A Perceptually Evaluated Signal Model: Collisions Between a Vibrating Object and an Obstacle

Samuel Poirot, Stefan Bilbao, Mitsuko Aramaki, Sølvi Ystad, Richard Kronland-Martinet

A Perceptually Evaluated Signal Model: Collisions Between a Vibrating Object and an Obstacle.

Samuel Poirot*, Stefan Bilbao†, *Senior Member, IEEE*, Mitsuko Aramaki*, *Senior Member, IEEE*, Sølvi Ystad*, and Richard Kronland-Martinet*, *Senior Member, IEEE*
* Aix Marseille Univ, CNRS, PRISM, Marseille, France
†Acoustics and Audio Group, University of Edinburgh, United Kingdom

***Abstract*—The collision interaction mechanism between a vibrating string and a non-resonant obstacle is at the heart of many musical instruments. This paper focuses on the identification of perceptually salient auditory features related to this phenomenon. The objective is to design a signal-based synthesis process, with an eye towards developing intuitive control strategies. To this end, a database of synthesized sounds is assembled through physics-based emulation of a string/obstacle collision, in order to characterize the effect of collisions on time-frequency content. The investigation of this database reveals characteristic time-frequency patterns related to the position of the obstacle during the interaction. In particular, a frequency shift of certain modes is apparent for strong interactions, which, alongside the generation of new frequency components, leads to increased perceived roughness and inharmonicity. These observations enable the design of a real-time compatible signal-based sound synthesis process, with a mapping of synthesis parameters linked to the perceived location of the obstacle. The accuracy of the signal model with respect to the physical model sound output and recorded sounds was evaluated through listening tests: time-frequency patterns reproduced by the signal model enabled listeners to precisely recognize the transverse location of the obstacle.**

***Index Terms*—Signal synthesis, Signal design, Acoustic signal processing.**

I. Introduction

THE numerical modeling of contact mechanics for realistic sound synthesis has applications in both music and game audio [1] [2]. For continuous interactions, the modeling of friction and self-oscillation phenomena [3] has led to the synthesis of convincing environmental sounds [4] [5] [6] as well as sounds of various musical instruments [7], including rubbed strings [8], the glass harmonica and the Tibetan bowl [9][10]. Collisions and the vibration of an object under a unilateral constraint generate audible behavior typical of many musical instruments. The most direct example is the case of the striking mechanism in instruments such as the piano [11] [12]. But collisions play a key role in many other instruments in which the resonator is in intermittent contact with a barrier. Examples include the string/fretboard interaction in the guitar [13], Indian instruments such as the sitar, the tanpura [14] and the rudra veena, for which the string is in partial contact with a sloping bridge, and also the slap bass [15], the prepared piano [16] and the snare drum [17].

The perturbation of a vibrating object by an obstacle is a complex non-linear phenomenon that gives rise to a wide variety of identifiable sound events. The interaction can be weak and manifest itself as a series of impacts causing a redistribution of energy among the modes without changing the modal parameters of the resonant object. In particular, this type of interaction is observed in the case of slight contact such as in the case of light rattling elements on a vibrating object, or in the case of the tanpura and the snare drum. On the other hand, modal parameters (frequencies and damping) can undergo significant changes when the interaction becomes less intermittent or more abrupt, as in the case of sounds produced by prepared strings (in a prepared piano or guitar). The objective of this study is to design a signal model for sound synthesis applications that can evoke these different events.

The development of physics-based sound synthesis algorithms has numerous applications, including the design of virtual musical instruments [18][19]. Comparisons of physical models of collisions for sound synthesis with measured sounds have been carried out with convincing results [20][21]. Furthermore, recent increases in the computational power of consumer-grade hardware allow real-time synthesis for increasingly complex physical models, though computational cost is still a concern [22] [23] [24] [25]. Other approaches inspired by physical models, also involving collision modeling, include mass-interaction networks [26][27]. However, the computational cost involved in simulating a physical model limits the range of sound output and the ability to control the synthesis process in an intuitive manner.

In contrast with physical models, signal models, based on perceptually relevant signal features, allow direct modelling of sound targeted to the way in which it is perceived. The challenge is to determine the perceptually salient features allowing the recognition of the sound event. The direct application is a sound synthesis model adapted to nonlinear musical instruments with low computational cost and can yield real-time event-driven synthesis of sounds in virtual or augmented reality environments, a particularly active field of research[28] [29]. For example, Gan proposes an interactive multimodal simulation platform based on impact sounds [30] [31] that could be compatible with the model proposed in our study. The outcomes of such a study can be useful for *sonic interaction design* [32] and the synthesis of *new* sounds [33].

Recently, various studies have used data-driven methods for sound generation with convincing results [34] [35]. Our approach is complementary to these approaches: the direct

Manuscript received ...

modelling of sound morphologies is fully interpretable and transparent but requires more information a priori and more care in the model design. In future studies it would be interesting to link the two approaches using physical and signal models to create large deep-learning training datasets.

Our approach is inspired by the ecological approach to auditory events [36] [37]. Adapted from the field of visual perception [38], it suggests the existence of invariant structures (specific patterns in the acoustic signal) that carry the relevant information to perceptually recognize sound events. More specifically, this study is in line with the action-object paradigm [39] [40] [41], which allows us to link the semantic description of a sound as the result of an 'action' on an 'object' to a sound synthesis process. This approach combines physics-based sound synthesis with detailed psychophysical tests, as proposed in recent studies [42] [43]. In this paper, we seek to characterize the 'action' of disturbing the vibrations of an object with an obstacle. In particular, we aim to define a signal model allowing the synthesis of sounds ranging from sparse collisions that do not modify the modal parameters of the vibrating object to the coupling between the vibrating object and the obstacle resulting in a variation of the modal frequencies of the vibrating object. Longer term, we seek an intuitive mapping between synthesis parameters and evocative semantic labels. Our methodology consists in analyzing a corpus of sounds representative of the phenomenon in order to determine the perceptually salient features. We then propose a synthesis model that we validate and calibrate by means of listening tests.

The simple but representative case of a 1-D resonant object (a stiff string) colliding with a unilateral pointwise obstacle has been selected. This example has the advantage of a representation in terms of a small number of modes, which simplifies observations on the time-frequency representation. However, our observations can easily be extrapolated to more complex objects. Also, we do not distinguish the disturbance resulting from a direct action of a user or from an interaction with an object in the environment. We believe that the sounds generated in these two situations can be modeled in the same way even if the control issues should be considered differently.

A sound database that is representative of sounds produced with this type of nonlinear interaction has been assembled. To produce the sounds, we have chosen a physical approach based on recent investigations of numerical modeling of collisions in musical instruments [44] [45] that generates realistic sounds defined by physical parameters [20] (see Section II). The use of a physical model rather than recorded sounds allows us to generate an infinite number of samples under perfectly controlled conditions. We have nevertheless added a few recorded sounds to the corpus for verification purposes. A reduced number of sounds considered characteristic of the particular phenomenon studied in this paper are presented. We then investigate perceptually-relevant morphologies responsible for the evoked nature of the interactions (see Section III), design a signal-based synthesis process (see Section IV), and evaluate it through a listening test with a view towards perceptual control (see Section V). In the final section, we provide some general conclusions and outline future research directions.

The stimuli and sound examples are available online at the following address [46].

II. Physical modeling and synthesis of the sound corpus

In this section, a physical model of the collision of a vibrating string with an obstacle is presented, (see Fig. 1). A numerical simulation model, allowing for the synthesis of a corpus of realistic sounds, is outlined.

A. String Model

The transverse dynamics of a linear stiff string are described by the following equation, commonly used in physics-based sound synthesis [47][48][49][50]:

$$\partial_t^2 u = \gamma^2 \partial_x^2 u - \kappa^2 \partial_x^4 u - 2\sigma_0 \partial_t u + 2\sigma_1 \partial_t \partial_x^2 u + \frac{1}{\rho S}\left(\delta(x-x_0)F_0(t) - \delta(x-x_1)F_1(t)\right) . \quad (1)$$

Here, $u(x,t)$ is the transverse displacement of the stiff string, as a function of spatial coordinate $x \in [0, L]$, for a string of length L, and for time $t \geq 0$. ∂_t and ∂_x indicate partial differentiation with respect to time t and spatial coordinate x, respectively. Initial conditions are assumed quiescent, and boundary conditions are chosen to be of simply supported type, so that $u = \partial_x^2 u = 0$ at $x = 0, L$.

This equation of motion incorporates various effects: the first term on the right hand side is due to tension in the string, the second due to stiffness, and the third and fourth allow two-parameter control over frequency-dependent loss. In this study, parameters are chosen to correspond to a steel guitar string tuned to $G\#4$, and are as indicated in Table I below. The fundamental frequency f_1, in Hz, is approximately

$$f_1 = \frac{\gamma}{2L} . \quad (2)$$

The string is assumed excited by a vertical downward force of amplitude $F_1 = F_1(t)$, acting pointwise at $x = x_1$, and modeled through the use of a spatial Dirac delta function $\delta(\cdot)$. We approximate the excitation force due to plucking at time $t = 0$ through:

$$F_1(t) = \begin{cases} \frac{A_1}{2}\left(1 - \cos(\frac{\pi t}{\Delta t})\right) & 0 \leq t < \Delta t \\ 0 & \text{else} \end{cases} \quad (3)$$

where A_1 is the maximum amplitude of the excitation, and Δt is the duration. See Table I.

B. Collision Modeling

The final term on the right-hand side of (1) represents the vertical collision force (amplitude $F_0 = F_0(t)$, located at $x = x_0$) due to a pointwise barrier positioned at vertical height $y = y_0$ above the string's rest position (see Fig. 1). Collisions with a rigid barrier are modeled as the contact with a stiff unilateral non-linear spring. We use the following model :

$$F_0(t) = -H(t - t_0)\frac{d\Phi}{du} \quad (4)$$

TABLE I
PARAMETER SET FOR THE STRING MODEL, INCLUDING MATERIAL AND GEOMETRIC STRING PARAMETERS (CORRESPONDING TO A STEEL STRING OF 1MM DIAMETER, SEE [51]), THE EXCITATION AND COLLIDING OBJECT (SEE EQ.(1) TO (5)).

Parameter	Role	Value
γ	wave speed	404.02 m·s^{-1}
κ	stiffness constant	1.297 m^2·s^{-1}
σ_0	loss parameter	0.05 s^{-1}
σ_1	loss parameter	0.002 m^2·s^{-1}
ρ	density	7.8×10^3 kg· m^{-3}
S	cross-sectional area	7.85×10^{-7}m^2
L	string length	0.5 m
x_1	excitation location	$3L/20$ m
A_1	excitation amplitude	200 N
Δt	excitation duration	1 ms
x_0	object location	$\in$ [0, L] m
y_0	object height	$\in \mathbb{R}$ in m
K	object stiffness	5×10^{10}N·m$^{-\alpha}$
α	object nonlinear exponent	1.4
t_0	object activation time	0.5 s

Fig. 1. Representation of the string configuration, as given in (1). An excitation force F_1 is applied at spatial coordinate $x = x_1$, and an obstacle is located at (x_0, y_0) and modeled by a stiff non-linear spring, of stiffness K and nonlinear exponent α.

where a collision potential Φ is defined as

$$\Phi = \frac{K}{\alpha+1}[u(x_0) - y_0]_+^{\alpha+1} \geq 0 \tag{5}$$

with a stiffness parameter $K \geq 0$, and exponent $\alpha > 1$, as per standard models of elastic collision [52]. In this article, the notation $[\cdot]_+$ indicates the "positive part of", i.e., $[\zeta]_+ = \frac{1}{2}(\zeta + |\zeta|)$. The Heaviside function H is used here to indicate that the object is assumed activated at time $t = t_0$.

An efficient finite difference scheme is employed in order to simulate (1) and is described in full in [44]. It is used in order to generate a sound corpus, allowing for the study of the perceptual effects of the position of the obstacle at (x_0, y_0).

The generation of a large number of samples led us to identify common features for a rigid obstacle ($K > 10^9$ N·m$^{-\alpha}$) located far from the ends of the string ($x_0 \in$ [L/8, 7L/8] m). The following section (III) presents time-frequency representations of a few selected sounds from this corpus, allowing a better appreciation of this particular phenomenon. Also, we selected 27 5-second sounds for 3 different values of x_0 and 9 different values of y_0 to be used as stimuli in the listening test described in Section V. The sampling frequency is set to 44100 Hz for the generation of the corpus.

III. INVESTIGATIONS ON SOUND MORPHOLOGIES

Given the assembled corpus, it is possible to examine time-frequency behavior in detail, and particularly variations with respect to the object location (x_0, y_0). A time-frequency

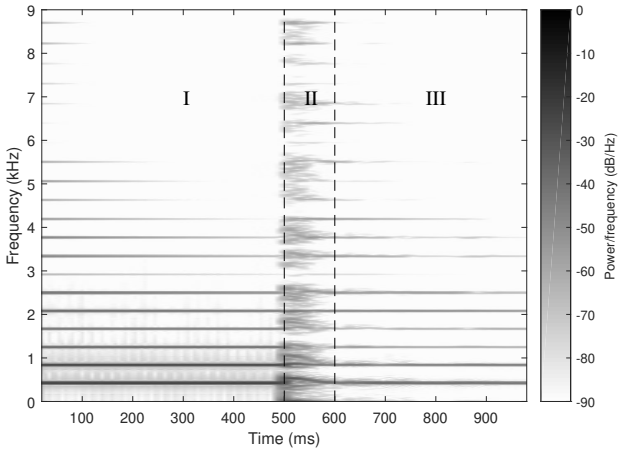

Fig. 2. Spectrogram of a signal synthesized with the physical model for $x_0 = 5L/12$; $y_0 = 0.1 \times U(x_0, t_0)$ (see Eq.(6) for a definition of $U(x,t)$). 2048-point Kaiser windows ($\beta = 5$) with 1800 samples of overlap are employed. The frequency range 0-9000 Hz is displayed. Shown are the three phases of string vibration induced by the interaction with the obstacle: I-before interaction, II- during interaction, III- after interaction.

representation for a typical sound is shown in Fig. 2. Three main phases [53] may be distinguished:

I When the string is plucked, the signal is quasi-harmonic and vibrates according to model (1) under linear conditions.

II When the rigid barrier collides with the vibrating string, the number of spectral components increases (the *interaction phase*). The interaction lasts until there is no more contact between the string and the obstacle.

III Finally, string vibration returns to the linear regime with an altered modal state.

A. Interaction phase

The interaction phase II is the main point of interest in this paper, and results from repeated irregular collisions between the string and obstacle. Due to the high spring stiffness of the obstacle, each collision gives rise to a burst of high frequency energy propagating through the string (see Fig. 2).

As a running measure of string vibration amplitude, it is useful to define

$$U(x,t) = \max_{t' \in W(t)} |u(x,t')| \qquad W(t) = H(t+\tfrac{1}{2f_1}) - H(t-\tfrac{1}{2f_1}) \tag{6}$$

which is the maximum absolute value of the displacement of the string at x, averaged over a single period duration (note that the fundamental frequency f_1 is as defined in (2)). If $U(x_0,t) > y_0$, collisions with the obstacle will normally occur at each oscillation of the string, leading to an increase in high frequency energy. However, even though the collision model itself is lossless here, over-all damping of string vibration is increased, as losses in the string model are stronger at high frequencies.

By examining the time-frequency behavior shown in Fig. 2, the spectral content of the signal is modified as soon as the obstacle appears (from $t = t_0 = 0.5$ s, $U(x_0,t) \gg y_0$). In particular, the interaction induces frequency shifts of the

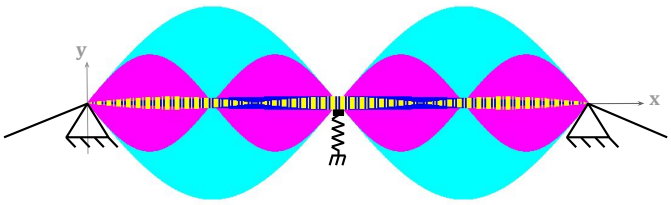

Fig. 3. Representation of the mode shapes that can vibrate without interacting with an obstacle positioned near the middle of the string. We observe large admissible amplitudes for the even-numbered modes (mode 2: cyan; mode 4: magenta). Conversely, the odd-numbered modes interact with the obstacle at low amplitudes (mode 1: blue; mode 3: yellow) and will therefore be strongly modified by the presence of the obstacle. In general, a mode with a node near the obstacle is not affected much by the obstacle.

modes, harmonic distortion and mode coupling between the distorted harmonics and the other modes. Perceptually, the simultaneous presence of these new frequency components creates beating tones, roughness, and a noise-like signal. This process, causing significant losses, implies a rapid decrease in the displacement amplitude of the string at the point of interaction $U(x_0, t)$ until it gets lower than y_0. The duration of this phenomenon depends on the distance y_0, the damping coefficients of the string $\{\sigma_0, \sigma_1\}$, the stiffness of the barrier K, and the non-linear exponent α. An overview of the effects of these parameters on the interaction phase is available in [51]. In this paper, we focus on the perceptual cues related to the position of the obstacle (x_0,y_0).

After the interaction phase, the signal corresponds to the natural vibration of the stiff string (quasi-harmonic), but with a different distribution of modal energy. Residual collisions may occur, inducing slight harmonic distortion.

B. Evolution of the morphologies with respect to the location of the rigid barrier

During the interaction phase II, some collision-related effects (frequency shift, rapid decrease in amplitude, harmonic distortion) vary depending on the position of the rigid barrier (x_0,y_0). In particular, when x_0 is an integer fraction of the string length L, and for $y_0 \approx 0$, any mode that does not have a node at this location is extinguished (see Fig.3). Several examples are shown in Figures 4 (for $x_0 = L/2$ and $x_0 = L/3$). This phenomenon is used by guitar players to play natural harmonics. We also observe a specific pattern during the interaction phase for an obstacle located in the middle of the string: the frequency components corresponding to the odd modes (modified by the interaction) are each replaced by a pair of new components which appear on both sides in a symmetrical way. The pattern is more complex for other positions of the obstacle, as the signal is noisier and its energy is distributed around more partials (See Fig.4).

On the other hand, when y_0 increases, the interaction duration decreases and the mode amplitudes are less affected. The modes remain at levels close to their initial values before the interaction (see Fig. 5, left). For $y_0 = 0.99 \times U(x_0, t_0)$, the interaction phase is barely visible on the spectrogram (Fig. 5, right), but we can still identify slight harmonic distortion due to sparse collisions causing the string to buzz.

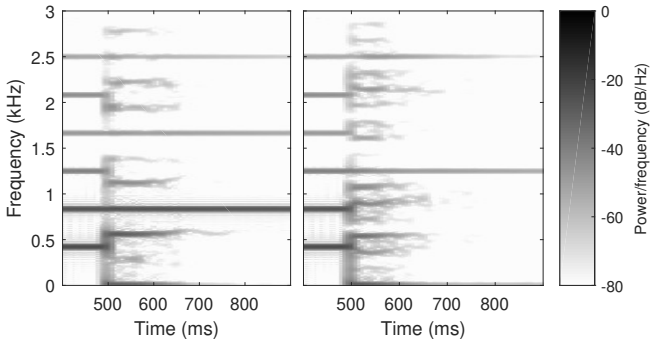

Fig. 4. *Spectrograms of two signals synthesized with the physical model for $x_0 = L/2$; $y_0 = 0$ (left) and $x_0 = L/3$; $y_0 = 0$ (right). 2048-point Kaiser windows ($\beta = 5$) with 1800 samples of overlap. The frequency range 0-3kHz is displayed. Partial tones corresponding to the modes whose modal shape includes a node at the position of the obstacle (multiple of 2 on the left, multiple of 3 on the right) are not modified by the interaction.*

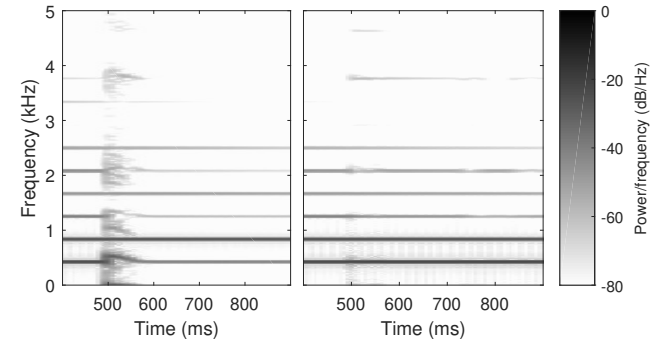

Fig. 5. *Spectrograms of two signals synthesized with the physical model for $x_0 = L/2$; $y_0 = 0.1 \times U(x_0, t_0)$ (left) and $x_0 = L/2$; $y_0 = 0.99 \times U(x_0, t_0)$ (right). 2048-point Kaiser windows ($\beta = 5$) with 1800 samples of overlap. The frequency range 0-5kHz is displayed. The effect of the interaction on the spectral content diminishes as the obstacle moves away. The rough and inharmonic aspect of the signal no longer appears for very subtle contacts (right).*

From these observations, it is clear that the important features associated with the interaction phenomenon are the the generation of high-frequency components (related to the loss processes), the rough and inharmonic character of the signal during the interaction phase (especially for a strong interaction), and the return to quasi-harmonic sound after the interaction. The aspects of the signal that encode information related to the position of the collisions are the duration of the interaction phase and the distribution of the effects on the string modes.

From a perceptual point of view, the interaction phase is characterized by roughness, as defined by Vassilakis [54]. For instance, a strong interaction leads to a buzzy, harsh sound that is produced by the presence of several tonal frequency components in narrow frequency intervals. Indeed, if the frequency difference between two components is smaller than the critical bandwidth, then a single tone is perceived either as fluctuating loudness (beating) or as roughness.

A verification of the morphologies was done by adding recorded sounds to the corpus (available for listening on the accompanying site [46]). The sounds were recorded with an

electric guitar (original Fender stratocaster, American made) recorded directly through a sound card (RME babyface). String perturbations were performed manually with a finger or a pick. The sound morphologies observed for the synthesis model are very similar to those observed for the sounds recorded when using the pick. When using the finger to choke the string, the generation of frequency components is very limited and the energy of the modes interacting with the finger is dissipated.

IV. Signal-based synthesis process

In this section, we present the design of the signal model aimed at reproducing the sound morphologies determined in the previous section. The process should accurately replicate the three main phases and particularly the energy transfer to higher frequencies during the interaction phase. One of the challenges here is to design a synthesis process mimicking physical energy dissipation mechanisms.

As a default case, the method should replicate the sound of a stiff string vibrating normally. The method should further be able to imitate the effect of a rigid barrier at a specified location (x_0, y_0) from $t = t_0$.

A. Additive synthesis of the resonant object

To design a process that generates the sound radiated by a stiff string, consider the following additive representation (for a modal model, see e.g. [55]):

$$s(n) = \sum_{i=1}^{N} \underbrace{A_{0(i)} e^{-\alpha_i n k}}_{A_i(n)} \sin(2\pi f_i n k)\,. \tag{7}$$

Here, f_i are the frequencies of the sinusoidal components, in Hz, α_i are the damping coefficients, and $A_{0(i)}$ is the initial amplitude associated with component i, for $i = 1, \ldots, N$. The sampling frequency is fixed at $f_s = 44100$ Hz. The time step k, in s, is defined as $k = 1/f_s$.

The expressions for the frequency of each mode and the damping law are obtained from the physical model defined in Eq. (1), for small values of the parameters σ_0 and σ_1 (see [50] p177-179):

$$f_i = \frac{\gamma i}{2L}\sqrt{1 + \frac{\kappa^2 \pi^2 i^2}{\gamma^2 L^2}} \qquad \alpha_i = \sigma_0 + \sigma_1 \frac{\pi^2 i^2}{L^2}\,. \tag{8}$$

The initial amplitude of each mode $A_{0(i)}$ is measured from the signal generated by the physical model. We synthesize partials up to $f_s/2$.

B. Modeling signal behavior due to collisions

When the rigid barrier interacts with the vibrating string (at $t = t_0$), the signal-based method must generate high frequency components, induce losses, and inharmonic and rough timbral content from the beginning of the interaction. For that purpose, signal energy is redistributed towards high frequency components.

If we consider each tonal component as the signal resulting from the oscillation of a mass/spring/damper system, the energy to be conserved is proportional to the power of the sinusoidal signal. In this case, the energy of the global system including all the elementary oscillators is conserved on the condition that the sum of the powers of the tonal components is not modified during the energy transfer. Thus, the following recurrence equation can be established for the power of each tonal component P_i:

$$P_i(n+1) = \left(P_i(n) + \underbrace{T_i(n)}_{\text{transfer}} \right) \underbrace{e^{-2\alpha_i k}}_{\text{losses}} \tag{9}$$

with $\sum_{i=1}^{N} T_i(n) = 0$.

During the transfer (interaction phase), each mode loses part of its power ($T_{i-}(n)$) which is redistributed simultaneously to all other modes (contribution to $T_{i+}(n)$). The amount of power lost by each mode i is proportional to the amount by which a threshold p_i is exceeded, and a fraction of this power is recovered by the other modes. We can express the transfer term $T_i(n)$ as follows :

$$T_i(n) = \underbrace{-\lambda \left[P_i(n) - p_i\right]_+}_{T_{i-}(n)} + \underbrace{\sum_{j=1}^{N} \lambda \theta_j \left[P_j(n) - p_j\right]_+}_{T_{i+}(n)} \tag{10}$$

The parameter $0 < \lambda < 1$ controls the speed at which the transfer occurs, which has an effect on the duration of the interaction. Another way to affect the speed of redistribution is to perform the redistribution once every N_d samples. The transitions between each redistribution are smooth if λ is at most of the order of $1/N_d$. We set $N_d = 800$ and $\lambda = 1/800$ so that the interaction duration is similar to that observed for the sounds synthesized with the physical model. The coefficients θ_j weight the redistribution from the mode j to the mode i. One must set $\sum_{j=1}^{N} \theta_j = 1$ to ensure the conservation of power during the transfer ($\sum_{i=1}^{N} T_i(n) = 0$).

We define the threshold p_i from an amplitude $\hat{A}_i$ that corresponds to the limit at which the mode interacts with the obstacle:

$$p_i = \frac{\hat{A}_i^2}{2} \tag{11}$$

We define the limit amplitude $\hat{A}_i$ of each mode i by the following equation:

$$\hat{A}_i = \frac{y_0^*}{\sin(\frac{i\pi}{L} x_0^*)}\,. \tag{12}$$

We therefore consider that a mode i interacts with a fictitious obstacle located at (x_0^*, y_0^*) and transmits energy to the other modes as long as its amplitude is greater than the limit amplitude $\hat{A}_i$. We introduce here a fictitious modal deformation of amplitude $\hat{A}_i$ passing through the obstacle (see Fig. 7).

The weighting coefficients θ_i are calculated as follows:

$$\theta_i = \frac{|\sin(\frac{i\pi}{L} x_0^*)|}{\sum_{j=1}^{N} |\sin(\frac{j\pi}{L} x_0^*)|} \tag{13}$$

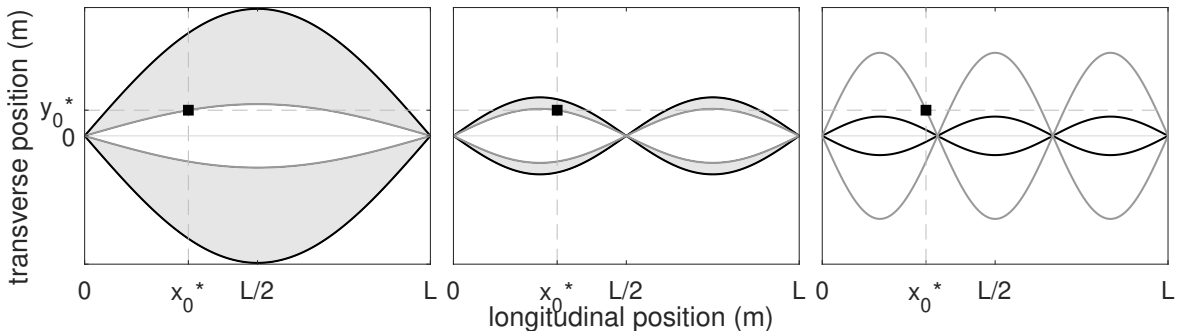

Fig. 6. The black curves represent a modal deformation (for modes $1, 2, 3$, from left to right) at a given time step. The obstacle is displayed at (x_0^*, y_0^*) as a black marker. The grey curves passing through the obstacle represent the modal deformation for the limit amplitude. The left and middle graphs shows grey areas that highlight the difference between black and grey curves when the amplitude of the mode is greater than the limit amplitude (interaction area). The right graph shows the case when the amplitude of the mode is lower than the limit amplitude: in this case, the mode does not distribute energy to other modes.

Here, the process affects the amplitude of the modes according to their modal shape for $x = x_0^*$. Modes with a vibration node at $x = x_0^*$ are not affected by any energy transfer ($\theta_i = 0$). In addition, because $\sum_{i=1}^{N} \theta_i = 1$, the sum of the power redistributed to all modes $\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda\theta_j \left[P_j(n) - p_j\right]_+$ is equal to the total power exceeding the limit $\sum_{i=1}^{N} \lambda \left[P_i(n) - p_i\right]_+$. We define $\Delta P_{tot}(n)$ as the sum of the power redistributed to all modes:

$$\Delta P_{tot}(n) = \sum_{i=1}^{N} \underbrace{\sum_{j=1}^{N} \lambda\theta_j \left[P_j(n) - p_j\right]_+}_{T_{i+}(n)} = \sum_{i=1}^{N} \underbrace{\lambda \left[P_i(n) - p_i\right]_+}_{-T_{i-}(n)} \tag{14}$$

Note that the limit amplitude $\hat{A}_i$ of the modes with a node at x_0^* tends towards infinity and thus the power variation $T_i(n)$ is zero, so they are not affected by the redistribution process. Two scenarios can be identified for the other modes:

- if $A_i(n) > \hat{A}_i$, this mode will be considered as interacting with the rigid barrier. Part of its power will be distributed to the other modes of the string.
- if $A_i(n) < \hat{A}_i$, this mode only receives power from interacting modes. Its amplitude may increase if the distribution process is stronger than the damping process at the mode's frequency.

It is also interesting to note here that it is possible to introduce losses into the redistribution by multiplying the term $T_{i+}(n)$ by a number between 0 and 1 to limit or even eliminate the appearance of frequency components during the interaction phase. We can also choose to weight the redistribution in certain frequency bands (for example, limiting the redistribution towards high frequencies to approach muted sounds) by acting on the coefficients θ_i (ensuring that $\sum_{i=1}^{N} \theta_i \leq 1$ to guarantee the stability of the synthesis process).

In addition to the redistribution process, we generate up to two frequency components associated with each mode during the interaction in order to create a rough and inharmonic signal. We here seek to reproduce the particular pattern described in Sec.III-B for an obstacle located in the middle of the string (see Fig.4).

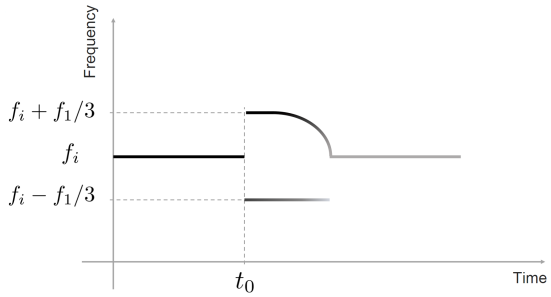

Fig. 7. Schematic time-frequency representation of the two frequency components generated for each mode i of the string during the interaction (the grayscale represents the amplitude of the components).

This reproduction is sufficient to encode the sound event and the location of the rigid barrier. We have used the following equations to generate this part of the signal:

$$s(n) = \sum_{i=1}^{N} \left[B_i(n)\sin(\Phi_{i+}(n)) + C_i(n)B_i(n)\sin(\Phi_{i-}(n))\right] \tag{15}$$

with :

- $\Phi_{i+}(n)$, the phase of the upper tonal component associated with the mode i:

$$\Phi_{i+}(n) = \Phi_{i+}(n-1) + 2\pi(f_i + C_i(n)\frac{f_1}{3})k$$

- $\Phi_{i-}(n)$, the phase of the lower tonal component associated with the mode i:

$$\Phi_{i-}(n) = \Phi_{i-}(n-1) + 2\pi(f_i - \frac{f_1}{3})k$$

- $B_i(n)$, the amplitude of the upper tonal component associated with the mode i:

$$B_i(n) = \sqrt{\frac{2P_i(n)}{1 + C_i(n)^2}}$$

- $0 \leq C_i(n) \leq 1$, a function allowing the continuous transition from the phase where two tonal components per mode are generated ($C_i > 0$) to the phase where

only one tonal component is associated to each mode ($C_i = 0$). C_i becomes greater than 0 if the redistributed power exceeds a threshold value $\hat{P}$ and the rate at which it approaches 1 is driven by the coefficient c_p:

$$C_i(n) = \theta_i \left[1 - \exp\left(-c_p(\Delta P_{tot}(n) - \hat{P})\right)\right]_+ \quad (16)$$

The power $P_i(n)$ associated with each mode i at time step n is distributed in two components. If $\Delta P_{\text{tot}}(n)$ is greater than an arbitrary limiting value $\hat{P}$ (i.e. if there is a strong interaction), the process generates two distinct components for each mode at frequencies $f_i + C_i(n) f_1/3$ and $f_i - f_1/3$. Then, as $\Delta P_{\text{tot}}(n)$ decreases, the component at $f_i - f_1/3$ gradually disappears and the frequency of the other component decreases to the initial frequency of the mode f_i.

For $0 < \Delta P_{\text{tot}}(n) < \hat{P}$, only one component remains for each mode at f_i, but $\Delta P_{\text{tot}}(n)$ is still distributed over all the modes. This corresponds to the effect of sparse collisions on the signal. For $\Delta P_{\text{tot}}(n) = 0$, the process generates one component for each mode, and the redistribution halts.

These variations are driven by the function $C_i(n)$, ranging between 0 and 1. We made the choice to use an exponential function here so that the frequency of the upper component remains stable around $f_i + f_1/3$ at the beginning of the interaction and then rapidly drops to its final value f_i (see Fig. 7). The coefficient c_p characterizes the slope of the frequency drop. Also, $C_i(n)$ is weighted by the distribution coefficient θ_i (the modes with a node at x_0^* are not affected).

Below is an overview of the proposed synthesis method:

from $t = t_0$

0. initialization for $n = \lfloor t_0 \times f_s \rfloor$:
$P_i(n) = A_i(n)^2/2$

1. Calculation of the power transfers for all the modes (Eq. (10)):
$T_i(n) = -\lambda \left[P_i(n) - p_i\right]_+ + \sum_{j=1}^{N} \lambda \theta_j \left[P_j(n) - p_j\right]_+$

2. Calculation of the signal at the nth time step, splitting the mode into two sinusoidal components (Eq. (15)).

3. Calculation of the power of the modes at $n+1$ (Eq. (9)):
$P_i(n+1) = \left(P_i(n) + T_i(n)\right) e^{-2\alpha_i k}$

4. Update of the time step and return to step 1.

The spectrogram of a sound synthesized by this model is shown in Fig. 8 for $x_0^* = L/2$.

Preliminary listening experiments indicate that this signal-based synthesis process is capable of generating a variety of sounds that evoke the perturbation undergone by a string interacting with an obstacle and the parameters (x_0^*,y_0^*) appropriately control the model. The profile of the frequency components during the interaction can be modified through the parameters $\hat{P}$ and c_p.

This signal model offers a simple and controllable way to transfer energy between different tonal components. This phenomenon is an essential feature of the non-linear behaviour of sound sources. It is possible to extend the use of this model

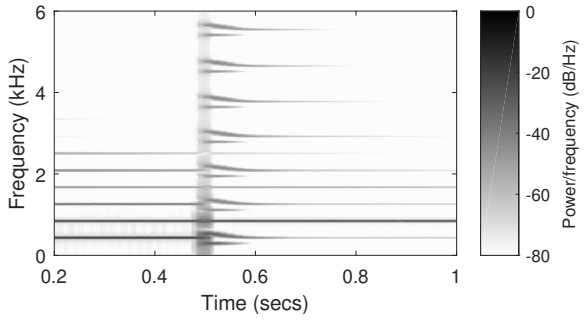

Fig. 8. Spectrogram of a sound synthesized with the signal-based synthesis process for $x_0^* = L/2$; $y_0^* = 0.42 \times |\sin(\pi x_0^*/L)|$; $\hat{P} = 340$; $c_p = 6 \times 10^{-4}$. Note that the frequency components corresponding to the modes 2 and 4 are not modified by the process during the interaction. 2048-point Kaiser windows ($\beta = 5$) are used, with 1800 samples of overlap.

to other configurations. For instance, it is possible to *choke* the string by removing positive contributions from transfers ($T_{i+} = 0$). Also, we can extend the model to two-dimensional objects and generate snare drum sounds [17]. The definition of random weighting coefficients θ_i for the redistribution allows an approach to the emulation of the sound of the tanpura [14]. More generally, the addition of random amplitude and frequency modulations during interaction allows more realistic sounds to be generated (sound examples are available at [46]).

V. Perceptual evaluation

In this section, we describe several perceptual evaluations of sounds resulting from the collisions between a string and an obstacle. In the first two experiments, we seek to better understand our ability to retrieve the position of the obstacle from the perception of the sound resulting from this phenomenon, and further evaluate the abilities of the signal model to encode the signal morphologies that allow the phenomenon to be identified and localized with respect to the control parameters (x_0^*, y_0^*, $\hat{P}$, c_p). Also, we seek to propose a semantic description to describe the perturbation undergone by the string. The aim is to work towards an intuitive control of the sound synthesis method presented in this paper. In a last experiment, we seek to evaluate the quality of the synthesized sounds in terms of evocation and realism by comparing them to recorded sounds.

A. Experiment 1: evaluation of perceived location of the obstacle

In this experiment, we compare sounds generated by the signal and physical models through a listening test. We test the hypotheses and choices made when designing the signal model (e.g. the addition of tonal components for roughness, frequency variation for inharmonicity, and thresholds and weighting coefficients for selective redistribution). In particular, we evaluate how the signal model under different configurations allows the recognition of the position of an obstacle. The sounds generated by the physical model are taken as a ground truth for the signal model.

1) Experimental design: The experiment is a full factorial design. We study the influence of three factors on the perceived location of the obstacle. The j^{th} level of the factor ζ is indicated by ζ_j. The three factors ($\mathcal{M}$, $\mathcal{X}$, $\mathcal{Y}$) are described below:

• $\mathcal{M}$ (3 levels) corresponds to the model used to synthesize the sound: the physical model and the signal model with 2 different profile functions $C_i(n)$, chosen to have different evolutions of the rough and inharmonic signal.

• $\mathcal{X}$ (3 levels) corresponds to the longitudinal position of the obstacle.

• $\mathcal{Y}$ (9 levels) corresponds to the transverse position of the obstacle. The 9 level values were chosen from an informal calibration conducted by the authors (through a listening test) aiming at a perceptually linear transition along the whole range of y_0 and ${y_0}^*$ for both the physical and the signal models. It thus considers the fact that the sound rapidly changes for small variations of the position close to the extremities.

TABLE II
LEVELS FOR FACTORS $\mathcal{M}$, $\mathcal{X}$ AND $\mathcal{Y}$.

Factor $\mathcal{M}$

Level	Description
$\mathcal{M}_1$	physical model
$\mathcal{M}_2$	signal model: $c_p = 6 \times 10^{-4}$, $\Delta P_{lim} = 340$
$\mathcal{M}_3$	signal model: $c_p = 1 \times 10^{-4}$, $\Delta P_{lim} = 4000$

Factor $\mathcal{X}$

Level	x_0 for $\mathcal{M}_1$, x_0^* for $\mathcal{M}_2$, $\mathcal{M}_3$
$\mathcal{X}_1$	$L/2$
$\mathcal{X}_2$	$L/3$
$\mathcal{X}_3$	$5L/12$

Factor $\mathcal{Y}$

Level	$\frac{y_0}{U(x_0,t_0)}$ for $\mathcal{M}_1$	$\frac{y_0^*}{\lvert\sin(\pi x_0^*/L)\rvert}$ for $\mathcal{M}_2$, $\mathcal{M}_3$
$\mathcal{Y}_1$	0.0065	0.005
$\mathcal{Y}_2$	0.013	0.01
$\mathcal{Y}_3$	0.13	0.1
$\mathcal{Y}_4$	0.315	0.26
$\mathcal{Y}_5$	0.49	0.58
$\mathcal{Y}_6$	0.675	0.74
$\mathcal{Y}_7$	0.875	0.9
$\mathcal{Y}_8$	0.921	0.99
$\mathcal{Y}_9$	0.985	0.995

The function $C_i(n)$ is defined differently for $\mathcal{M}_2$ and $\mathcal{M}_3$ (see Eq. (16)). Fig. 9 shows the frequency gap at the beginning of the interaction Δf for the different levels of $\mathcal{Y}$ for the two different profile functions (corresponding to $\mathcal{M}_2$ and $\mathcal{M}_3$). Here, we generate a rough and inharmonic signal during the interaction when $\Delta f > 0$ (levels $\mathcal{Y}_{j\leq 8}$ for $\mathcal{M}_2$ and $\mathcal{Y}_{j\leq 7}$ for $\mathcal{M}_3$). Also, the frequency of the upper component tends to increase more smoothly and the rough signal duration is shorter for $\mathcal{M}_3$ than for $\mathcal{M}_2$.

In addition to these synthesized sounds, two baseline sounds simulating the vibrating string without an obstacle were synthesized: one from the physical model and the other using the signal model. In summary, a total of 83 sounds

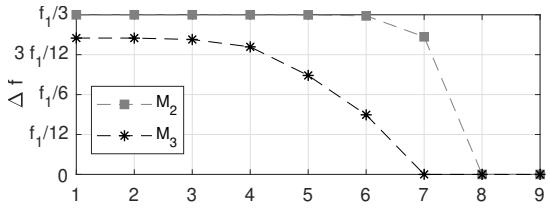

Fig. 9. The frequency gap $\Delta f = C_i(n) f_1/3$ of the affected modes at the beginning of the interaction for the signal model for different levels of $\mathcal{Y}$.

($= 3 \times 3 \times 9 + 2$) were generated for the listening test.

2) Participants: 22 participants (nine female) took part in the experiment. Six of them regularly practiced a string instrument at an amateur level and ten worked in audio-related fields (as a researcher or technician). The age range was from 22 to 67 years old (with a mean of 32). None had hearing problems (as tested using a calibrated audiometer).

3) Procedure: Testing was performed in a quiet room, and participants used closed headphones (Sennheiser HD280pro). The 83 sounds were presented to the participants in a random order. For each sound, participants were asked to retrieve the position of the obstacle and to indicate it visually on a screen displaying the string and the obstacle. The visual indication was performed using two sliders for the longitudinal and transverse positions of the obstacle, noted R_x and R_y respectively. R_y was evaluated on a scale from 0 to 100, where 0 was for the equilibrium position of the string and 100 for the maximum vibration amplitude of the string at x_0. R_x was evaluated on a scale from 0 to $L/2$. Three markers permitted the localisation of specific positions along the x-axis ($L/2$, $L/3$, $L/4$). The participants had the possibility to tick a box labeled "no idea" if they were not able to locate the obstacle along the x-axis. Also, the participants were asked to give a short semantic description of the action evoked by each sound (with a verb and/or adverb). This semantic report was optional. The total duration of the listening tests was between 35 and 75 minutes per participant.

4) Results:

a) Perception of the transverse position R_y: We conducted a repeated measures Analysis of Variance (ANOVA) on the R_y values including $\mathcal{M}$ and $\mathcal{Y}$ as factors. The results are displayed in Fig. 10. We observe a strictly increasing curve of the mean value for the level of the factor $\mathcal{Y}$ for all the models ($\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$).

In the comparison of results between the models, scores did not differ significantly for the first four and last two levels of $\mathcal{Y}$ (there was no significant variation between results for the models for $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4, \mathcal{Y}_8, \mathcal{Y}_9$, $p > 0.937$). In particular, one can note that the perceptual distance between the models is very low (because the model correctly transcribes the sound event) for the first levels of $\mathcal{Y}$ (see $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4$, Fig.10).

It is interesting to note that the extreme values for R_y can be reached with the signal model. Thus, encoding through the signal-based method allows the recognition of the transverse position over the whole range of $\mathcal{Y}$. We found significant differences between results for the models for the middle

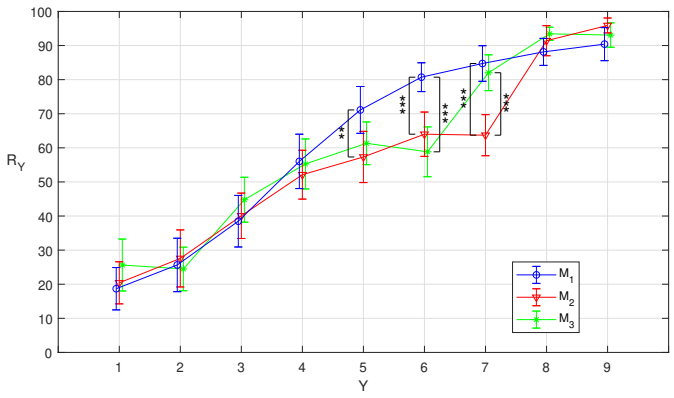

Fig. 10. Combined effects of factors $\mathcal{Y}$ and $\mathcal{M}$ on the measure R_y. Least Square Means. Current effect: $F(18, 378) = 7.7548$, $p < 0.0001$. Vertical bars denote +/- standard errors. Tukey's HSD Post Hoc test : $** = p < 0.01$, $*** = p < 0.001$.

levels of $\mathcal{Y}$ ($\mathcal{Y}_5$-$\mathcal{Y}_7$). Considering the curve for the physical model ($\mathcal{M}_1$) as the reference, these deviations can be directly related to the different profiles chosen for the signal models ($\mathcal{M}_2$, $\mathcal{M}_3$). Indeed, one can observe that R_y falls between levels 8 and 7 (resp. 7 and 6) for $\mathcal{M}_2$ (resp. $\mathcal{M}_3$), directly corresponding to the levels of appearance of the rough and inharmonic part of the signal during the interaction for these respective models (see Fig. 9).

Thus, this result shows that the rough and inharmonic aspect of the signal has a strong impact on the perceived transverse location R_y.

Finally, the semantic labels proposed by the participants can be summarized as follows:

- for high levels of $\mathcal{Y}$: "*effleurer, frôler*" (French verbs for "touch softly"), "*légèrement, faiblement*" (French adverbs for "slightly") and "*doucement*" (French for "softly")
- for low levels of $\mathcal{Y}$: "*étouffer*" (French for "choke"), "*appuyer*" (French for "push") and "*fortement*" (French for "strongly")

These word choices show that the transverse position is often associated with the evoked intensity (or force) of interaction. We touch slightly (resp. strongly) the string for high (resp. low) levels of $\mathcal{Y}$. Thus, the rough and inharmonic aspect of the sound evokes a strong interaction.

b) Perception of the longitudinal position R_x: To evaluate the ability of participants to identify the longitudinal position of the obstacle, the distance to the target was calculated ($d = |R_x - \mathcal{X}_j|$) for any participant who did not tick the "no idea" box. If $d \leq L/30$, the R_x score was considered as correct (within a margin of error of one step to the left and right of the marker). In addition, we examined the scores at particular locations along the string: $\mathcal{X}_1 = L/2$, $\mathcal{X}_2 = L/3$, and $\mathcal{X}_3 = 5L/12$. Indeed, since the variation of sounds according to $\mathcal{X}$ follows a specific logic (see Section III-B), some sounds could be more easily identified at particular locations, as when playing a natural harmonic on a guitar. Thus, the longitudinal position could be more easily recognizable for low levels of $\mathcal{Y}$ particularly at $L/2$, $L/3$, $2L/3$, $L/4$... For other values of x_0, the sound is entirely dissipated during the interaction, leaving no possibility of retrieving the longitudinal position with the remaining harmonics. For instance, there is no sound after the interaction with an obstacle located at $x_0 = 5L/12$ and $y_0 = 0$ as the 12^{th} harmonic of the string has already disappeared at $t = t_0$. Also, it is increasingly difficult to distinguish the strongest harmonics after the interaction when y_0 increases.

First, we evaluated the results for the physical model ($\mathcal{M}_1$). We found that the target identification scores were poor and that the longitudinal position of the obstacle was generally difficult to recognize. The participants had no idea in 38.72% of the cases ("no idea" box ticked) and the values were close to the target in 8.08% of the cases only. When focusing on the 3 first levels of y_0, the participants had no idea in 28.79% of the cases, and they were close to the target in resp. 9.09%, 16.67%, and 6.06% of the cases for resp. $x_0 = L/2$, $x_0 = L/3$, and $x_0 = 5L/12$.

Interestingly, if we focus on participants who play a string instrument regularly (i.e. 6 participants), we observe a significant improvement in the results. When focusing on the 3 first levels of y_0, these participants had no idea in 16.67% of the cases, and they were close to the target in resp. 22.22%, 50%, and 5.56% of the cases for resp. $x_0 = L/2$, $x_0 = L/3$, and $x_0 = 5L/12$. One can also note that these participants reported $R_x = L/4$ for $x_0 = L/2$ in 38.89% of cases. This confusion is due to the fact that the remaining harmonics after the interaction for $x_0 = L/4$ have a double frequency of the remaining harmonics for $x_0 = L/2$. After the interaction, the two sounds are separated by an octave, meaning that they are perceptually close (due to the octave-equivalence effect).

Since these results were obtained on scores from 6 participants who play a string instrument regularly and a reduced number of stimuli in the experiment, a more in-depth study has been conducted.

B. Experiment 2: Effect of expertise and training on the perception of the longitudinal position

We conduct a similar experiment, with participants having a background in acoustics and music, for specific obstacle positions in order to investigate more precisely if the recognition of the longitudinal position is perceived by this type of person and if specific training can significantly improve this recognition. Also, we attempt to verify whether the signal model correctly transcribes the phenomenon.

1) Experimental design: We study the influence of three factors described below on the perceived longitudinal position of the obstacle:

- $\mathcal{T}$ (2 levels) corresponds to the training of the participant ($\mathcal{T}_1$: before training; $\mathcal{T}_2$: after training).

- $\mathcal{M}$ (2 levels) corresponds to the model used to synthesize the sound ($\mathcal{M}_1$: physical model; $\mathcal{M}_2$: signal model).

- $\mathcal{X}$ (8 levels) corresponds to the longitudinal position of the obstacle ($\mathcal{X}_j$: $x_0 = L/(j+1)$).

The sounds are generated for an obstacle located at $y_0 = 0$. Each sound is repeated twice for a total of 64 stimuli.

2) Participants: 11 participants (4 female) took part in the experiment, all part of a master's degree in acoustics and musicology. They have a scientific background allowing them to understand the modal representations of string vibrations and a musical background allowing them to recognize different musical intervals. The age range was from 20 to 24 years old (with a mean of 22). None of them reported any hearing problems.

3) Procedure: Testing was performed in a quiet room and participants used closed headphones. The experiment was organized in three distinct phases:

- ($\mathcal{T}_1$) First, 32 stimuli were presented to the participants in a random order (16 conditions $\mathcal{T}_1\mathcal{M}_i\mathcal{X}_j \times 2$ repetitions). For each sound, participants were asked to retrieve the longitudinal position of the obstacle by moving a slider ranging from 0 to $L/2$. Seven equidistant markers were positioned on the slider (0, $L/12$, $L/6$, $L/4$, $L/3$, $5L/12$, $L/2$).

- Then the participants were informed about the physical phenomenon involved in the interaction process and how they could retrieve the position from the difference in pitch between the beginning and the end of the sound.

- ($\mathcal{T}_2$) The first phase was repeated, post training.

4) Results: As for the previous experiment, we computed a percentage of correct answers from the distance to the target for each condition. We used binomial tests to evaluate the statistical significance of deviations between the conditions and a theoretical random distribution computed from random answers. The probability of having a correct answer by answering randomly is different for each level of the longitudinal position factor $\mathcal{X}$ and is equal to the number of slider positions corresponding to a correct answer divided by the total number of slider positions.

The average correct response rate across all conditions is low (23.72%) but still significantly higher than for random responses (p<0.00001). The correct response rate is significantly higher than for random for all conditions except for $\mathcal{X}_6$ (p=0.1026), $\mathcal{X}_7$ (p=0.0762) and $\mathcal{X}_8$ (p=0.5340). For these specific locations, the intensity of the harmonics remaining after the interaction is low and the interval is difficult to recognise ($L/7$, $L/8$, $L/9$). The correct response rate excluding these conditions is 31.13%. It is therefore a difficult but not impossible task for $\mathcal{X}_i$ with i<6.

If we compare the different conditions with a binomial test, we observe that the correct response rate is significantly higher after the training phase (32.39%>15.06%, p<0.00001). We can conclude that training has a significant effect and that the task can be learned quickly. If we observe only the conditions $\mathcal{X}_i$ with i<6, the rate of correct answers after training is 42.27%. The task in this case becomes quite feasible, even if the correct response rate remains below 50%.

C. Experiment 3: Evaluation of the quality of the sounds generated by the signal model

The previous experiments have informed us about the ability of our model to evoke an obstacle colliding with a string at a particular position. However, they provide little evidence to show whether the sounds produced are plausible or realistic. The aim of this final experiment is to find out how closely (and under what conditions) sounds generated by the signal model can approach real sounds. In particular, we noticed that the sounds generated by the signal model could sound synthetic despite the correct transcription of the position information and we wish to evaluate whether the use of random processes could improve the plausibility of the generated sounds.

1) Experimental design: We used the MUSHRA methodology to compare different versions of the signal model with recorded sounds.

We used 6 different recorded sounds from the corpus (D-string of an electric guitar), for 3 different obstacle positions ($\mathcal{X}_1$: L/2, $\mathcal{X}_2$: L/3, $\mathcal{X}_3$: L/4) and 2 different obstacles ($\mathcal{O}_1$: finger, $\mathcal{O}_2$: pick).

We calibrated the signal model with two different sets of parameters to reproduce the recorded sounds corresponding to an interaction between a string and a finger or a pick. For each recorded sound, we compared the 6 following conditions:

- $\mathcal{C}_1$ Randomized phase of each tonal component during the interaction phase to generate a noisy-like signal (low anchor),
- $\mathcal{C}_2$ Signal model without additional tonal component during the interaction phase for strong interactions,
- $\mathcal{C}_3$ Signal model as presented in Sec.IV (2 tonal components per mode),
- $\mathcal{C}_4$ Signal model without additional tonal component during the interaction phase, random frequency modulation and random modification of non-zero weighting coefficients at regular time intervals,
- $\mathcal{C}_5$ Signal model with 2 tonal components per mode, random frequency modulation and random modification of non-zero weighting coefficients at regular time intervals,
- $\mathcal{C}_6$ hidden reference (recorded sound).

We added a background corresponding to an empty recording to the synthesized sounds in order to ensure that the noise associated with the recording would not be a factor in the evaluation of the stimuli. The test interface was designed with the webMUSHRA framework [56].

2) Participants: 18 participants (5 female) took part in the experiment. The age range was from 23 to 72 years old (with a mean of 38). None of them reported any hearing problems.

3) Procedure: After listening alternately to the reference and the different conditions, participants had to evaluate the similarity between a reference sound and 6 stimuli corresponding to the 6 conditions mentioned above. Participants were asked to rate the similarity, noted R, using sliders ranging from 0 to 100 with the following indications: 0-20 “Bad" ; 20-40 “Poor" ; 40-60 “Fair" ; 60-80 “Good" ; 80-100 “Excellent".

4) Results: We conducted a repeated measures Analysis of Variance on the response values R including $\mathcal{C}$, $\mathcal{O}$ and $\mathcal{X}$ as factors. The results for the low anchor and the reference sound are consistent (see Fig.11, $\mathcal{C}_1$ and Ref). We focus our statistical analysis on the other 4 conditions ($\mathcal{C}_2$, $\mathcal{C}_3$, $\mathcal{C}_4$ and $\mathcal{C}_5$).

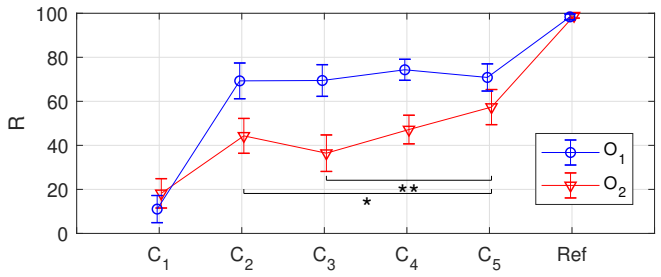

Fig. 11. Combined effects of factors $\mathcal{C}$ and $\mathcal{O}$ on the measure R. Least Square Means. Current effect: $F(5, 85) = 24.968$, $p < 0.0001$. Vertical bars denote +/- standard errors. Tukey's HSD Post Hoc test : $* = p < 0.05$, $** = p < 0.01$.

On average, we observe scores around 60 and little variation between conditions. Interesting results can be observed in the case of the interaction between the factors $\mathcal{O}$ and $\mathcal{C}$. The evolution of the quality of sound reproduction by the signal model for the different conditions is significantly affected by the nature of the object used for the collisions with the string. In the case of the finger ($\mathcal{O}_1$), the reproduction is rated equally good for all conditions. The redistribution phenomenon during the interaction between the string and the finger is subtle and the appearance of a frequency component is limited. We can conclude that a basic model gives good results and that it is not necessary to add tonal components to increase the roughness or random modulations to improve the natural aspect of the sound for this kind of interaction.

Conversely, the quality of the reproduction of sounds recorded with the pick ($\mathcal{O}_2$) varies significantly according to the conditions (see Fig.11). In this case, the model corresponding to the condition $\mathcal{C}_5$ (between "fair" and "good") is significantly better than the models presenting no random modulations ($\mathcal{C}_2$ and $\mathcal{C}_3$ between "poor" and "fair"). The interaction between the string and the pick is strong and leads to a rich and complex sound. For this type of interaction, the models corresponding to conditions $\mathcal{C}_2$ and $\mathcal{C}_3$ generate unnatural sounds even though it is clear that the transcription of the sound event was correct in the previous experiments. In particular, we can observe that the appearance of a second tonal component seems to degrade the sound quality if it is not accompanied by random modulations.

D. Discussion: towards a perceptual control

The results of the perceptual evaluations are of great interest for the design of perceptually controlled sound synthesis processes. In practice, a perceptual control can be designed by defining semantic labels "soft"/"strong" (as proposed by the participants, see Sec.V-A4a) corresponding to the evocation of the nature of the interaction and by mapping them to the previous signal patterns and to the control parameters reflecting the physical location of the rigid barrier (x_0^*, y_0^*). Note that the synthesis process allows the generation of different profiles with the same global energy distribution during the interaction (profile function $C_i(n)$ defined in Eq. (16), controlled by c_p and $\hat{P}$). For this purpose, results obtained from the listening test (see Fig. 10) allowed the calibration of the perceived transversal position with respect to the physical position, notably large perceptual variations for low values of y_0. It was observed that the interaction was described as 'strong' when the signal was inharmonic and rough ($C_i(n) > 0$). Conversely, when energy is distributed to high-frequency components, but with no frequency shift of the modes, the perceived transverse location was close to its maximum value and the interaction was described as 'soft'. In contrast, we found that the longitudinal position was difficult to retrieve by listeners and required expertise (i.e. a trained ear), since it is based on the identification of the harmonics remaining after the interaction for low levels of $\mathcal{Y}$ and for specific values of x_0 ($L/2$, $L/3$ ...). The signal model yields a consistent transcription of these acoustic cues, allowing trained users to recognise the longitudinal position of the obstacle.

Further, the proposed signal-based synthesis process allows a low-parameter alternative to physical constraints in terms of control while retaining the perceptual characteristic of interacting with an obstacle. In practice, the sound morphologies reproduced by the signal model and characterizing the non linear interaction could be applied to different virtual objects or sound textures by modifying the initial signal (defined in section III.A). Hence, new tools could be developed for sound designers, giving them an alternative to databases of recorded sounds for various applications such as video games. This would lead to real-time event-driven synthesis of sounds in virtual or augmented reality environments. Also, the possibility emerges of generating entirely new sounds that carry information contained in the sound invariants.

However, there is still work to be done to develop ergonomic tools that can be used in a mainstream sound design context. In particular, it is necessary to explore the sound space that can be covered by the model and to distinguish interesting cases from problematic ones (redundant and/or critical). Also the mapping of the synthesis parameters to a smaller, more perceptually meaningful set may require some additional work.

VI. Conclusion & perspectives

In this paper, we have indentified the sound morphologies responsible for the recognition of nonlinear interactions such as a string colliding with a rigid barrier. We have proposed a signal-based sound synthesis process allowing the retranscription of the disturbance undergone by a vibrating object when it collides with a rigid obstacle. We have determined typical signal behaviors that carry the perceptual information of the nature of the interaction with respect to the location of the rigid barrier by investigating a sound data bank synthesized with a physical model. We have observed a generation of frequency components during the interaction due to harmonic distortion and mode coupling and a frequency shift of some modes for strong interactions which, added to the generation of frequency components, causes a rough and inharmonic signal. We then designed a signal-based sound synthesis process that reproduces these patterns and evaluated subjects' perceptual ability to locate the interaction point from listening to the sounds generated by both physical and signal models. Results from the physical model informed us about our ability to locate the

obstacle, and we validated the transcription of the phenomenon with the signal model by comparing the results obtained with both models. The comparison of the sounds generated by the signal model with recorded sounds allowed us to determine that the quality of the reproduction was considered "good". The results of the listening tests showed that we are able to precisely recognize the transverse location which we associate with the notion of "intensity" of interaction. Conversely, the recognition of the longitudinal location requires a trained ear.

This type of interaction brings together various identifiable phenomena observed in musical instruments (tanpura, fret buzz, rattling element) and may be the subject of future work on interactions with various sound-producing objects. We also aim to investigate the perceptual control of structural features of the barrier. Indeed, it may be of interest to study how we perceive the size, shape and material of the obstacle through the radiated sound of the string. Finally, as discussed earlier in section V.D, we also intend to explore how these sound invariants could be applied to other types of objects (e.g. membranes, etc.), or to other types of sound textures in the context of the creation of *new* sounds.

REFERENCES

[1] C. Zheng and D. L. James, "Toward high-quality modal contact sound," in *ACM SIGGRAPH 2011 papers*, Aug. 2011, pp. 1–12.

[2] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, Dec. 2001, pp. 61–66.

[3] A. Akay, "Acoustics of friction," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1525–1548, April 2002.

[4] F. Avanzini, S. Serafin, and D. Rocchesso, "Interactive simulation of rigid body interaction with friction-induced sound generation," *IEEE Trans. Speech Audio Proces.*, vol. 13, no. 5, pp. 1073–1081, Sept. 2005.

[5] E. Thoret, M. Aramaki, C. Gondre, R. Kronland-Martinet, and S. Ystad, "Controlling a non linear friction model for evocative sound synthesis applications," in *Proc. Int. Conf. on Digital Audio Effects*, Sept. 2013, pp. 1–7.

[6] S. Conan, O. Derrien, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "A synthesis model with intuitive control capabilities for rolling sounds," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 22, no. 8, pp. 1260–1273, Aug. 2014.

[7] S. Serafin and D. Young, "Toward a generalized friction controller: from the bowed string to unusual musical instruments," in *Proc. Conf. New Interfaces Mus. Expr.*, June 2004, pp. 108–111.

[8] M. McIntyre and J. Woodhouse, "On the fundamentals of bowed-string dynamics," *Acta Acustica united with Acustica*, vol. 43, no. 2, pp. 93–108, 1979.

[9] S. Serafin, C. Wilkerson, and J. Smith, "Modeling bowl resonators using circular waveguide networks," in *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, Sept. 2002, pp. 117–121.

[10] O. Inácio, L. L. Henrique, and J. Antunes, "The dynamics of tibetan singing bowls," *Acta Acustica United with Acustica*, vol. 92, no. 4, pp. 637–653, 2006.

[11] J. Laroche and J.-L. Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *IEEE Trans. Speech Audio Proces.*, vol. 2, no. 2, pp. 329–344, 1994.

[12] D. Russell and T. Rossing, "Testing the nonlinearity of piano hammers using residual shock spectra," *Acta Acustica United with Acustica*, vol. 84, no. 5, pp. 967–975, 1998.

[13] S. Bilbao and A. Torin, "Numerical modeling and sound synthesis for articulated string/fretboard interactions," *Journal of the Audio Engineering Society*, vol. 63, no. 5, pp. 336–347, 2015.

[14] M. van Walstijn, J. Bridges, and S. Mehes, "A real-time synthesis oriented tanpura model," in *Proc. Int. Conf. Digital Audio Effects*, 2016, pp. 175–182.

[15] E. Rank and G. Kubin, "A waveguide model for slapbass synthesis," in *Proc. Int. Comf. Acoust. Speech Sig. Proces.*, May 1997, pp. 443–446.

[16] S. Bilbao and J. Fitch, "Prepared piano sound synthesis," in *Proc. of the 9th Int. Conference on Digital Audio Effects*, Sept 2006, pp. 77–82.

[17] S. Bilbao, "Time domain simulation and sound synthesis for the snare drum," *J. Acoust. Soc. Am.*, vol. 131, no. 1, pp. 914–925, Jan. 2012.

[18] S. Willemsen, N. S. Andersson, S. Serafin, and S. Bilbao, "Real-time control of large-scale modular physical models using the sensel morph," in *16th Sound and music computing conference*. Sound and Music Computing Network, May 2019, pp. 151–158.

[19] "Physical audio website, real-time physical modelling synthesis audio plug-ins," https://physicalaudio.co.uk/, accessed: 2020-12-17.

[20] C. Issanchou, S. Bilbao, J.-L. Le Carrou, C. Touzé, and O. Doaré, "A modal-based approach to the nonlinear vibration of strings against a unilateral obstacle: Simulations and experiments in the pointwise case," *J. Sound Vib.*, vol. 393, pp. 229–251, Feb 2017.

[21] C. Issanchou, J.-L. Le Carrou, C. Touzé, B. Fabre, and O. Doaré, "String/frets contacts in the electric bass sound: Simulations and experiments," *Applied Acoustics*, vol. 129, pp. 217–228, 2018.

[22] C. Issanchou, J.-L. Le Carrou, S. Bilbao, C. Touzé, and O. Doaré, "A modal approach to the numerical simulation of a string vibrating against an obstacle: Applications to sound synthesis," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Sept. 2016, pp. 5–9.

[23] J. Bridges and M. Van Walstijn, "Modal based tanpura simulation: combining tension modulation and distributed bridge interaction," in *Proc. Int. Conf. Digital Audio Effects*, Sept. 2017, pp. 299–306.

[24] M. Ducceschi, "A numerical scheme for various nonlinear forces, including collisions, which does not require an iterative root finder," in *Proc. Int. Conf. Digital Audio Effects*, Sept. 2017, pp. 80–86.

[25] M. Ducceschi and S. Bilbao, "A physical model for the prepared piano," *Proc. 26th Int. Cong. of Sound Vib., Montreal, Canada*, July 2019.

[26] N. Castagné and C. Cadoz, "Genesis: a friendly musician-oriented environment for mass-interaction physical modeling," in *Proc. Int. Comp. Music Conf.*, Sept. 2002, pp. 330–337.

[27] J. Villeneuve and J. Leonard, "Mass-interaction physical models for sound and multi-sensory creation: Starting anew," *Sound and Music Computing Conference*, May 2019.

[28] L. Pruvost, B. Scherrer, M. Aramaki, S. Ystad, and R. Kronland-Martinet, "Perception-based interactive sound synthesis of morphing solids' interactions," in *SIGGRAPH Asia 2015 Technical Briefs*. ACM, Aug. 2015, pp. 1–4.

[29] C. Gan, D. Huang, P. Chen, J. B. Tenenbaum, and A. Torralba, "Foley music: Learning to generate music from videos," in *European Conference on Computer Vision*. Springer, 2020, pp. 758–775.

[30] C. Gan, J. Schwartz, S. Alter, D. Mrowca, M. Schrimpf, J. Traer, J. D. Freitas, J. Kubilius, A. Bhandwaldar, N. Haber, M. Sano, K. Kim, E. Wang, M. Lingelbach, A. Curtis, K. T. Feigelis, D. Bear, D. Gutfreund, D. D. Cox, A. Torralba, J. J. DiCarlo, J. B. Tenenbaum, J. Mcdermott, and D. L. Yamins, "ThreeDWorld: A platform for interactive multi-modal physical simulation," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021. [Online]. Available: https://openreview.net/forum?id=db1InWAwW2T

[31] C. Gan, Y. Gu, S. Zhou, J. Schwartz, S. Alter, J. Traer, D. Gutfreund, J. B. Tenenbaum, J. H. McDermott, and A. Torralba, "Finding fallen objects via asynchronous audio-visual integration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 523–10 533.

[32] K. Franinovic and S. Serafin, *Sonic interaction design*. MIT Press, 2013.

[33] P. Susini, O. Houix, and N. Misdariis, "Sound design: an applied, experimental framework to study the perception of everyday sounds," *The New Soundtrack*, vol. 4, no. 2, pp. 103–121, 2014.

[34] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson, and W. T. Freeman, "Visually indicated sounds," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2405–2413.

[35] Y. Zhou, Z. Wang, C. Fang, T. Bui, and T. L. Berg, "Visual to sound: Generating natural sound for videos in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3550–3558.

[36] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[37] S. McAdams and E. Bigand, "Thinking in sound: The cognitive psychology of human audition." in *Based on the fourth workshop in the Tutorial Workshop series organized by the Hearing Group of the French Acoustical Society.* Clarendon Press/Oxford University Press, 1993.

[38] J. J. Gibson, "The ecological approach to visual perception," 1979.
[39] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, "Controlling the perceived material in an impact sound synthesizer," *IEEE Trans. Audio, Speech, Language Proces.*, vol. 19, no. 2, pp. 301–314, 2011.
[40] 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.
[41] S. Conan, E. Thoret, M. Aramaki, O. Derrien, C. Gondre, S. Ystad, and R. Kronland-Martinet, "An intuitive synthesizer of continuous-interaction sounds: Rubbing, scratching, and rolling," *Comp. Music J.*, vol. 38, no. 4, pp. 24–37, 2014.
[42] J. Traer, M. Cusimano, and J. H. McDermott, "A perceptually inspired generative model of rigid-body contact sounds," in *Digital Audio Effects (DAFx)*, vol. 1, no. 2, 2019, p. 3.
[43] V. Agarwal, M. Cusimano, J. Traer, and J. McDermott, "Object-based synthesis of scraping and rolling sounds based on non-linear physical constraints," in *2021 24th International Conference on Digital Audio Effects (DAFx)*. IEEE, 2021, pp. 136–143.
[44] S. Bilbao, A. Torin, and V. Chatziioannou, "Numerical modeling of collisions in musical instruments," *Acta Acustica united with Acustica*, vol. 101, no. 1, pp. 155–173, May 2015.
[45] M. Ducceschi, "A numerical scheme for various nonlinear forces, including collisions, which does not require an iterative root finder," in *Proc. Int. Conf. Digital Audio Effects*, Sept 2017.
[46] "Sound examples and stimuli," https://www.prism.cnrs.fr/publications-media/IEEEPoirot/, accessed: 2021-04-12.
[47] L. Hiller and P. Ruiz, "Synthesizing sounds by solving the wave equation for vibrating objects," *J. Audio Eng. Soc.*, vol. 19, pp. 463–470, 1971.
[48] A. Chaigne and A. Askenfelt, "Numerical simulations of piano strings. i. a physical model for a struck string using finite difference methods," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1112–1118, 1994.
[49] J. Bensa, S. Bilbao, R. Kronland-Martinet, and J. O. Smith III, "The simulation of piano string vibration: From physical models to finite difference schemes and digital waveguides," *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 1095–1107, 2003.
[50] S. Bilbao, *Numerical sound synthesis: finite difference schemes and simulation in musical acoustics*. John Wiley & Sons, 2009.
[51] S. Poirot, S. Bilbao, M. Aramaki, and R. Kronland-Martinet, "Sound morphologies due to non-linear interactions: Towards a perceptual control of environmental sound synthesis processes," in *Proc. Int. Conf. Digital Audio Effects*, Sept. 2018.
[52] S. Papetti, F. Avanzini, and D. Rocchesso, "Numerical methods for a nonlinear impact model: a comparative study with closed-form corrections," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2146–2158, 2011.
[53] D. Kartofelev, "Kinematics of ideal string vibration against a rigid obstacle," in *Proc. Int. Comf. Digital Audio Effects*, Edinburgh, UK, Sept. 2017, pp. 40–47.
[54] P. N. VASSILAKIS and R. A. KENDALL, "Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications," in *Proceedings of SPIE, the International Society for Optical Engineering Vol.7527*. Society of Photo-Optical Instrumentation Engineers, Feb. 2010.
[55] K. Van Den Doel, P. G. Kry, and D. K. Pai, "Foleyautomatic: physically-based sound effects for interactive simulation and animation," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, Aug. 2001, pp. 537–544.
[56] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, "webmushra—a comprehensive framework for web-based listening tests," *Journal of Open Research Software*, vol. 6, no. 1, 2018.

Samuel Poirot received M.S. degrees in mechanics (from I.N.S.A., Lyon, France ; A.M.U., Marseille, France and E.N.S., Cachan, France) and holds the agrégation to teach engineering sciences. He is currently a PhD student in acoustics in the PRISM laboratory under the direction of Stefan Bilbao and Richard Kronland-Martinet.

Stefan Bilbao (B.A. Physics, Harvard, 1992, MSc., PhD Electrical Engineering, Stanford, 1996 and 2001 respectively) is currently Professor of Acoustics and Audio Signal Processing in the Acoustics and Audio Group at the University of Edinburgh, and previously held positions at the Sonic Arts Research Centre, at the Queen's University Belfast, and the Stanford Space Telecommunications and Radioscience Laboratory. He is an Associate Editor of the IEEE/ACM Transactions on Audio Speech and Language Processing. He was born in Montreal, Quebec, Canada.

Mitsuko Aramaki received the Ph.D. degree from 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). Since 2017, she is head of the "Perception Engineering" team at the laboratory PRISM "Perception, Representations, Image, Sound, Music". Her research mainly focuses on sound modeling, perceptual and cognitive aspects of timbre, and multimodal interactions in the context of virtual/augmented reality.

Sølvi Ystad received her degree as a civil engineer in electronics from NTH (Norges Tekniske Høgskole), Norway in 1992. In 1998 she received her Ph.D. degree in acoustics from the University of Aix-Marseille II, Marseille. After a post doctoral stay at the University of Stanford - CCRMA, California, she obtained a researcher position at the CNRS (Centre National de la Recherche Scientifique) in Marseille in 2002. In 2017 she co-founded the interdisciplinary art- science laboratory PRISM - Perception, Representations, Image, Sound, Music (www.prism.cnrs.fr) in Marseille. Her research activities mainly focus on investigations of auditory and multimodal perception through so-called perceptual engineering which consists of crossing different disciplines to link physical and signal knowledge with human perception and cognition.

Richard Kronland-Martinet has a background in theoretical physics and acoustics. He got the "Doctorat d'Etat ès Sciences" degree (habilitation) in 1989 from Aix Marseille University, France, for his pioneer work on analysis and synthesis of sounds using time-frequency and time-scale (wavelets) representations. He is currently Director of Research at the National Center for Scientific Research (CNRS), and head of the Interdisciplinary laboratory PRISM (Perception, Representations, Image, Sound, Music). His primary research interests are in analysis and synthesis of sounds with a particular emphasis on high-level control of synthesis processes. He published more than 250 journal articles and conference proceedings in this domain. He recently addressed new scientific challenges linked to the semantic description of sounds and to their synthesis control based on sound invariants, using an interdisciplinary approach associating signal processing, physics, perception and cognition.