PHYSICALLY-BASED AUDIO RENDERING OF CONTACT

Federico Avanzini

Dipartimento di Elettronica ed Informatica
Università degli Studi di Padova
Via Gradenigo 6/A, 35131 - Padova, Italy.
avanzini@dei.unipd.it

Matthias Rath, Davide Rocchesso

Dipartimento di Informatica
Università degli Studi di Verona
Strada Le Grazie, 37134 - Verona, Italy
{rath;rocchesso}@sci.univr.it

ABSTRACT

This paper describes an algorithm for real-time synthesis of contact sounds for interactive simulations and animation. The algorithm is derived from a physically-based impact model, and the acoustic characteristics of colliding objects can be realistically simulated by properly adjusting the physical parameters of the model. A technique for describing the spatial dynamics of a resonating object is proposed, which allows simulation of position-dependent interaction. It is shown that the numerical implementation leads to an efficient sound synthesis module, that runs in real-time on low cost platforms. The effectiveness of the model is demonstrated, and its applications are discussed.

1. INTRODUCTION

Sound has been recognized to be an effective channel of information within human-computer interfaces. It can be used to convey alarm and warning messages, status and monitoring indicators, and encoded messages [1]. It has been shown that the usability of interfaces can be significantly improved by adding carefully designed sounds to graphical menus, buttons and icons [2]. In animation and interaction, non-speech sound plays an important role in integrating visual information. It affects the way the user perceives events, and provides the user with a sense of presence and immersion in a synthetic environment.

Research in audio for multimedia systems has traditionally focused on techniques related to auralization of environments. Properly designed reverberation and sound spatialization algorithms provide information on the size and shape of an environment, as well as the location of auditory events [3]. Far less attention has been devoted to the audio *sources*, and the interaction mechanisms that are involved in sound production.

Experimental psychologists have provided a great deal of interesting results on the perception of sound sources. Research in ecological acoustics shows that listeners typically tend to describe sounds in terms of the sound-producing events rather than acoustic parameters. Gaver [4] refers to this attitude as "everyday listening", and discusses how to use everyday sounds to synthesize *auditory icons*. To this end, it is necessary to identify those acoustic cues which convey information about specific physical properties of objects (such as size, shape, material) and their interactions.

Recent research in sound modeling has exploited the above results. Klatzky *et al.* [5] have shown that auditory information can be used in simulated contact with virtual objects, to elicit perception of material. Van den Doel *et al.* [6] describe models for audio rendering of collisions and continuous contact (friction and rolling). Convincing results are obtained, however the contact models used in these works do not fully rely on a physical description, and as a consequence the attack transients and the overall realism are affected. Moreover, due to the lack of physical description of contact forces, the control parameters of the sound models are not easily associated to physical dimensions.

A fully physical approach has been adopted by O'Brien *et al.* [7], who have simulated the behavior of three-dimensional objects using a finite element method. The computation is used for generating both visual and audio animation, hence a high degree of coherence and perceptual consistency can be achieved. On the other hand, finite elements have high computational costs and are possibly "too" accurate, i.e. the models take into account also those sound features that are not perceivable or relevant.

An alternative approach to physical modeling amounts to first investigating what acoustic cues are significant for auditory perception, and designing the sound synthesis algorithms on the basis of such investigation. This way, only the important sound features are incorporated into the models, while less salient cues are discarded. This design process naturally leads to "cartoon" sound models, i.e. simplified models of real mechanisms and phenomena, in which some features are magnified. One advantage is that veridical audio output and physically based control are achieved using simple and computationally efficient synthesis algorithms.

Using such a "cartoon" approach, we have developed a model in which a hammer and a resonator interact through a non-linear contact force. The model is controlled through physical parameters, such as impact velocity and position, and produces realistic impact sounds. Moreover, it allows control on salient acoustic cues: specifically, we have shown that the resonator model can be "tuned" to various materials [8], while the parameters of the non-linear force can be adjusted to different levels of "hardness" [9].

This paper addresses the issue of position-dependent interaction. Hitting an object at different positions results in different responses, since its resonances are excited to different extents. After reviewing in Sec. 2 the hammer-resonator system and the numerical implementation, in Sec. 3 we propose a strategy to integrate position-dependent interaction in the model with little computational overhead. Results and applications are discussed in Sec. 4.

2. IMPACT MODEL

2.1. Hammer, resonator, interaction

The resonator is modeled using modal synthesis techniques [10], where a resonating structure is described in terms of its normal modes. The state of the system is here written as the vector of its,

0-7803-7304-9/02/$17.00 C2002 IEEE

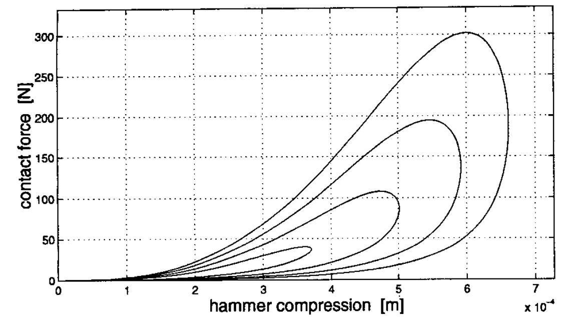

Figure 1: Compression/force characteristics.

generally infinitely many, modal states (as it would, in the spatial description be seen as the vector of states of spatial points). Modal description is in principle equivalent to a description in spatial variables: modal and spatial parameters are related through a linear basis transformation. Modal parameters, though, relate differently to human perception, which is of great importance in terms of implementation and especially of simplification/abstraction[1].

Each modal state $\boldsymbol{w}_j = \binom{x_j}{\dot{x}_j}$ follows a differential equation

$$\ddot{x}_j + r_j \dot{x}_j + k_j x_j = f_j, \tag{1}$$

where $r_j \geq 0$ and $k_j > 0$ are the damping and the elastic constant of the jth mode, respectively, while f_j is the sum of external forces on the mode. For sufficiently small damping ($r_j^2 < 4k_j$), the impulse response of system (1) is given by

$$x_j^{(impulse)}(t) = e^{-t/t_j} \sin(\omega_j t) \ , \tag{2}$$

where the resonance frequency ω_j and the decay time t_j are

$$k_j = \omega_j^2 + 1/t_j^2, \quad r_j = 2/t_j \ . \tag{3}$$

For $\omega_j^2 \gg 1/t_j^2$, i.e. "very small" damping ($r_j^2 \ll k_j$), the resonance frequency is approximated by $\omega_j \simeq \omega_0 \triangleq \sqrt{k_j}$.

In practice modes are always of finite number n, since the ear bandwidth as of any system of processing/reproduction is finite. The transformation from the mode states to a spatial state variable in a point P is then $\boldsymbol{w}_P = \sum_{j=1}^{n} a_{P\,j}\, \boldsymbol{w}_j$. Equivalently:

$$x_P = \sum_{j=1}^{n} a_{P\,j} x_j = \boldsymbol{a}_P \boldsymbol{x}' \quad \text{and} \quad \dot{x}_P = \boldsymbol{a}_P \dot{\boldsymbol{x}}' \ , \tag{4}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{a}_P = (a_{P\,1}, \ldots, a_{P\,n})$. In a similar way, a force f applied at a spatial point Q is distributed to the separate modes according to

$$f_j = a_{Q\,j}\, f, \qquad j = 1, \ldots n \tag{5}$$

The hammer is modeled as a lumped mass which moves freely except when it collides with the resonator. Hence its position x_h is simply described by the equation

$$m_h \ddot{x}_h = f, \tag{6}$$

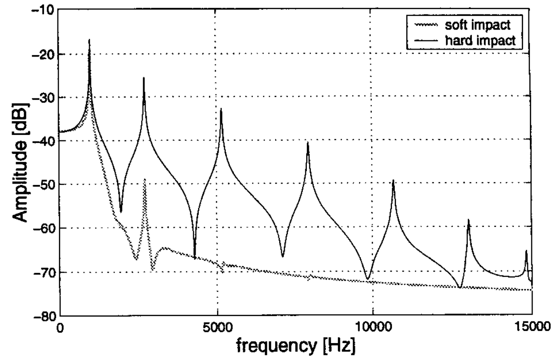

Figure 2: Sound spectra obtained when hitting a resonator with a soft mallet (low m_h/k) and with a hard hammer (high m_h/k).

where m_h is the hammer mass. Note that Eq. (6) is equivalent to Eqs. (1), (5), where $n = 1$, $r_1 = k_1 = 0$, and $a_{Q\,1} = 1/m_h$. Thus, the hammer model can be regarded as a special case of the modal resonator. In fact, the final implementation of the impact model treats *both* the interacting objects as modal resonators.

What is left is an equation which describes the interaction between the two objects. Hunt and Crossley [11] proposed a model for the contact force between two colliding objects, under the hypothesis that the contact surface is small:

$$f(x, \dot{x}) = \begin{cases} -kx^{\alpha} - \lambda x^{\alpha} \cdot \dot{x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{7}$$

where the value of α depends only on the local geometry around the contact surface. The variable x stands for the hammer compression, i.e. the difference between the resonator displacement and the hammer position. Therefore, when $x > 0$ the two objects are in contact. Marhefka and Orin [12] have used this model for describing contact with the environment in dynamic simulations of robotic systems. Figure 1 shows the compression/force characteristics for a hammer hitting a hard surface with various impact velocities. Note that hysteresis occurs, i.e. the paths during loading and unloading are different. This effect is due to the presence of a dissipative term in Eq. (7).

2.2. Properties

When the system is discretized, the modal resonator appears as a filter bank of n second-order bandpass filters, each accounting for one specific mode of the resonator. The filter parameters (center-frequency and quality factor) are accessed through the physical quantities r_j, k_j described above. Due to the non-linear nature of the contact force, computational problems occur in the numerical hammer-resonator system. These can be handled by computing the contact force iteratively at each time step [8, 9].

The sound model has been tested in previous studies in order to assess its ability to convey perceptually relevant information to a listener. A study on materials [8] has shown that the decay time is the most salient cue for material perception. This is very much in accordance with results by Klatzky *et al.* [5]; however, the physical model used here is advantageous over using a signal-based sound model as in [5], in that more realistic attack transients are

---

[1] It might be stated that the spatial description of an object rather refers to its visual appearance whereas modal properties have a closer relationship to auditory perception.

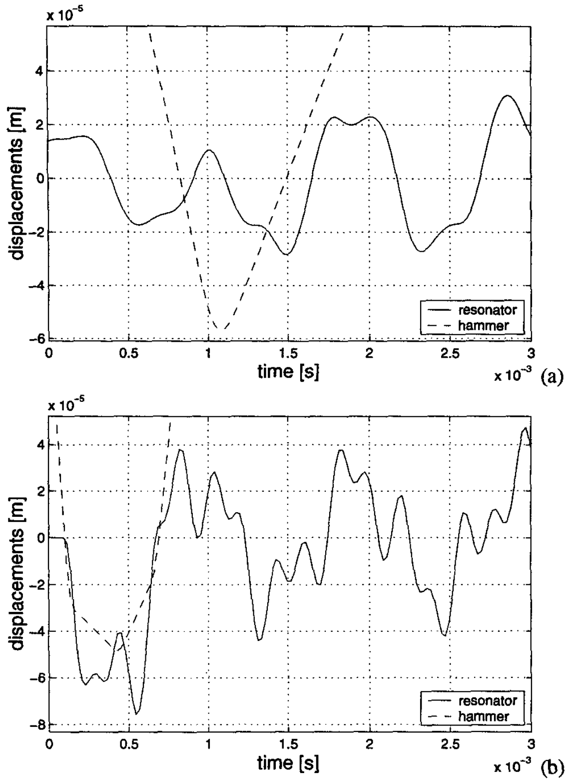

Figure 3: Numerical simulations; (a) impact on an oscillating resonator; (b) micro-impacts in a hard collision. Intersections between the solid and the dashed lines denote start/release of contact.

obtained. A study on hammer hardness [9] has shown that the contact time (i.e. the time after which the hammer separates from the resonator) can be controlled using the physical parameters. Specifically the ratio m_h/k is found to be the most relevant parameter in controlling contact time. Figure 2 shows an example of soft and hard impacts, obtained by varying m_h/k.

Due to the physical description of the contact force, realistic effects can be obtained from the model by properly adjusting the physical parameters. Figure 3(a) shows an example output from the model, in which the impact occurs when the resonator is already oscillating: the interaction, and consequently the contact force profile, differs from the case when the resonator is not in motion before collision. This effect can not be simulated using pre-stored contact force profiles as in [6]. Figure 3(b) shows an example of "hard collision", obtained by giving a very high value to k in Eq. (7), while other model parameters have the same values as in Fig. 3(a). It can be noticed that several micro-collisions take place during a single impact. This is qualitatively in accordance with the remarks about hard collisions by van del Doel *et al.* [6].

Gesture-based control models can be designed for the sound synthesis algorithm, where the impact velocity of the hammer is used as the main control parameter. In a recent work, a virtual percussion instrument has been designed, where the sound model is based on the hammer-resonator system described above and the control model is implemented through a gestural interface [13].

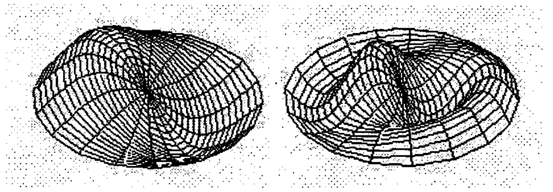

Figure 4: A circular membrane displaced from its rest position along the axes of mode(1,1) (left) and mode(1,2) (right). The frequencies of vibration along these axes are 1.593 and 2.917 times that of mode(0,1) (the "fundamental").

The subtle sound nuances achieved by the model provide a rich timbral palette that can be very useful in sonification and auditory display. Its responsiveness to user gestures is especially suitable for feedback sounds, where the audio information should give confirmation about the extent and quality of the performed action.

3. POSITION-DEPENDENT INTERACTION

Figure 4 shows a membrane displaced from its rest position along isolated modal axes. The distance of each point of the membrane from the "rest plane" is here proportional to the weighting factor of the mode at this position. Note that the section lines of the mode–shape with the rest plane stay fixed through the whole movement along this modal axis, since the weighting factors at these positions are obviously 0. Correspondingly, an external force applied at these node lines does not excite the mode at all.

Determination of position dependent weighting factors

There are several possible approaches to gain the position dependent weights. In the case of a finite one-dimensional system of point masses with linear interaction forces, modal parameters are exactly found through standard matrix calculations. Most systems of interest of course do not fit these assumptions. In some cases the differential equations of distributed systems can be solved analytically, giving the modal parameters; this holds for several symmetrical problems as circular or rectangular membranes.

Alternatively, either accurate numerical simulations (e.g. waveguide mesh methods) or physical measurements on real systems can be used. Impulse responses computed (or recorded) at different points then form a basis for the extraction of modal parameters. The acoustic "robustness" of the modal description allows convincing approximations on the basis of microphone-recorded signals of e.g. an objects struck at different points, despite all the involved inaccuracies: spatially distributed interaction, as well as wave distribution through air, provide signals that are quite far from impulse/frequency responses at single points.

The following considerations illuminate the possibility of aforementioned estimations. Equations (1) and (2) correspond to the frequency response of a resonant lowpass filter. The transfer function connected to one or a pair of spatial points of the system is a weighted sum of these responses with position dependent factors. Even in non-ideal recording conditions, the prominent modes can be identified from peaks in the response. The level of the peak j reflects the position dependent weight while its width is related to the time factor t_j. Decay times can though be extracted more easily from STFT values at different temporal points. The clear perceptual character of these parameters finally allows "tuning by ear" which in many situations is the final judging instance.

Qualitative observations on modal shapes, exemplified in Fig. 4, can be effectively used in a context of cartoonification: for modes of higher frequencies the number of nodes increases and its spatial distance accordingly decreases.

1. One consequence is that for higher modes even small inaccuracies in interaction or pickup position may result in strongly different weighting factors, so that an element of randomization can here add "naturalness".

2. For interaction positions close to a boundary, which is a common node for all modes, the lowest modes gradually disappear and higher modes (with smaller "regions of weight") relatively gain in importance. This phenomenon can be well noticed for a drum: if the membrane is struck close to the rim the excited sound gets "sharper", as the energy distribution in the frequency spectrum gets shifted upwards ("rimshots"). For a clamped bar higher partials are dominant near the fixed end, whereas lower frequencies are stronger for strokes close to the free vibrating boundary (noticeable in sound adjustments of electromechanical pianos).

Similar considerations apply to points of symmetry: the resonant modes with modal shapes antisymmetric to central axes are not present in the center of a round or square membrane. They consequently disappear "bottom–up" when approaching the center point.

4. IMPLEMENTATION AND USE

The impact model has been implemented as a module for the real–time sound processing software *PD*[2]. Controls are handled with the temporal precision of an audio buffer length; values for the *PD* buffersize are e.g. 32 or 64 samples, that provide temporal accuracies of ≈ 0.75 ms and ≈ 1.5 ms, respectively. The iterative algorithm used for solving the non-linear interaction has been observed to exhibit a high speed of convergence: the number of iterations at each time step is never higher than four. As a consequence, the module runs in real-time on low-cost platforms.

The quality of the audio generated from the model has been assessed through both informal evaluations and formal listening tests [8]: in general, the impact sounds are perceived as realistic. The control on the impact location provides convincing results.

The algorithm is being used as the kernel of a variety of sound models. It is known that certain sounds are very effective to convey information about continuous processes (e.g., the sound of a vessel being filled provides a precise display of the level of liquid [14]). Specifically, phenomena like sliding, rolling, scraping, crumpling, and so on are easily recognized by ear, thus being perfect candidates for auditory display. If properly cartoonified, they can be used as dynamic auditory icons that give information on events and processes. Just by temporal organization and dynamic control of micro-impacts, sound models have been developed for a variety of these phenomena (namely, bouncing, breaking, and rolling).

Again, the veridical control that can be exerted on the contact model, and the possibility to manipulate the control variables in real-time, gives much more flexibility to the sound designer. In the end the sound cartoons of complex processes such as scraping or rolling turn out to be more engaging and pleasant as compared to the results of signal-based models.

5. ACKNOWLEDGMENTS

This work has been supported by the European Commission under contract IST-2000-25287 (project "SOb - the Sounding Object": `www.soundobject.org`).

6. REFERENCES

[1] W. Buxton, "Using our Ears: an Introduction to the Use of Nonspeech Audio Cues," in *Extracting Meaning from Complex Data: Processing, Display, Interaction*, E. J. Farrel, Ed., pp. 124–127. Proceedings of SPIE, Vol 1259, 1990.

[2] S. A. Brewster and M. G. Crease, "Correcting Menu Usability Problems with Sound," *Behaviour and Information Technology*, vol. 18, no. 3, pp. 165–177, 1999.

[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, 1994.

[4] W. W. Gaver, "How Do We Hear in the World? Explorations in Ecological Acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, Apr. 1993.

[5] R. L Klatzky, D. K. Pai, and E. P. Krotov, "Perception of Material from Contact Sounds," *Presence*, vol. 9, no. 4, pp. 399–410, Aug. 2000.

[6] K. van del Doel, P. G. Kry, and D. K. Pai, "FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation," in *Proc. ACM Siggraph 2001*, Los Angeles, Aug. 2001.

[7] J. F. O'Brien, P. R. Cook, and G. Essl, "Synthesizing Sounds from Physically Based Motion," in *Proc. ACM Siggraph 2001*, Los Angeles, Aug. 2001.

[8] F. Avanzini and D. Rocchesso, "Controlling Material Properties in Physical Models of Sounding Objects," in *Proc. Int. Computer Music Conf. (ICMC'01)*, La Habana, Sept. 2001, Available at `http://www.soundobject.org`.

[9] F. Avanzini and D. Rocchesso, "Modeling Collision Sounds: Non-linear Contact Force," in *Proc. COST-G6 Conf. Digital Audio Effects (DAFX-01)*, Limerick, Dec. 2001, pp. 61–66, Available at `http://www.soundobject.org`.

[10] J. M. Adrien, "The Missing Link: Modal Synthesis," in *Representations of Musical Signals*, G. De Poli, A. Piccialli, and C. Roads, Eds., pp. 269–297. MIT Press, 1991.

[11] K. H. Hunt and F. R. E. Crossley, "Coefficient of Restitution Interpreted as Damping in Vibroimpact," *ASME J. Applied Mech.*, pp. 440–445, June 1975.

[12] D. W. Marhefka and D. E. Orin, "A Compliant Contact Model with Nonlinear Damping for Simulation of Robotic Systems," *IEEE Trans. Systems, Man and Cybernetics-Part A*, vol. 29, no. 6, pp. 566–572, Nov. 1999.

[13] M. Marshall, M. Rath, and B. Moynihan, "The Virtual Bhodran — The Vodhran," in *Int. Workshop on New Interfaces for Musical Expression*, Dublin, May 2002.

[14] G. B. Vicario, "Phenomenology of Sound Events," in *Deliverable of the Sounding Object Project Consortium*, D. Rocchesso, Ed. 2001, Available at `http://www.soundobject.org`.

---

[2] `http://www.pure-data.org/doc/`