Real-time Contact Sound Synthesis for Multisensory Interaction

JUAN LIU[† ††], and HIROSHI ANDO[† ††]

1. Introduction

The sense of hearing plays many important roles in our daily lives. We may produce abundant sounds by interacting with objects in our reach. These sounds convey plenty of information about the object, such as material, roughness, stiffness, the speed and type of our actions et al. Concurrent sound feedback provides us a sense of presence and realism. In virtual reality systems, users may immediately notice the unnatualness if the interface has no sound effect or provides mismatched sound.

This paper addresses the problem of real-time audio synthesis for haptic interface and provides a physically motivated integrated model for various contact interactions, such as impacting, scratching and inhibiting. The prototype system based on a force feedback device (PHANTOM®) is able to offer tightly synchronized visual-haptic-auditory stimuli, which makes users feel like manipulating the real object.

2. Related Work

For continuously interactive sound synthesis, playing back prerecorded sounds is unsatisfactory and inapplicable since we cannot predict the changes in the sound. People excite sound-producing objects with a variety of other objects in a variety of ways. A lot of researches have been carried out in acoustics, physics and digital signal processing for physical model of sound production and parameterized sound synthesis methods [1]. Van den Doel, Kry and Pai [2] mapped analytically computed vibration modes onto object surfaces for automatically synthesizing sound effects for animation and simulation. However, they used different methods for impact, scraping and rolling scenes and the objects they simulated are made of one kind of material and simple structure. Because in animation the transition among motions can be predicted from the dynamic model, it is unclear whether their models can work well for human-machine interaction, in which users' intention and active action introduce new problems. O'Brien, Cook and Essl [3] described an off-line system to compute sound and motion from a single physical model of deformable bodies. The sound is generated by computing the propagation of acoustic pressure waves induced by motion. Barrass and Adcock [4] implemented scraping sound using granular synthesis for haptic workbench. But the impact sound was not integrated effectively in their system. Since our aim is to synthesize realistic sound for arbitrary contact actions carried out through haptic interface, our work is closely related to on-line algorithms based on a physical model.

† Multimodal Communication Group, Universal Media Research Centre, National Institute of Information and Communications Technology (NiCT)

†† Cognitive Information Science Laboratories, Advanced Telecommunications Research Institute International (ATR)

3. Integrating Sound Models and User Actions

Contact sounds we hear during our interaction with solid objects are waves radiated by vibrating structures. Various materials and shapes deform in different ways so that we can tell the difference of their sounds. To simulate the interactive sounds emitted from a solid object, we may use a physically motivated modal synthesis model $\mathbf{M} = \{\mathbf{f}, \mathbf{d}, \mathbf{A}\}$, which describes the object by a bank of damped harmonic oscillators with modal frequencies f_n, decay rates d_n and amplitudes a_n where $n = 1, ..., N$ [1, 2]. The vibrating structure is taken as a linear time-invariant system (LTI), which can be characterized by its impulse response $h(t)$.

$$h(t) = \sum_{n=1}^{N} a_n e^{-d_n t} \sin(2\pi f_n t) \quad (1)$$

Contact sound $y(t)$, the output of the system, is the convolution of the input stimulus $x(t)$ and the impulse response, i.e. $y(t) = (h * x)(t)$. Therefore the sound synthesis problem can be decomposed into two parts, i.e. to obtain vibration model parameters and to get input excitation model.

In many situations if the way of interaction is different, such as striking or scraping, contact sounds are totally different even when we touch at the same location. Therefore several models (i.e. several sets of parameters) are necessary in order to provide a more complete description of the acoustic features of one surface. Then how should we trigger these models to generate sound? It is intuitive to choose a top-down strategy that excites corresponding model according to the type of an action (impacting or scratching). We call it the branched method. It might work for sound effect synthesis for animation with predefined plot, but in real-time interaction, the intention of the user is blind to the system. The system has to judge what kind of action is carrying out. However, to gather certain amount of data for the judgment takes time, while the sound has to be generated at the same moment when the user's stylus touches the surface. Since human's auditory temporal resolution is much higher than that of somatic sensation, in a system that updates haptic output at 1kHz and generates 44.1kHz audio output, the latency of judging process can be noticeable and annoying. Meanwhile, the fluctuation at the boundary of actions and sound models will also make the synthesized sound noise and unnatural.

We argue that a bottom-up strategy is more suitable. If we put the user's intention aside, and just analyze the vibration of the surface, in a process of interaction, the way of vibration is not pure impact or scratching mode. For example, when the stylus touches the surface, normal force will trigger the modes of vibration in impact model, even the user intend to scrape the surface. Or in some cases, the impact behavior may also introduce tangential element that excites the scratching sound model. These models are superposed rather than exclusive. So we propose a physically motivated way to integrate the

excitation of sound models. An inhibition factor is also introduced to characterize a typical interactive action that people intent to attenuate the sound by pushing the surface.

Take two kinds of interactive actions and sound models as an example (impact and scratching). Assume there is only one contact point at one moment. The synthesized sound is:

$$y(t) = (h_{imp} * x_{imp})(t) + (h_{scr} * x_{scr})(t) \quad (2)$$

where the input stimuli x_{imp} and x_{scr} are functions of contacting time, normal force exerted on the surface F_n and tangential force F_t. Along the increase of contact duration, the possibility of the action being an impact decreases, meanwhile that of the scratching action increases. The ratio of sounds changes accordingly.

$$x_{imp} = w_{imp}(t)(1-\rho)F_n(t),$$

$$x_{scr} = w_{scr}(t)(1-\rho)F_t(t),$$

$$w_{imp} = e^{-\alpha(t-t_0)\phi},\ w_{scr} = 1 - e^{-\alpha(t-t_0)\phi}.$$

$\phi = \{0,1\}$ is used to indicate whether the proxy of the stylus is touching the surface or not.

0: not contacting; 1: contacting the surface since t_0.

α is a coefficient which is related to the material.

$\rho = \{0,1\}$ is used to indicate whether the contact is an inhibition action or not. The change of position of the proxy on the surface in consecutive 2 cycles (totally 2 ms) is taken as a criterion.

0: more than or equal to threshold δ, no inhibition;

1: less than threshold δ, taken it as an inhibition behavior.

The inhibition action dramatically damps the sound:

$$h_{imp} = \sum_{n=1}^{N} a_n^{imp} e^{-(\rho D_{imp}+1)d_n^{imp}t} \sin(2\pi f_n^{imp} t),$$

$$h_{scr} = \sum_{n=1}^{N} a_n^{scr} e^{-(\rho D_{scr}+1)d_n^{scr}t} \sin(2\pi f_n^{scr} t),$$

where $D_{imp} > 1$, $D_{scr} > 1$, related to the material and the contact area of the stylus and the surface.

4. Manipulating Objects: from Real to Virtual

The purpose of our work is to duplicate a real object in the virtual environment, and when people interact with the virtual object, they may have the feeling that the real one is present in front of them. That requires us to provide multi-sensory stimuli and synchronize them closely. The prototype system as shown in Fig.1 consists of Reachin® 3D display, PHANTOM force-feedback device and the integrated audio part utilizing our real-time sound synthesis method proposed in above section.

Firstly the 3D shapes and texture images of the object are captured by a non-contact 3D digitizer (Vivid 910 from Konica Minolta), and various contact sounds of the real object are recorded in an anechoic room. The shapes and images taken from different viewpoints are stitched together forming a complete 3D model for visual and haptic rendering. The parameters of contact sound models are obtained by fitting them to the sampled sound files. For each sound model, around 200 mode frequencies are selected from the result of windowed discrete Fourier transformation. Then the damping and coupling amplitude parameters are estimated using least square algorithm. Finally each sound model is associated with the part of the object where their samples are taken. The data from the haptic device are used to control the output of sound models as described in Section 3.

Users wearing a stereo shutter glasses may see the 3D image of the object, touch it with the stylus in various ways, feel the force-feedback and simultaneously hear the contact sounds of high fidelity.

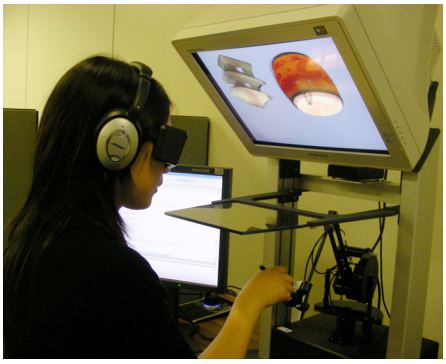

Fig. 1 Multi-sensory interaction system showing Japanese drum demonstration

5. Conclusions and Future Work

A method to utilize consistent sound synthesis models for various contact interactions is proposed and implemented in a multi-sensory interaction system. Since we can precisely control the frequency component, damping and amplitude parameters as well as the excitation of sound models with haptic feedback, we are going to devise some psychophysical experiments applying this system to investigate auditory cues that influence people's perceptual judgments and their relationship with other modalities. The next version of the system may incorporate spatial information of the sound sources and radiation efficiency. We are also working on aerodynamic sound and liquid sound to provide a wider application prospect for the system in design, entertainment, education, training and telecommunication.

References

1) Cook, P. R.: *Real Sound Synthesis for Interactive Applications*, A K Peters, Wellesley, MA (2002)

2) van den Doel, K., Kry, P. G., and Pai, Dinesh K.: FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation, in *Computer Graphics (ACM SIGGRAPH 01 Conference Proceedings)*, pp. 537-544 (2001).

3) O'Brien, J. F., Cook, P. R., and Essl, G.: Synthesizing Sounds from Physically Based Motion, in *Proceedings of SIGGRAPH 2001*, Annual Conference Series, (Los Angeles, California), pp. 529-536 (2001)

4) Barrass, S., Adcock, M.: Interactive Granular Synthesis of Haptic Contact Sounds, in *Proceedings of the Audio Engineering Society 22nd International Conference on Virtual, Synthetic and Entertainment Audio (AES22)*, Espoo, Finland. AES. pp. 270-277 (2002)