to the *rolling curve* x_{offset}, thus evoking a rolling movement.

B. Signal Models

Several authors have derived signal models from physically inspired considerations in order to synthesize rolling sounds. These models are mainly based on source-filter schemes, assuming that the physical exciter and the resonator are uncoupled (i.e. that the resonator vibrations do not influence the exciter's behavior). Thus, they can be modeled by an excitation signal (which represents the force applied on the resonator by the exciter) and resonant filters (which represent the natural frequencies of the vibrating object). Such methods do not perfectly imitate physical reality, but instead aim to reproduce sound effects that are perceptually relevant.

[1] In the original model for rolling sounds [15], the authors do not take into account the position-dependent interaction of the ball on the resonator. However, as proposed in [34], the effect of the position-dependent interaction (i.e. the emphasis and attenuation of certain resonant modes according to the exciter location on the resonant surface) can be taken into account within the terms $m_r^{(i)}$.

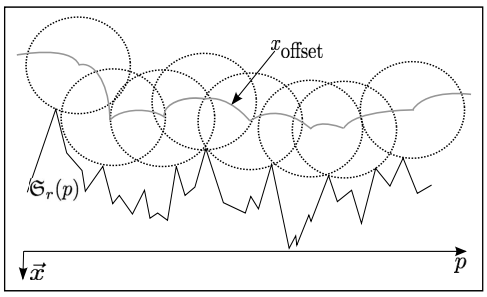


Fig. 1. Rolling curve x_{offset}, determined by the surface profile $\mathfrak{S}_r$ and the ball size. Adapted from [15].

The FoleyAutomatic synthesis scheme, proposed by Van den Doel *et al.* [16], reproduces various continuous interactions between solid objects such as rolling or scraping, using a source-filter approach. The general synthesis process consists of playing a noise at varying speeds according to the relative velocity of the interacting objects. The noise is assumed to model the surface roughness. The authors consider a fractal surface model for this purpose, which turns out to give satisfactory results in modeling real rough surfaces [35]. Fractal surfaces have a power spectrum proportional to ω^β ($\beta < 0$), where β can be interpreted as a surface roughness parameter. For rolling sound synthesis, the authors propose an additional filtering stage that enhances the spectral envelope near the object's resonance modes, in line with the observations made by Hermes [17]. Hermes hypothesized that the contact between a ball and a surface is smoother for rolling than for friction interactions. To account for this, the sharp onset of a conventional impact model that usually consists of a sum of exponentially damped sine waves is replaced by a more gradual onset modeled by a sum of impulse response-like gammatones, i.e. $s(t) = \sum_{i=1}^{N} a_i t^{\gamma-1} \exp(-t/\tau_i) \sin(2\pi f_i t)$, with $\gamma = 2$. This impact model is then convolved with a time series of pulses with a Poisson distribution. In order to enhance the auditory perception of a rolling object, the sound envelope is modulated. This can be justified by the fact that a rolling object is neither perfectly spherical nor homogeneous. This idea is also taken into account in [15].

Amplitude modulations in the synthesis of rolling sounds have proved to be perceptually relevant: in [36], artificial amplitude modulations were applied to recorded rolling sounds and perceptual experiments showed that they clearly influenced the perceived size and speed of the ball.

C. Analysis/Synthesis Schemes

Lagrange *et al.* [18] and Lee *et al.* [19] apply analysis/synthesis schemes for sounds generated by a sustained contact between rigid objects to the synthesis of rolling sounds. Both these models are source-

filter approaches. The main idea is to examine recorded sounds in order to extract the filter part parameters (characterizing the resonances of the interacting objects) and an excitation pattern used as the input into the filter part. They both assume that interaction sounds are the result of a succession of micro impacts on a resonant surface. However, the two studies use different methods to estimate the filter parameters and to perform the detection of contact events.

Lee *et al.* [19] propose first estimating the time location of contact events by high-pass filtering the signal, and then using the envelope of this filtered signal to detect the maxima of the events. The filter parameters are estimated between each pair of successive contact events. Each contact sound is then separated into subbands in which zeros are estimated using notch filtering. Thus, the notches observed in the spectrum of rolling sounds are modeled over a rigid, finite length surface, since the modes are selectively excited and damped or suppressed depending on the location of the excitation. This also has the advantage of reducing the Linear Predictive Coding (LPC) order used to estimate the poles. Synthesis can then be performed.

Lagrange *et al.* [18] propose extracting filter parameters related to the object's resonances using a high resolution method (ESPRIT [37], [38]). This is performed on short-time frames of recorded sounds. Each frame is determined such that the vibrating object remains in a free regime as far as possible. The source signal is then extracted using inverse filtering in the frequency domain. Each impact amplitude and time location is determined on the extracted source signal. The excitation is then encoded by modeling the distributions of the impact amplitudes and time intervals between impacts by exponential and gamma distributions respectively. Each impact is also modeled using a Meixner window [39].

III. Perceptually Relevant Cue for the Rolling Evocation

In this section, we focus on the perceptual features responsible for evocation of a rolling interaction. We assume that the interaction force is a transformational invariant and consequently, carries the perceptually relevant information related to the rolling interaction (independently of the resonator's properties). To test our hypothesis, we investigated synthetic sounds instead of recorded sounds, because they can be accurately controlled. The physics-based model investigated by Rath and Rochesso [15] described in Section II-A was chosen. Firstly, in our view, this model produces the most convincing sounds (some sound examples, computed as proposed in [15], [33], are available on this paper's companion webpage). Moreover, in [40], it has been shown that this model produces rolling sounds that are spontaneously recognized as such by listeners.

This model allows us to focus on the excitation, i.e. the interaction force, by reducing the contribution

of the resonant object. Hence, we consider that forces are computed on a rigid surface i.e. without any resonator contribution. The equations are discretized and solved using an explicit fourth-order Runge-Kutta method (RK4) [41], which is known to be an accurate method for solving differential equations. As shown by Papetti *et al.* [42], RK4 solutions for the physics of a bouncing ball are always close to the analytical solution. The surface is discretized with a resolution of 0.1 mm and the sampling rate is f_s=44.1 kHz. In the next part of this section, we use a perceptual experiment to verify whether, for certain combinations of parameters, the relevant information linked to the rolling interaction might be conveyed by the interaction force. This test also investigates the contribution of the modal characteristics during the rolling interaction, as explained below.

A. Subjects

Seventeen participants took part in the experiment: 12 men, 5 women, 30 years old on average (standard deviation: 8.7 years). They were all volunteer students and researchers working at the laboratory (CNRS-LMA). None of them had prior knowledge of the stimuli used in the experiments and none of them reported hearing problems.

B. Stimuli

From equations (2) and (3), it can be seen that during contact between the rolling object and the surface (x>0), the pair of coefficients $(\frac{k}{m}, \frac{\lambda}{k}) \stackrel{\text{def}}{=} (\kappa, \mu)$ parameterizes the behavior of the interaction force. We investigated whether certain pairs of coefficients produce a perception of rolling or not. The parameter space (κ, μ) was sampled regularly with all the combinations between $\kappa = [\kappa_1, \kappa_2, \kappa_3] = \left[5 \cdot 10^7, 5 \cdot 10^9, 5 \cdot 10^{11}\right]$ and $\mu = [\mu_1, \mu_2, \mu_3] = [0.1, 1, 10]$. The κ values were chosen in line with the literature: in [27], the values range between 10^4 and 10^9, while in [31] the value is in the order of 10^{12} N.m$^{-\alpha}$.kg^{-1}. For each pair, 5-second simulation was run with two fractal surfaces (β=-0.5 and β=-1). Eighteen stimuli associated with the raw forces were thus generated. In order to investigate the influence of the resonator on the perception of rolling, 18 supplementary stimuli were obtained by convolving these raw forces with an impulse response evoking a metal object (obtained with [24]), giving a total of 36 stimuli for the test (both raw forces and convolved forces). For all simulations, we assumed a constant velocity of 20 cm/s, a constant ball radius of 1 cm and a maximum asperities height of 0.1 μm. An amplitude envelope was applied to all the stimuli (to evoke a sound source approaching, then moving away from the listener).

C. Procedure

The test was performed in a quiet office on a standard laptop with Sennheiser HD-650 headphones. A graphical interface was specially designed for the experiment using Max/MSP software[2] . A subjective rating of rollingness was evaluated by experiment participants. For each stimulus, the listener used a slider to evaluate, on a scale spanning from 0 to 100, how much the sound evoked a rolling object. Verbal descriptors for rollingness along the slider were used as in [43], from Not at all like rolling (0), A little bit like rolling (25), Somewhat like rolling (50), Close to rolling (75) to Exactly like rolling (100). Before the evaluation session, 6 different sounds were presented in a random order to familiarize the listener with the sounds. During the evaluation, the 36 stimuli were presented in a random order to avoid the effect of presentation order.

D. Results

The data were analyzed using a repeated measures Analysis of Variance (ANOVA) with surface irregularity (2 levels), resonator contribution (2 levels), κ (3 levels) and μ (3 levels) as factors. Effects were considered significant if the p value was equal to or less than 0.05. A significant main effect was identified for the resonator contribution ($F(1,16)$=16, p=0.001). The mean "rollingness" was 40.9% for the raw forces and 53.7% for the convolved forces, i.e. about 13% higher when the force was convolved with an impulse response from a resonant object. The effect of the surface was not significant (p=0.13). In addition, the $\kappa\times\mu$ interaction was found to be significant ($F(4,64)$=5.08, p=0.001). The sounds generated with the parameter set (κ_2, μ_2) obtained the highest ratings and were perceived as close to rolling. This set of parameters significantly differed from all the other combinations ($p<0.01$ at least) except with (κ_2, μ_3) (p=0.07). In Fig. 2, mean "rollingness" ratings for the raw forces, averaged over the two surface conditions are presented in the plane (κ, μ).

E. Discussion

The results of this perceptual experiment revealed that sounds generated with the parameters (κ_2, μ_2) were more often perceived as having been produced by rolling objects than the other combinations (67.2%, see Fig. 2). The ratings were already high for sounds based on the raw forces only, showing that the interaction force is a relevant cue for perception of a rolling sound. Moreover, convolving the interaction force with the impulse response of an object slightly improves the perception of "rollingness"

[2]cycling74.com

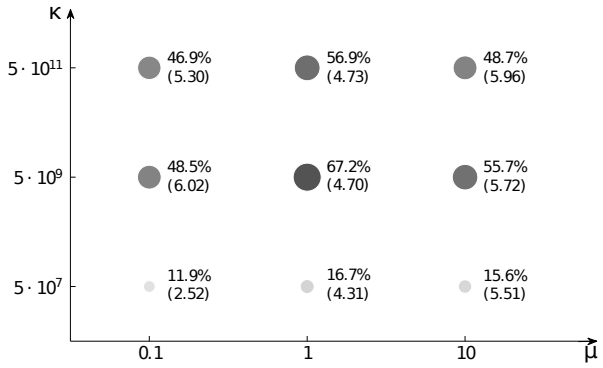


Fig. 2. Perceptual experiment results for sounds corresponding to the raw forces only (no convolution with an impulse response from a resonant object). The mean "rollingness" rating (standard error between brackets) is displayed for each set (κ, μ). The grey scale and circle size are proportional to the mean rating.

(rating of 83.5% for the (κ_2, μ_2) condition, which again has the highest mean rating in the convolved condition). One might argue that the perceptual experiment is biased due to the fact that we explicitly told the listeners they were to listen to "rolling sounds". However, after the experiment, most subjects said that some sounds evoked small balls rolling over hard surfaces (concrete, floor tiles) and others over a metallic plate, meaning that they were able to imagine the action and even a scenario from the presented sounds. Thus, we can deduce that the interaction force has a great impact on the "rollingness" perception. We will use the set (κ_2, μ_2) and some combinations in its close vicinity as reference values for the physical model synthesis presented throughout this paper.

IV. Proposed Model

The listening test described above revealed that the interaction force f itself carries relevant information for the evocation of a rolling interaction. Moreover, convolving this force with an impulse response from a vibrating object enhances this perception. From a signal point of view, this interaction force can be considered as a series of impacts characterized by particular statistics for impact amplitudes and trigger times, whereby each impact presents a specific local impact shape. Modeling the impact series involves analyzing and reproducing these statistics.

We here propose a synthesis model that includes an analysis-synthesis framework for the impact series and two additional processes: 1) the amplitude modulation effect that was shown to be a relevant cue to enhance the perception of speed (see Section II-B) and 2) the local impact interaction model that takes into account the dependency of contact duration (i.e. duration of impacts) on amplitude. In the following

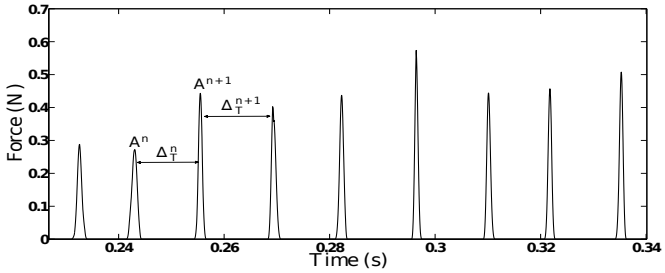


Fig. 3. Typical exemplar of the behavior of the nonlinear interaction force associated with a clear perception of rolling interaction. Close-up analysis over a short time.

sections, a signal characterization for the interaction force is described, along with the main contributions of the proposed synthesis model. Further details on the estimation of parameters are then provided.

A. Signal Characterization of the Interaction Force

We here focus on the interaction force f, with the aim of characterizing its behavior from a signal point of view. As can be seen in Fig. 3, f can be considered as a succession of micro-impacts. This leads to the following model:

$$f(t) = \sum_n A^n \, \phi^n(t - T^n) \tag{5}$$

where A^n and T^n are respectively the amplitude and time-position of each impact. They both characterize the so-called "impact sequence" and will be considered as random variables in the proposed model. ϕ^n represents the so-called "impact patterns". Because of the non-linear behavior of the force, the shape of the impact patterns changes with the impact index n. Such shapes have already been modeled in the literature and will be described in further detail in Section IV-D.

Modeling the statistical properties of the impact sequence is a more challenging problem. For this purpose, we assume that the ball rolls over an "infinite" surface with constant statistical properties. This implies that, when the ball speed is constant, the stochastic process that models the impact sequence is stationary, i.e. its statistical properties do not depend on the time origin. Thus, it is more relevant to consider the time-interval between impacts than the absolute position of each impact. For this purpose, we introduce the time-interval series defined as $\Delta_T^n{=}T^{n+1}{-}T^n$. The impact sequence is consequently characterized by the 2-dimensional time-series (A^n, Δ_T^n).

We examined the statistical behaviors of the A^n and Δ_T^n series from several simulations of the interaction force computed using the contact model (equations (2) and (3)). The physical parameters were chosen from within a range of values coherent with the perception of rolling (cf. Section III): Ball velocity:

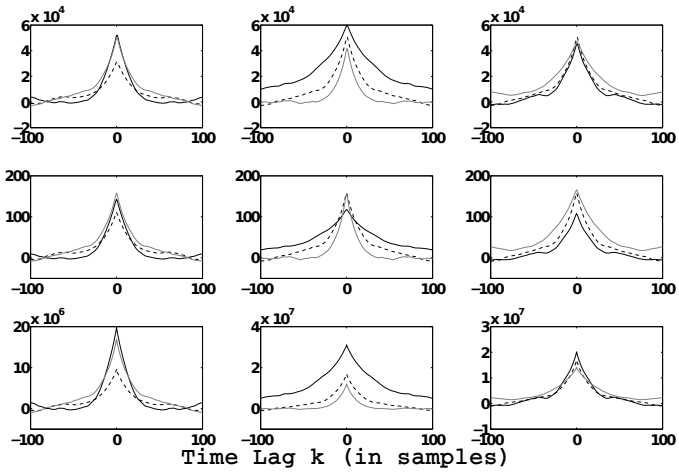


Fig. 4. Correlation functions of impact sequence for a ball rolling on a fractal surface, for different set of parameters (κ, μ, β). *Top row:* Cross-correlation between amplitude and time-interval. *Middle row:* auto-correlation of amplitude. *Bottom row:* Auto-correlation of time-interval. *Left column:* κ=10^9 (dashed black), $5{\cdot}10^9$ (grey), 10^{10} (black) N.m$^{-\alpha}$.kg^{-1}. *Middle column:* μ=0.5 (black), 1 (dashed black), 2 (grey) s.m^{-1}. *Right column:* β=0 (black), -0.5 (dashed black), -1 (grey). When not varying, the parameters are: κ=$5 \cdot 10^9$ N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1} and β=-0.5.

20 cm/s, ball radius: 1 cm, κ= $\left[10^9, 5{\cdot}10^9, 10^{10}\right]$ N.m$^{-\alpha}$.kg^{-1}, μ= $[0.5, 1, 2]$ s.m^{-1}, β= $[0, -0.5, -1]$ and maximum asperity height: 0.1 μm. We extracted the A^n and Δ_T^n series by detecting the local maxima of the computed forces. By definition, A and Δ_T have strictly positive values, and thus strictly positive mean values. To analyze these stochastic processes, we computed the centered variables of A and Δ_T by subtracting the mean value. We note ${}^cA^n$=$A^n$$-\mu_A$ and ${}^c\Delta_T^n$=$\Delta_T^n$$-\mu_{\Delta_T}$, where μ_X is the mean of the process X. We then evaluated their correlation properties. We estimated the auto-correlation functions $C_{({}^cA^n, {}^cA^n)}(k)$ and $C_{({}^c\Delta_T^n, {}^c\Delta_T^n)}(k)$, of ${}^cA^n$ and ${}^c\Delta_T^n$ respectively. These functions, plotted in Fig. 4 (middle and bottom rows), show that both variables are clearly non-white, i.e. the present X^n value depends on past X^{n-i} values. This memory process is coherent with physics, because the impact series is structured in time: the rolling ball model is derived from the bouncing ball model, which presents an organized interaction pattern. We also evaluated the cross-correlation function $C_{({}^cA^n, {}^c\Delta_T^n)}(k)$ between ${}^cA^n$ and ${}^c\Delta_T^n$. This function, plotted in Fig. 4 (top row), reveals that the two variables are highly correlated. This observation seems also coherent with physics: impacts of low amplitudes (or high amplitudes, respectively) are generally followed by close bounces (or distant bounces, respectively) characterized by low (or high, respectively) Δ_T values. This particular structure of the correlation functions is likely to characterize the rolling interaction, and its reproduction is crucial for perceptually convincing synthesis.

B. Analysis-Synthesis Scheme for the Impact Sequence

The block-diagram of the analysis-synthesis scheme for the impact sequence is shown in Fig. 5. It should be noted that the synthesis scheme shown at the bottom of Fig. 5 covers the whole synthesis process, i.e., once the amplitude and time interval series are synthesized, the amplitude is modulated to enhance the perception of velocity and then a local impact model is applied. The two latter steps are described in Sections IV-C and IV-D. We only describe the main points of the analysis-synthesis scheme here. Details of the estimation of parameters are given in Section IV-E.

1) Analysis Scheme: The amplitude A^n and time interval Δ_T^n series are first extracted from the simulation with the physics-based model [15] and zero-centered by removing their mean values. Then, the analysis consists of removing the auto-correlation of each centered variable with whitening filters $H_A^{-1}(z)$ and $H_{\Delta_T}^{-1}(z)$. The whitened versions of the time series are named ${}^c\tilde{A}^n$ and ${}^c\tilde{\Delta}_T^n$ respectively. The cumulative distribution functions of ${}^c\tilde{A}^n$ and ${}^c\tilde{\Delta}_T^n$ ($F_A(.)$ and $F_{\Delta_T}(.)$ respectively) are estimated in order to transform these time series so that the outputs W_A^n and $W_{\Delta_T}^n$ both follow a uniform law.

Finally, the cross-correlation $C_{(W_A^n, W_{\Delta_T}^n)}$ between the residual signal W_A^n and $W_{\Delta_T}^n$ is estimated and plotted in Fig. 6. Given the long-term auto-correlation of ${}^cA^n$ and of ${}^c\Delta_T^n$ (see Fig. 4), we assume that $C_{(W_A^n, W_{\Delta_T}^n)}$ is nearly a Dirac delta function, in other words that W_A^n and $W_{\Delta_T}^n$ are totally correlated and represent a single stochastic process (and that this behavior remains true for all combinations of parameters considered). This assumption is in line with the physics: the only random variable in the physics-based model [15] described in Section II-A is the surface profile. The rest of the model is deterministic.

2) Synthesis Scheme: For a process X (either A or Δ_T), the synthesis scheme is obtained by inverting the analysis scheme. As explained in the previous section, we use the same white process W with $\mathcal{U}([0,1])$ to synthesize both ${}^c\hat{A}$ and ${}^c\hat{\Delta}_T$. ${}^c\hat{X}^n$ can be modeled by:

$$ {}^c\hat{X}(t) = \left[F_X^{-1}(W) * h_X\right](t) \tag{6}$$

where the symbol $\hat{}$ stands for "estimated process" (i.e. simulated process), h_X is the impulse response of the inverse whitening filter H_X, and $*$ denotes the convolution product. Note that t stands for discrete time, determined according to the sampling frequency.

C. Amplitude Modulation

Numerical simulations with the physics-based model led to the conclusion that modifying the actual ball velocity weakly affects the sensation of velocity. In line with this observation, variation in the synthesis parameters with respect to perceived velocity does not reveal a particular behavior (plots and sounds

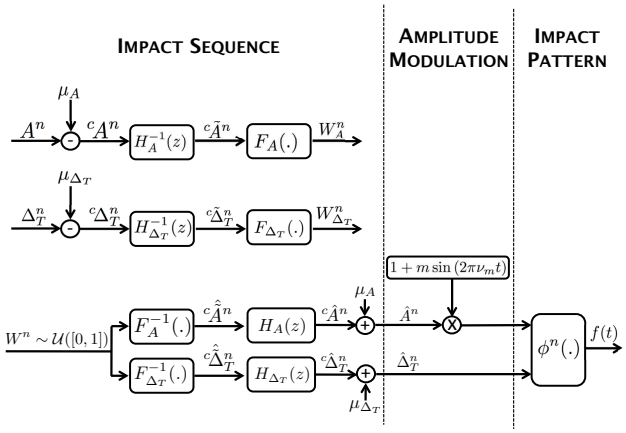


Fig. 5. Top: Analysis scheme of the impact series (A^n, Δ_T^n). Bottom: Synthesis scheme. Note that the amplitude series is modulated and then a local impact pattern is applied.

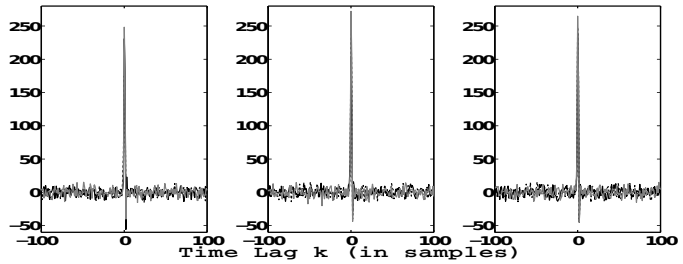


Fig. 6. Cross-correlation between the outputs W_A^n and $W_{\Delta_T}^n$ of the analysis scheme. The used parameters and colors are the same as in Fig. 4.

available on companion webpage). In contrast, the synthesis parameters exhibit significant variations with respect to the sensation of ball mass or surface roughness (see Section V). From a perceptual point of view, the asymmetry of rolling objects is likely to contribute to the sensation of velocity. However, the physics of asymmetric rolling objects is highly complex (see for instance [45]) and the link between the physical process and the perception of rolling has not yet elucidated. Some authors have proposed that a rolling object should never be considered perfectly spherical or perfectly homogeneous and have therefore applied a sinusoidal modulation to the interaction force [15], [17]. In line with these studies, we introduced an amplitude modulation that has been shown to strongly influence the perception of size and speed [36] (see Section II-B):

$$f_{out}(t) = [1 + m\sin(2\pi\nu_m t)]\,.f_{in}(t) \tag{7}$$

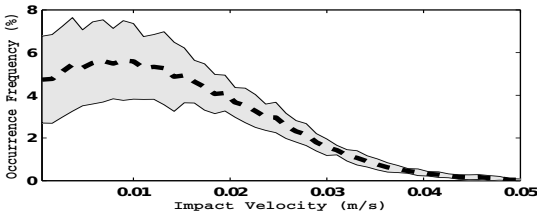


Fig. 7. Distribution of the impact velocities during the rolling interaction. Dashed black line: mean. Grey area: standard error.

with $m \in [0, 1]$ the modulation depth. The modulation frequency follows:

$$\nu_m \propto \frac{v}{R} \tag{8}$$

where v is the transversal velocity and R the ball radius.

D. Modeling the Impact Patterns

In this section, we focus on the shape of the impact patterns ϕ^n. A simplified impact model defined by a raised-cosine function is proposed in [16]:

$$\phi^n(t) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi t}{t_0^n}\right)\right] & , t \in \left[-\frac{t_0^n}{2}, \frac{t_0^n}{2}\right] \\ 0 & , \text{otherwise} \end{cases} \tag{9}$$

where t_0^n is the impact duration. In our model, t_0^n is not constant because of the non-linear behavior of the force, but is still not considered as an independent random variable. We assume that t_0^n is obtained from the impact sequence through a deterministic law that we model below. In order to simplify the notations, we have focused on a single impact pattern and omitted the impact index n.

The second term $\lambda x^\alpha \dot{x}$ in the interaction force model (3) is known to introduce hysteresis (see e.g. [44]), such that the asymmetry of impact increases with impact velocity. To investigate whether the chosen impact pattern is valid in our case, the typical impact velocities involved in a rolling interaction were first examined. Several simulations of a rolling ball with different rolling speeds (from 10 cm/s to 100 cm/s) were computed with all the combinations of the parameters κ, μ and β proposed on Fig. 4. Fig. 7 shows the distribution of the impact velocities. One can note that the impact velocities do not exceed 5 cm/s. Thus, for impact velocity from 0.1 to 5 cm/s, and with κ and μ values as on Fig. 4, the proposed model (9) fits impacts computed using equation (3) with a goodness of fit $0.90 < R^2 < 0.99$. The raised-cosine model is thus well suited for our concern.

As revealed by the non-linearity α in the contact model presented in (3), the duration of contact between the ball and the plate depends on the transversal velocity of impact [31], [33], [46]. In [33], it is shown that the contact duration varies as follows:

$$t_0 = \underbrace{\left(\frac{m_e}{k}\right)^{\frac{1}{\alpha+1}}}_{\mathcal{A}} \underbrace{\left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}}}_{\mathcal{B}} \cdot \underbrace{\int_{v_{out}}^{v_{in}} \frac{dv}{(1+\mu v)\left[-\mu\left(v-v_{in}\right)+\log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}}_{\mathcal{C}} \tag{10}$$

where v_{in} and v_{out} are the velocities at the beginning and end of the contact respectively. It can be established that v_{out} is a function of μ and v_{in} so the integral $\mathcal{C}$ depends only on μ and v_{in}. In [44], it is shown that contact duration is weakly (if at all) affected by μ, while v_{in} has a strong influence on t_0. In line with the analytical calculations from Chaigne and Doutaut who found $t_0 \propto v_{in}^{-1/5}$ [46] (in their study, the authors do not consider the dissipative term μ, which is set to 0), we propose to model the contact duration as varying with the power of the impact velocities. Since variation in μ will not affect the term $\mathcal{B}\times\mathcal{C}$, contact duration can be rewritten as:

$$t_0 = \zeta \cdot v_{in}^{-\theta} \tag{11}$$

where ζ is a constant determined by the ball mass m_e, contact stiffness k and the non-linearity factor α. Linear regression on numerical simulations of impacts with the physical model leads to $\theta \approx 0.23$ (goodness of fit $R^2 > 0.99$). Several fitting simulations on θ for various sets of parameters revealed that θ is always close to 0.23. According to the expression for t_0, increasing the ball mass (m_e) will increase the impact duration, while increasing the surface hardness (i.e. a larger k) will shorten the impact.

However, impact velocity v_{in} cannot be directly estimated from observation of the interaction force. This model is not therefore adapted for our concern. One solution consists of modeling the impact duration as a function of the impact amplitude A using the same law, but different parameters:

$$t_0 = \zeta' . A^{-\theta'} \tag{12}$$

It is obvious that impact amplitude and velocity are directly linked: the faster the ball moves upwards, the higher it goes. There is no theoretical evidence for a linear relationship between amplitude A and input velocity v_{in}. However, we assume that the non-linearity is sufficiently weak for our model to be approximately valid. It should be noted that, due to the compliance of the struck object, Avanzini and Rochesso showed that the contact duration was always higher when the exciter is coupled with a resonator

[44], and increased with the resonator mass ($m_r^{(i)}$ in (2)) [33]. However, as highlighted by the previous perceptual experiment, the interaction force f itself is perceptually relevant for the rolling interaction, and our aim is not to reproduce all the subtleties of the physical model.

E. Parameter Estimation

In this section, we describe how the parameters of our signal model are estimated from the force signals obtained with the physics-based model (cf. Section II-A). These parameters will be further used for synthesis. We consider two categories: parameters describing the impact sequence and parameters describing the impact pattern.

1) Impact Sequence Parameters: As explained in Section IV-B1, the impact sequence is characterized by the mean values μ_X, the inverse whitening filters $H_X(z)$ and the inverse cumulative distribution functions F_X^{-1}, X being the amplitude A or the time interval Δ_T. The mean values can easily be estimated using the empirical mean estimator $\mu_X = 1/N \sum_{n=1}^{N} X^n$. For the inverse whitening filters, we assume Auto-Regressive Moving Average (ARMA) models:

$$X(z) \approx H_X(z)\tilde{X}(z) \quad , \quad H_X(z) = \frac{1 + \sum_{i=1}^{p} b_i z^{-i}}{1 + \sum_{i=1}^{q} a_i z^{-i}} \tag{13}$$

The usual strategy for optimizing whitening filters consists of minimizing the energy of the residual signal. In practical terms, the coefficients a_i and b_i are obtained by minimizing $\sum_{n=1}^{N} \left[h_X^{-1} * X\right]^2 (t)$ according to a_i and b_i. Optimization is performed using an iterative Gauss-Newton strategy [47].

By applying this method to several simulations of both amplitude and time interval series, we found that one pole and one zero (p=q=1) are sufficient to whiten both processes satisfactorily. This result is supported by observations of the auto-correlation functions of the residual signals ${}^c\tilde{A}^n$ and ${}^c\tilde{\Delta}_T^n$ (same physical parameters as in Section IV-A) in Fig. 8. It can be seen that, for all the considered combinations of parameters, the auto-correlations are very close to the Dirac function, meaning that the signal whitening process has succeeded.

The cumulative distribution functions F_X^{-1} are directly related to the probability density functions of ${}^c\tilde{X}$. As Fig. 9 shows, ${}^c\tilde{A}$ and ${}^c\tilde{\Delta}_T$ follow approximately centered Gaussian distributions. Their variance σ_X^2 can be easily estimated with a least squares method. Thus, in implementing our synthesis scheme, the white noise input W and the inverse cumulative distribution functions F_X^{-1} can be successfully replaced by a single centered and normalized Gaussian random generator, scaled with standard deviations σ_A and σ_{Δ_T}, to directly generate the centered white processes ${}^c\hat{\tilde{A}}$ and ${}^c\hat{\tilde{\Delta}}_T$.

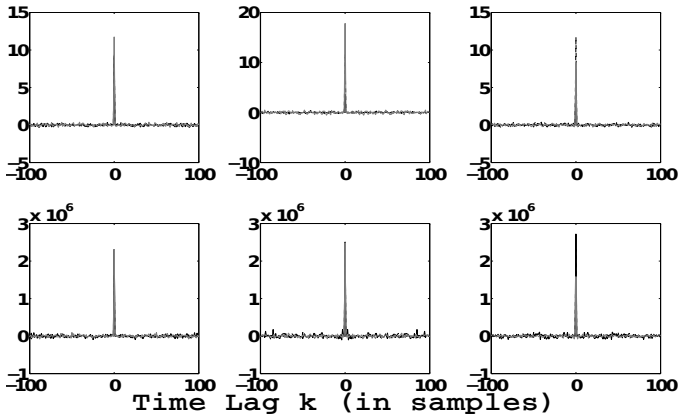


Fig. 8. Auto-correlation of the whitened signals ${}^c\tilde{A}^n$ *(Top row)* and ${}^c\tilde{\Delta}_T^n$ *(Bottom row)* for one pole and one zero whitening filters. The used parameters and colors are the same as in Fig. 4.



Fig. 9. Bars: Estimated probability density functions of the centered processes ${}^c\tilde{A}$ *(Left)* and ${}^c\tilde{\Delta}_T$ *(Right)*. Solid line: Gaussian fit. For the sake of figure readability, the results are only displayed for the parameter set κ=5·10^9 N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1} and β=-0.5. Readers can find figures for the others parameter sets on the accompanying website.

2) Impact Pattern Parameters: In the proposed model (see Section IV-D), the impact pattern duration t_0 depends on the impact amplitude A and the function has two parameters: θ' and ζ' (see Eq. (12)). We estimated t_0 as a function of A on the output of the physics-based model, and tried to fit the curve using the proposed model. We observed that for some rolling realizations, the process can lead to absurd fits from a physical point of view, due to outliers.

To avoid these situations, we first fit the law on single impacts, where no outliers can be observed. This leads to optimal estimated value for both θ' and ζ'. In the rolling-interaction case, we kept this θ' value and performed an optimization only on the parameter ζ'.

As an example, we here give the results for the simulation parameters κ=2·10^9 N.m$^{-\alpha}$.kg^{-1} and μ=1 s.m^{-1}, a ball radius of 1 cm, a rolling velocity of 20 cm/s and a fractal surface exponent β=-1. The first fitting process leads to θ'=0.29 (goodness of fit R^2=0.93). The modeled curve for the rolling

interaction fits with R^2=0.70.

V. Intuitive Control Strategy

In this part, an intuitive control strategy for the rolling sound model is described. We focus on three main perceptually relevant attributes of the rolling interaction: ball size and velocity, and surface roughness. Although the mass of the ball also could have been considered, it was not taken into account, since we assumed that there is a direct link between the perception of size and mass. For each of these high-level features, the relationships between the physical parameters and the synthesis parameters will be detailed. Then the variation range of the synthesis parameters in the proposed mapping strategy will be discussed.

A. Controlling the Ball Size

The size of the rolling object directly influences the impact duration t_0. From (10), it can be seen that t_0 is determined by ball mass m_e, contact stiffness k and the dissipative term μ, which all depend on the ball radius. Recalling that α is set to $3/2$, recalling (4), and expressing the sphere mass in function of its density, it can be shown that t_0 is proportional to the ball radius R (with both the exciter and the resonator material properties kept constant). Thus, (11) can be expressed as:

$$t_0 \propto R \cdot v_{in}^{-0.23} \tag{14}$$

The contact duration can thus be expressed as a function of impact amplitude A (see Section IV-E2):

$$t_0 \propto R \cdot A^{-0.29} \tag{15}$$

On the basis of this expression, we defined a high-level control for the perceived size S, defined as the normalized radius $S \in [0.1, 1]$, which influences the impact duration with the following mapping:

$$t_0 = 7.88 \times 10^{-4}\, S\, A^{-0.29}\ , \quad [\text{s}] \tag{16}$$

so that the bigger the S, the longer the t_0 and the bigger the perceived size.

B. Controlling the Ball Velocity

As seen in Section IV-C, the perception of velocity can be enhanced by an interaction force amplitude modulation. This sound effect is characterized by a modulation frequency, i.e. the faster the ball (or slower respectively), the faster (or slower respectively) the amplitude modulation. We thus defined a high-level

control for perceived velocity V, which is a normalized parameter $V \in [0.1, 1]$ that acts on the amplitude modulation frequency ν_m according to equation (8) and with the following mapping:

$$\nu_m = 3\,\frac{V}{S} \tag{17}$$

The multiplicative constant is set empirically. The modulation depth is set to $m = 0.3$ so that the resulting effect is audible but still natural.

C. Controlling the Surface Roughness

In the physics-based model, the roughness of the fractal surface was set by the parameter β. However, changing the value of β led to a complete change in the set of parameters describing the impact sequence in the signal model, i.e. the statistics of the time series for amplitude (μ_A,σ_A) and for time interval (μ_{Δ_T},σ_{Δ_T}) and the coefficients of the inverse whitening filters $H_A(z)$ and $H_{\Delta_T}(z)$. We therefore conducted some numerical simulations to investigate the variation of each of these synthesis parameters with respect to β. These observations are based on a rolling simulation with the following parameters: κ=$\frac{1}{3}\cdot 10^{10}$ N.m$^{-\alpha}$.kg^{-1}, μ=1 s.m^{-1}, R=1 cm, v=20 cm/s and a maximum asperities height of 0.1 μm. The roughness parameter β varies between -1.5 (smooth surface) and 0 (rough surface). The set (κ, μ) was chosen in line with the previous perceptual experiment, and then refined by the authors.

In the left-hand column in Fig. 10, the estimated variances σ_A and σ_{Δ_T} are represented as a function of β. One can note that σ_A does not exhibit a specific behavior with respect to β, while σ_{Δ_T} increases with β. The variation of σ_{Δ_T} is coherent with the physics: a rough surface with weak correlation between the asperities leads to high variability in the time intervals between impacts. In the middle column in Fig. 10, variation in the first order coefficients of the inverse whitening filters $H_A(z)$ and $H_{\Delta_T}(z)$ is shown as a function of β. For both time and amplitude series, the coefficients a_1 and b_1 increase with the roughness. In the right-hand column on Fig. 10, the estimated mean dissipative terms μ_A and μ_{Δ_T} are represented as a function of β. It can be seen that μ_A does not exhibit a specific behavior with respect to β, whereas μ_{Δ_T} increases with the roughness. The variation in μ_{Δ_T} is coherent with physics: a rough surface generates an impact sequence with longer intervals between impacts.

By simultaneously modifying all these parameters according to the above observations, we were then able to reproduce sounds evoking different surface characteristics from extremely smooth (quasi-squealing sounds, e.g. a marble sliding on a window pane) to rough and chaotic ones (like a ball rolling on damaged concrete and encountering many large asperities).

We defined a high-level control for surface roughness ρ which varies between 0 (smooth surface) and 1 (rough surface). This control is different from the fractal-surface roughness parameter β in the

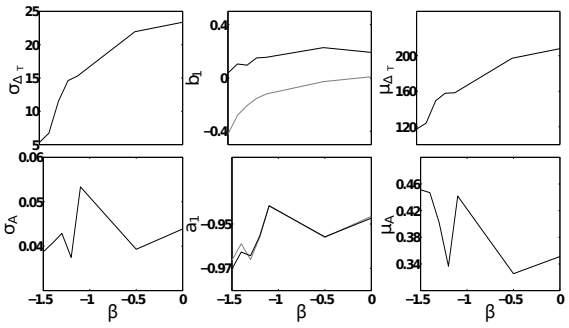


Fig. 10. Variances (left-hand column) and mean values (right-hand column) of the impact series of the signal model as a function of the roughness parameter of the physics-based model (β=-1.5: smooth surface. β=0: rough surface). Middle column: Inverse whitening filter coefficients for amplitude series A (in black) and for time-interval series Δ_T (in grey). μ_{Δ_T} and σ_{Δ_T} in samples (sample rate: 44100 Hz).

TABLE I

EXTREME VALUES FOR THE LINEAR MAPPING BETWEEN THE SURFACE ROUGHNESS CONTROL ρ AND THE SYNTHESIS PARAMETERS RELATED TO THE IMPACT SERIES (AMPLITUDE A AND TIME INTERVAL Δ_T).

		$\rho_{min}=0$	$\rho_{max}=1$
A	σ_A	0.04	0.04
	a_1	-0.97	-0.93
	b_1	0.07	0.32
	μ_A	0.43	0.27
$\mathbf{\Delta_T}$	σ_{Δ_T}	0.19 (ms)	0.85 (ms)
	a_1	-0.97	-0.93
	b_1	-0.34	0.35
	μ_{Δ_T}	3.1 (ms)	6.4 (ms)

physics-based model, which varies between $-\infty$ (perfectly smooth surface) and 0 (rough surface). The normalized range for ρ is more natural in the context of a high-level perceptual control. We assume linear mapping between ρ and each synthesis parameter. The extreme values were set according to the experiments related in the previous section, and are summarized in Table I.

VI. SUBJECTIVE EVALUATION OF THE CONTROL STRATEGY

In this section we evaluated the relevance of the proposed mapping between high-level controls and the synthesis parameters, i.e. whether modifying a high-level control had the desired perceptual effect.

To this aim, we used an experimental procedure based on a paired comparison test.

A. Subjects

Fourteen participants took part in the experiment: 9 men, 5 women, 29 years old on average (standard deviation: 9.9 years). They were all volunteer students and researchers involved with the laboratory (CNRS-LMA). None of them had prior knowledge of the stimuli used in the experiments and none of them reported hearing problems.

B. Stimuli

We considered a five-step variation for each high-level control parameter. Each synthesized stimulus lasted for $3s$ and was normalized with respect to its maximum amplitude. For each high-level control, 15 stimuli (i.e. 3 sets of 5 stimuli) were synthesized.

For the ball size assessment, the surface roughness control ρ was set to 0.5 and three values for velocity control V were used (0.1, 0.5 and 1). For the velocity assessment, ρ was set to 0.75 and three values for size control S were used (0.1, 0.3 and 0.5). For the surface roughness assessment, the velocity control V was set to 0.5 and three values for size control S were used (0.1, 0.5 and 0.9). The five steps were: $[0.1, 0.3, 0.5, 0.7, 0.9]$ for ball size assessment, $[0.1, 0.3, 0.5, 0.8, 1]$ for ball velocity assessment, and $[0, 0.25, 0.5, 0.8, 1]$ for surface roughness assessment.

C. Procedure

The test was performed in a quiet office on a standard laptop with Sennheiser HD-650 headphones. A graphical interface was specially designed for the experiment using Max/MSP software. The subjective evaluation was effectuated by conducting a paired comparison experiment, a so-called A-B listening test. For each high-level control, the sounds were evaluated in pairs within 3 distinct blocks in which only the high-level control parameter varied (e.g. when evaluating the size, only the size parameter varied within given a block, and the velocity parameters were different for the 3 blocks). This was done to avoid interaction effects between different factors. Indeed, as shown by Houben et al. [48], when both size and speed vary between two sounds, the ability to discriminate the faster or the bigger ball is impaired. For each set, 10 pairs were presented corresponding to all combinations of the 5 stimuli (identical pairs were not tested). Thus, for each high-level control, a total of 30 pairs (i.e. 3 sets of 10 pairs) were presented.

Each high-level control was evaluated separately through 3 independent sessions. In each session, sound pairs were presented to participants in a random order. They were asked to listen to each pair A-B

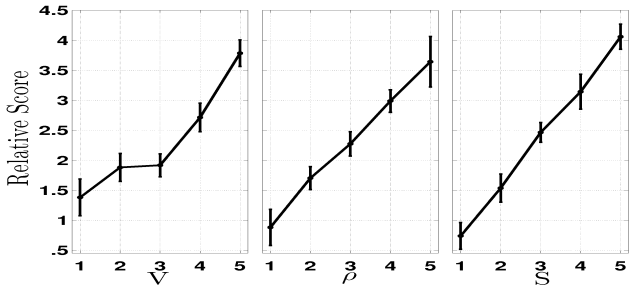


Fig. 11. Relative scores (with 95% confidence intervals) averaged across 14 subjects, for five values of each high-level control: ball velocity (left), surface roughness (middle) and ball size (right).

as often as they wished and then choose between the 3 possibilities displayed on the graphical interface "*A is bigger/is faster/is rolling over a rougher surface* than B" (icon $A>B$), "*A is as big/is as fast/is rolling over a surface as rough* as B" (icon $A=B$) or "B *is bigger/is faster/is rolling over a rougher surface* than A" (icon $B>A$), by clicking on the corresponding icon. Each session was preceded by a training run of 3 trials, in order to familiarize the participants with the sounds, the task and the interface. No feedback on their answers was given to participants during the training or during the test, and they did not know that only the high-level control values varied within a block. The sessions were balanced across subjects to annul the effect of presentation order.

D. Results

For each high-level control, data were collected into 5×5 matrices M_p^s for each participant p and for each set s of 5 stimuli (i.e. 3 matrices per subject for each high-level control). The matrices were filled as follows: if the subject p, answered $A>B$ for sound set s, (where A is the stimulus at step i, B the stimulus at step j), item $M_p^s(j,i)$ (i^{th} column, j^{th} row) was set to 1; if the subject answered $B>A$, item $M_p^s(i,j)$ was set to 1; if the subject answered $A=B$, item $M_p^s(j,i)$ was set to 0.5 if $i>j$, or $M_p^s(i,j)$ was set to 0.5 if $j>i$. The diagonal of each matrix was set to 0.5, since identical pairs were not tested. A one-dimensional subjective scale was then obtained by summing the rows of the matrix M_p^s, as proposed in [49]. To focus on an overall assessment of the high-level controls, we computed a mean one-dimensional subjective scale by averaging the 3 vectors obtained from each set of stimuli s for each high-level control. The scores obtained for each participant p reflect the perceived ball's relative velocity, size, or surface roughness according to the variation of the high-level control. Results are plotted in Fig. 11.

Repeated measures ANOVA was performed on the relative scores separately for each high-level control, with control level (5 levels) as within-subject factor. For all statistical analyses, effects were considered significant if the p value was less than or equal to 0.05. When interactions between 2 or more factors were significant, post hoc comparisons (Tukey test) between relevant condition pairs were computed.

The effect of the high-level size control was significant ($F(4, 52)$=98.895, p<0.001) and all the control levels significantly differed from each other (p<0.001 for all). Secondly, the effect of the high-level velocity control was significant ($F(4, 52)$=45.804, p<0.001). All the control levels significantly differed from each other (p<0.05), except between level 2 and 3 (p=0.867). Lastly, the effect of the roughness control was significant ($F(4, 52)$=47.491, p<0.001). All the control levels significantly differed from each other (p<0.01 for all).

E. Discussion

Our results showed that the three high-level controls produced the desired perceptual effect. Although the subjects were given no indication as the auditory cue(s) tested, the following observations were made. For the size control, subjects naturally associated lower frequency sounds (due to longer duration of impacts) and/or slower amplitude modulation with a perception of bigger balls. For the surface roughness control, participants reported a perception of surface roughness evoked by the sound, and in a coherent manner, they associated sounds with lower values for surface roughness with the evocation of an extremely smooth surface (on which the ball almost squeals), and sounds presenting more audible separate impacts with a perception of a rougher surface. For the velocity control, the results were also broadly satisfactory and were in line with Houben who concluded that amplitude modulation influences the perception of speed [36]. However, we found a lack of significant difference between sounds of level 2 and 3. This confusion may be due to the linear "discretization" of the velocity control V, which may not correspond to our perception, meaning that the perceived difference between certain velocity values is probably not precisely correlated to the numerical differences. A formal calibration experiment would be needed to adjust the high-level control mapping in order to propose perceptually-regular variation intervals for these controls. Simulating the time-varying amplitude of the surface's eigenmodes may also improve the perception of velocity, since it will result in an audible spectral cue [14].

VII. Conclusion and Future Perspectives

In this paper, we addressed the design of a rolling sound synthesizer. The goals of this study were to identify the perceptual features responsible for perception of rolling movement and to further propose a

generic, real-time compatible rolling sound synthesis model with intuitive controls.

In the first part of the study, a perceptual test was performed to identify the main perceptual features (i.e. transformational invariant) responsible for the perception of rolling. For this purpose, a physics-based model presented in [15] was used to generate the stimuli. The model was particularly interesting, since it produced evocative rolling sounds and made it possible to investigate the perceptual relevance of the different physical sources included in the model. For instance, it was possible to modify the contribution from the resonant object by simulating more or less rigid surfaces. The listening test showed that the non-linear interaction force has a great impact on the perception of "rollingness". This force can be viewed as series of impacts characterized by particular statistical properties.

Based on these observations, we proposed a signal model from a complete analysis-synthesis scheme. In particular, we provided a method to estimate and simulate the statistics of the impact sequences, and the contact duration versus the impact amplitude. In addition to the impact sequence, amplitude modulation effects were added to enhance the perception of velocity of the rolling object. Further, the shapes of the impact patterns were modeled based on [16].

Finally, high-level controls (i.e. velocity, size of the rolling ball and roughness of the surface on which the ball is rolling) were defined and validated using a perceptual experiment. The proposed synthesis model can generate perceptually satisfactory synthetic sounds (cf. sound examples on the companion web page). As pointed out by some authors [43], [50], the rolling sound could be rendered more realistic by taking into account additional effects such as the amplitude variation of the modes over time when the ball is rolling on a finite surface. This effect is currently being implemented in the synthesizer using a comb-filtering process. The low level of complexity in implementing the model meets the real-time requirements of interactive applications in virtual reality. In addition, the model can be directly controlled with high-level parameters related to the description of events in the scene in which the user is involved. Hence, sounds could be generated in line with the user actions in a virtual environment might be generated, meaning that this could offer a good alternative to prerecorded sound databases.

More interestingly, the proposed model is actually generic enough to generate impact series that comply with various statistics. Therefore, by defining appropriate statistical properties, the proposed model can synthesize excitation signals enabling the evocation of other kinds of continuous interactions such as scratching or rubbing. Furthermore, in line with the design of a navigation space for perceived materials [24], it should be possible to build an intuitive control space for continuous interactions, allowing the user to synthesize sounds that evoke rubbing, scratching, rolling and to continuously morph between them (e.g. synthesize a rubbing sound and slowly transform it into a rolling sound). The authors have already

published a preliminary study on the perceptual differences between rubbing and scratching sounds, and on the synthesis and morphing between these interactions [21]. For these types of sounds, the source signal was modeled as a series of micro-impacts characterized by time series with specific amplitudes and time intervals. A first navigation space version for perceived interactions was proposed in [22], addressing the design of a mapping strategy allowing continuous interpolation between the synthesis parameter values. Moreover, the calibration of each interaction could be improved through the use of analysis-synthesis methods on recorded sounds as proposed in [18], [19].

Going further in our goal of offering infinite possibilities for sound creation and manipulation, the proposed *action-object* framework is suitable for the creation of *sound metaphors*. This means that by freely combining evoked objects (based on structural invariants) with evoked actions (based on transformational invariants), unheard-of action-object combinations could be synthesized. To give an example, physically impossible action-object combinations such as "rolling water drops" could be simulated by combining synthesized water drops (obtained from [51], [52]) with the rolling action. With further experiments on new sound textures, the salience of the invariant sound morphologies related to the rolling interaction highlighted in this study (but also to other actions such as scratching for instance) can be accurately examined. Such transformational methods are useful for musical as well as for sonification applications. In particular, direct applications can be found in the current "MetaSon" project in which specific sound design issues are raised by the car industry. Since the electric cars of the future will be very quiet, they should be provided with sounds that clearly evoke a rolling object and that are recognizable as potentially dangerous objects. The proposed model together with the intuitive control strategy constitutes a relevant tool for such investigations.

ACKNOWLEDGMENTS

The authors would like to thank the French National Research Agency (ANR) which funded this work under the MetaSon: Métaphores Sonores (Sound Metaphors) project (ANR-10-CORD-0003) within the framework of the CONTINT 2010 program: http://metason.cnrs-mrs.fr/home.html. The authors are also grateful to the anonymous reviewers for their insightful comments and to Charles Gondre for real-time implementation.

REFERENCES

[1] N. Böttcher, "Current problems and future possibilities of procedural audio in computer games." *J. Gaming & Virtual Worlds*, vol. 5, no. 3, 2013.

[2] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju, "Sound synthesis for impact sounds in video games," in *Symp. on Interactive 3D Graphics and Games*. ACM, 2011.

[3] A. Farnell, *Designing sound*. MIT Press Cambridge, 2010.

[4] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS ONE*, vol. 8, no. 12, p. e82491, 2013.

[5] J. Danna, J.-L. Velay, V. Paz-Villagran, A. Capel, C. Petroz, C. Gondre, E. Thoret, M. Aramaki, S. Ystad, R. Kronland-Martinet, "Handwriting movement sonification for the rehabilitation of dysgraphia," in *Proc. of the 10th Int. Symp. on Computer Music Multidisciplinary Research (CMMR)*, 2013, pp. 200–208.

[6] M. W. Rodger, W. R. Young, and C. M. Craig, "Synthesis of walking sounds for alleviating gait disturbances in Parkinson's disease," *IEEE Trans. on Neural Syst. and Rehabil. Eng.*, vol. (preprint), no. 99, 2013.

[7] R. Kronland-Martinet, S. Ystad, and M. Aramaki, "High-level control of sound synthesis for sonification processes," *AI & society*, vol. 27, no. 2, pp. 245–255, 2012.

[8] C. Michaels and C. Carello, *Direct perception*. Prentice-Hall Englewood Cliffs, NJ, 1981.

[9] W. Warren and R. Verbrugge, "Auditory perception of breaking and bouncing events: A case study in ecological acoustics." *J. Exp. Psychol. Hum Percept. Perform*, vol. 10, no. 5, pp. 704–712, 1984.

[10] S. McAdams and E. Bigand, *Thinking in sound: The cognitive psychology of human audition.* Oxford Science Publications, 1993.

[11] G. Lemaitre and L. Heller, "Auditory perception of material is fragile while action is strikingly robust," *J. Acoust. Soc. Amer.*, vol. 131, pp. 1337–1348, 2012.

[12] B. Atal and S. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, no. 2B, pp. 637–655, 1971.

[13] X. Rodet, P. Depalle, and G. Poirot, "Diphone sound synthesis based on spectral envelopes and harmonic/noise excitation functions," in *Proc. Int. Comput. Music Conf.*, 1988, pp. 313–321.

[14] C. Stoelinga, "A psychomechanical study of rolling sounds," Ph.D. dissertation, ENSTA ParisTech, 2007.

[15] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE Multimedia*, vol. 12, no. 2, pp. 60–69, 2005.

[16] K. Van Den Doel, P. Kry, and D. Pai, "Foleyautomatic: physically-based sound effects for interactive simulation and animation," in *Proc. Int. Conf. Computer Graphics and Interactive Techniques*, 2001.

[17] D. Hermes, "Synthesis of the sounds produced by rolling balls," *Internal IPO report no. 1226, Center for User-System Interaction, Eindhoven, The Netherlands*, 1998.

[18] M. Lagrange, G. Scavone, and P. Depalle, "Analysis/synthesis of sounds generated by sustained contact between rigid objects," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 509–518, 2010.

[19] J. Lee, P. Depalle, and G. Scavone, "Analysis/synthesis of rolling sounds using a source-filter approach," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2010.

[20] W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecol. Psychol.*, vol. 5, no. 4, pp. 285–313, 1993.

[21] S. Conan, M. Aramaki, R. Kronland-Martinet, E. Thoret, and S. Ystad, "Perceptual differences between sounds produced by different continuous interactions," in *Proc. Acoustics*, 2012.

[22] S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, R. Kronland-Martinet, and S. Ystad, "Navigating in a space of

synthesized interaction-sounds: Rubbing, scratching and rolling sounds," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2013.

[23] M. Aramaki, C. Gondre, R. Kronland-Martinet, T. Voinier, and S. Ystad, "Thinking the sounds: An intuitive control of an impact sound synthesizer," in *Proc. Int. Conf. Auditory Display*, 2009.

[24] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Controlling the perceived material in an impact sound synthesizer," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 301–314, 2011.

[25] M. Mathews and J. Smith, "Methods for synthesizing very high Q parametrically well behaved two pole filters," in *Proc. Stockholm Musical Acoust. Conf. (SMAC)*, 2003.

[26] J. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, and J. Vion-Dury, "Categorization and timbre perception of environmental sounds in schizophrenia," *Psychiatry Research*, vol. 189, no. 1, pp. 149–152, 2011.

[27] C. Stoelinga and A. Chaigne, "Time-domain modeling and simulation of rolling objects," *Acta Acust. United Ac.*, vol. 93, no. 2, pp. 290–304, 2007.

[28] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates. I. Theory and experiments," *J. Acoust. Soc. Amer.*, vol. 109, pp. 1422–1432, 2001.

[29] C. Lambourg, A. Chaigne, and D. Matignon, "Time-domain simulation of damped impacted plates. II. Numerical model and results," *J. Acoust. Soc. Amer.*, vol. 109, pp. 1433–1447, 2001.

[30] P. Morse and K. Ingard, *Theoretical acoustics*. Mc Graw-Hill Book Company, 1968.

[31] E. Falcon, C. Laroche, S. Fauve, and C. Coste, "Behavior of one inelastic ball bouncing repeatedly off the ground," *Eur. Phys. J. B*, vol. 3, no. 1, pp. 45–57, 1998.

[32] K. Hunt and F. Crossley, "Coefficient of restitution interpreted as damping in vibroimpact," *J. Appl. Mech.*, 1975.

[33] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2001.

[34] F. Avanzini, M. Rath, and D. Rocchesso, "Physically-based audio rendering of contact," in *IEEE Int. Conf. on Multimedia and Expo, 2002.* , vol. 2.0.4emIEEE, 2002, pp. 445–448.

[35] R. Sayles and T. Thomas, "Surface topography as a nonstationary random process," *Nature*, vol. 271, pp. 431–434, 1978.

[36] M. Houben, "The sound of rolling objects: Perception of size and speed," Ph.D. dissertation, Technische Universiteteit, Eindhoven, 2002.

[37] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[38] R. Badeau, R. Boyer, and B. David, "EDS parametric modeling and tracking of audio signals," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, 2002.

[39] E. Schuijers, W. Oomen, B. d. Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Proc. 114th Conv. Audio Eng. Soc.*, 2003.

[40] M. Rath, "Interactive realtime sound models for human-computer interaction: A sound design concept and applications," PhD thesis, 2004.

[41] U. M. Ascher and L. R. Petzold, *Computer methods for ordinary differential equations and differential-algebraic equations*, Siam, 1998.

[42] S. Papetti, F. Avanzini, and D. Rocchesso, "Numerical methods for a nonlinear impact model: A comparative study with closed-form corrections," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2146–2158, 2011.

[43] E. Murphy, M. Lagrange, G. Scavone, P. Depalle, and C. Guastavino, "Perceptual evaluation of rolling sound synthesis," *Acta Acust. United Ac.*, vol. 97, no. 5, pp. 840–851, 2011.

[44] F. Avanzini and D. Rocchesso, "Physical modeling of impacts: Theory and experiments on contact time and spectral centroid," in *Proc. Conf. Sound, Music Comput.*, 2004.

[45] W. Theron, "Analysis of the rolling motion of loaded hoops," Ph.D. dissertation, University of Stellenbosch, 2008.

[46] A. Chaigne and V. Doutaut, "Numerical simulations of xylophones. II. Time-domain modeling of the vibrating bars," *J. Acoust. Soc. Amer.*, vol. 101, no. 1, pp. 539–557, 1997.

[47] B. Ninness, A. Wills, and S. Gibson, "The University of Newcastle identification toolbox (unit)," in *Proc. World Congr. Intern. Fed. Automat. Contr.*, 2005.

[48] M. Houben, A. Kohlrausch, and D. Hermes, "Perception of the size and speed of rolling balls by sound," *Speech Commun.*, vol. 43, no. 4, pp. 331–345, 2004.

[49] H. A. David, *The method of paired comparisons*. Oxford University Press, 1988.

[50] C. Stoelinga, D. Hermes, A. Hirschberg, and A. Houtsma, "Temporal aspects of rolling sounds: A smooth ball approaching the edge of a plate," *Acta Acust. United Ac.*, vol. 89, no. 5, pp. 809–817, 2003.

[51] K. van den Doel, "Physically based models for liquid sounds," in *Proc. Int. Conf. Auditory Display*, 2004.

[52] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone, "A 3-D immersive synthesizer for environmental sounds," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1550–1561, 2010.



Simon Conan was born in Caen, France, in 1988. He received a M.S. degree in Engineering from Ecole Centrale Marseille and a M.S. degree in Mechanics, Physics and Engineering (speciality in Acoustics) from the Aix-Marseille University in 2012. He is currently pursuing a Ph.D. degree at the Laboratoire de Mécanique et d'Acoustique, Marseille, France. He is currently involved in the research project "MetaSon", financed by the French National Research Agency. His main interests are analysis and synthesis of sounds, high-level control of synthesis processes and auditory perception.



Olivier Derrien (M'10) was born in Aix-en-Provence, France, in 1974. He received the State Engineering degree in 1998 and the Ph.D. degree in audio processing in 2002, both from Télécom ParisTech (formerly ENST), Paris, France. From 2002 to 2003, he was a Teaching and Research assistant at the University of Paris-XI, Orsay, France. He joined the University of Toulon, Toulon, France, in 2003 as an Associate Professor in the field of telecommunications. In 2008, he joined the Laboratory of Mechanics and Acoustics (LMA), Marseille, France, as an Associate Researcher. His research interests include audio signal processing, especially audio coding, audio synthesis for virtual reality, music synthesis and audio effects. He is a co-author of 4 journal papers and 16 international conference papers.



Mitsuko Aramaki (M'09) received the Ph.D. degree from the Aix-Marseille University, Marseille, France, in 2003, for her work on analysis and synthesis of impact sounds using physical and perceptual approaches. She is currently a researcher at the National Center for Scientific Research (CNRS). She has been working at the Institut de Neurosciences Cognitives de la Méditerranée (Marseille, France) from 2006 to 2011 and since 2012, she joined the Laboratoire de Mécanique et d'Acoustique (Marseille, France). Her research mainly focused on sound modeling, perceptual and cognitive aspects of timbre, neuroscience methods and multimodal interactions in the context of virtual/augmented reality. She supervised 4 PhD theses and has been involved in industrial contracts with renowned companies such as Orange Labs and PSA Peugeot-Citroën. She is a member of the CMMR (Computer Music Multidisciplinary Research) steering committee and is regularly involved in the chair panel of the conference. She published 20 articles in international peer reviewed journals, more than 30 articles in international conference proceedings and co-edited 4 books published by Springer in their series LNCS.



Sølvi Ystad received her degree as a civil engineer in electronics from NTH (Norger Tekniske Høgskole), Trondheim, Norway in 1992 and her Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille, France, in 1998. She is currently a Researcher at the National French Research Center (CNRS) in the Laboratoire de Mé canique et d'Acoustique in Marseille, France. Her research activities are related to sound modeling with a special emphasis on the identification of perceptually relevant sound structures to propose intuitive user interfaces for controlling synthesized sounds. She was in charge of the ANR funded research project "Towards the sense of sounds," (http://www.sensons.cnrs-mrs.fr) from 2006-2009 and is currently participating in the ANR funded projects Metason, Physis and Potion.



Richard Kronland-Martinet (M'09-SM'10) has a scientific background in theoretical physics, supplemented by a specialization in acoustics. In 1989 he obtained a Doctorate of Science (Habilitation Research) from the University of Aix-Marseille, for his work on analysis, synthesis and processing of sounds by time-frequency and time-scale (wavelet) approaches. Since 1998 he is Director of Research at the Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille-France. His scientific activity concerns the science of sounds addressed from a multidisciplinary sight. He has been an important actor in the development of time-scale analysis methods and in their use for analysis, processing and synthesis of sound signals and music. The combination of physical concepts and methods of non-stationary signal processing has enabled him to develop many paradigms of synthesis, especially for instrumental sounds (piano, flute) but also for environmental sounds. His interest in perceptual and cognitive aspects associated with sounds have more recently led him to undertake research on the intuitive control of sounds over the process able to reproduce perceptual effects corresponding to high-level attributes.