Active Measurement of Contact Sounds

Joshua L. Richmond and Dinesh K. Pai
Department of Computer Science
University of British Columbia
Vancouver, BC V6T 1Z4, Canada
{jlrichmo|pai}@cs.ubc.ca

Abstract

We describe a novel system for robotic measurement of contact sounds. This system is well suited for acquiring impulse-response sound models. A brief explanation of the sound models is included. We also describe experimental results to illustrate the process of acquiring sound models using our system, and a brief analysis of the results. The initial results are promising and could lead to applications in reality-based modeling and object recognition.

1 Introduction

Sound is a critical component of everyday physical contact. When an object is struck, the energy of impact causes deformations to propagate through the object, causing its outer surfaces to vibrate and emit sound waves. This sound carries important information about the material composition of the object, its shape and size, as well as the type and location of contact on the object. Humans can perceive these subtle properties (e.g., [2, 5]) and can often perform tasks relying only on sound. A typical everyday example is locating a stud inside a wall by tapping on it and listening to the sound response.

Our primary motivation in this work is to endow simulated objects in virtual environments with realistic contact sounds. Sound can greatly enhance the perception of contact in virtual environments; contacts occur at small spatial scales and are difficult to discern visually. Sounds can be used instead of, or as a complement to, haptic force feedback. Hardware for computer sound displays (sound cards, speakers) is far more mature, widespread, and inexpensive than that for haptic displays today.

In addition, the active sound measurement technique described here could be used for recognition and classification (e.g., [6]). Material composition could also be deduced by analysis of acoustic measurements [13, 6]. Other possible uses include diagnosis of structural defects and nondestructive testing.

These applications require a computational model of contact sounds to be completely effective. For virtual environments to provide realistic interactions, the sound generated by objects should reflect the location and strength of forces applied to them. Simply playing recorded samples does not offer this sort of flexibility. Object recognition and classification tasks also rely on models of the sound to form class prototypes.

While the physics of sound is well understood [7] and a number of computational models for contact sounds are known (e.g., [11]), a key challenge is how the parameters of these models are identified for existing objects. Real objects can have complex shapes and material composition, hidden internal structure, and subtle boundary conditions. Measurement of these properties for the purpose of constructing a contact sound model is extremely difficult.

In this paper we propose the direct robotic measurement of contact sound responses of existing objects. We describe our implementation of an automated sound measurement system, using the Active Measurement facility (ACME) [8] at the University of British Columbia. With this system we are able to collect a large number of registered sound samples automatically. These samples can then be used to estimate parameters of an impulse response sound model.

A brief explanation of an impulse response sound model is presented in Section 2. We then describe our active sound measurement approach in Section 3. The ACME facility is described, with emphasis placed on describing the end effector used to acquire the sound samples. The general process of acquiring registered sound samples with ACME is included in Section 3.4. The results of a typical data collection are presented and analyzed in Section 4. Section 5 presents a summary and conclusions.

1.1 Related Work

The physics of sound has been thoroughly studied; a good reference on acoustics is [7]. However, computer synthesis of sounds made by contact interactions is relatively recent, though the effectiveness of such sounds has been known for a long time. Indeed, radio and film productions have long added artificial "Foley" sounds to the sound track to emphasize contact events such as footsteps.

An overview of the physical factors involved in producing natural contact sounds was presented by Gaver [2]. Several synthesis methods for impact, scraping, and composite sounds were described in [3, 9, 1]. However, no general method was given to compute the free parameters of the synthesis methods, such as the set of eigenfrequencies, the relative amplitudes of the partials, and the bandwidths of the frequencies. Simulation of such shape dependent contact sounds was described by van den Doel and Pai [12]. They demonstrated real time synthesis of shape-dependent contact sounds, and computation of model parameters for simple shapes. A brief overview is given in Section 2.

There has been little work on estimating sound models of actual, existing objects, though some of the work in nondestructive testing, noise, and audio compression can be interpreted in this way. Identification of material properties of objects was described by [6]. [10] acquired sound models of the type used in this paper by hand. Reality-based modeling of contact sounds from instruments has attracted some attention in music (e.g., [4]). To our knowledge, the system described in the present paper is the first to automate the acquisition of sound models.

2 Sound Model

By *sound model* we mean a mathematical model which can be used to either simulate the actual sound response of an object, or for object recognition. The adequacy of a model depends on its intended use; for interactive applications the model should allow fast, real-time simulation, while being sufficiently accurate to represent important features of the sound.

Sound models of objects can be very complex. For instance model parameters could include the geometry, boundary conditions, mass distribution, and the Lamé constants of elasticity for the object and its surrounding medium. Assuming linear elasticity, which is reasonable for light contacts for many materials, the vibration of the object is described by a function $\boldsymbol{u}(\boldsymbol{x}, t)$, which represents the deviation from equilibrium of the point $\boldsymbol{x}$; $\boldsymbol{u}$ obeys a wave equation of the form

$$(A - \frac{1}{c^2}\frac{\partial^2}{\partial t^2})\boldsymbol{u}(\boldsymbol{x}, t) = F(\boldsymbol{x}, t) \qquad (1)$$

with c a constant (related to the speed of sound in the material), and A is a self-adjoint differential operator under the specified boundary conditions [7].

To simulate such a model would require the solution of hyperbolic PDEs to obtain the vibration of the object and propagation of sound in the medium. This is clearly unsuitable for real-time simulation and analysis. In [11, 10] we developed a simpler impulse-response model for synthesizing contact sounds. For linear elastic materials, the vibration of the object can be computed from the Green's functions, or impulse responses, of the object with specified boundary conditions. Finally using a model of material damping due to internal friction (following [13, 6]), the sound pressure $p(t)$ is given by

$$p(t) = e^{-t/\tau_0} \sum_{i=1}^{n_f} a_i e^{-t f_i \pi \tan\phi} \sin(2\pi f_i t). \qquad (2)$$

Here, f_i are the frequencies of the vibration modes of the object with the specified boundary conditions; we keep only the n_f frequencies in the human range of hearing (approximately below 20kHz). The amplitudes a_i depend on the contact location. The reader can easily hear the effect of this dependence by tapping a coffee mug at different locations. The damping factor, $f_i \pi \tan\phi$, is frequency dependent; for ideal crystalline materials with internal friction, $\tan\phi$ is a material constant [14]. Humans can distinguish between different materials using $\tan\phi$, as shown in [5]. To provide more modeling flexibility, we will assume $\tan\phi$ can vary with frequency and impact location, leading to the final simple model

$$p(t) = \sum_{i=1}^{n_f} a_i e^{-d_i t} \sin(2\pi f_i t). \qquad (3)$$

However, the model parameters for most real objects can not be obtained easily if the object has complex geometry and material properties. In [11], these parameters were derived for simple object geometries and boundary conditions. But for most objects, determining these parameters is difficult even if geometry and boundary conditions were readily available. Determining object geometry and boundary conditions (including the internal structure of the object) is non-trivial. Thus it is useful to directly estimate these location dependent sound parameters from direct, controlled measurements of an object.

3 Active Sound Measurement

Acquiring sound models from existing objects requires recording sound samples produced by striking the objects at precisely registered locations on their surface. Such sound measurement (sample acquisition) could, of course, be performed manually with a microphone, hammer and ruler, as

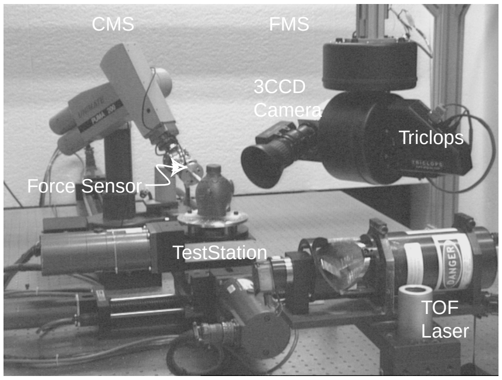


Figure 1: ACME Facility Overview

in [10]. But this is imprecise in both the location and intensity of the applied force; it is also a very tedious process.

Robotic measurement is therefore an attractive alternative. Such a robotic system should be able to acquire a model of the object shape, locate a dense set of points on a object to sample, move a sound measurement device to these points, make repeated light impacts at these points, and record the resulting sounds. We now describe our sound measurement system which is a part of the UBC Active Measurement facility (ACME) [8].

An overview of the ACME system is given in Section 3.1 below. Next we focus on the sound measurement; we do not address the acquisition of shape in this paper. Section 3.2 describes an end effector designed for sound measurement. Section 3.3 discusses the software used to capture the sound samples.

3.1 ACME

ACME is a telerobotic system with fifteen degrees of freedom. It is designed to acquire the rich set of measurements required for building reality-based models. A test station, gantry, pan/tilt unit and robot arm (see Figure 1) provide computer control over the motion of both the test object and the measurement sensors. Available sensors in the ACME system are a 3-CCD camera, a force/torque sensor, trinocular stereo vision system, laser range finder and microphones.

Every sensor and actuator within the system can be controlled remotely from the Internet using a Java^{TM}[1]-based teleprogramming interface. Using the current suite of available sensors, many different types of models can be created. These include deformation models, reflectance models and, as described in this paper, sound models. Each of the measurements is registered to a common frame of reference, making it easy to generate a multi-modal representation of objects.

Using the ACME facility simplifies acquiring sound samples from hundreds of locations on the surface of any object. It is also easy to obtain multiple samples at the same surface location in order to create a better estimate of the sound model. The trinocular stereo vision system may be used in conjunction with the sound sampling to provide a surface model by which the sound effector can be coarsely positioned by the robot arm. Fine motion to achieve a small offset from the surface can be performed using force sensing as described below.

3.2 Sound Effector for ACME

The acoustic impulse response is elicited from an object by striking it with an impulsive force at a point on the object. Since attaining a perfectly impulsive force is physically unrealisable, it must be approximated by a force that is well localized in time and space. The force must be powerful enough to produce a measurable sound while minimizing its contact time with the object. This is admittedly a very rough approximation, but has been shown to produce adequate sound models [10].

To acquire impulsive sound samples using ACME, a device was required that could deliver a near-impulsive impact at any orientation and at any location on an object. The 6-DOF Puma 260 robot arm provides a suitable base for such an end effector. The arm itself is not used to deliver the impulsive force, for two reasons. First, the inertia of the arm would make it difficult to quickly retreat after striking an object, resulting in large contact times and a large total impulse delivered. Second, repeated large impact loads could damage gearing and other mechanical elements in the robot.

Instead, our design uses a commercially-available Ledex push-solenoid. The solenoid is mounted in an aluminum bracket that connects it to the robot arm and to a microphone (See Figure 2). A spring is attached to the solenoid plunger to retract it after impact, and to retain the retracted position against gravity in certain orientations. To strike an object, the solenoid is powered under computer control for 30 ms. This design delivers a low-inertia, near-impulsive impact to objects 5 mm in front of the solenoid.

Two sensors are connected to the solenoid's mounting bracket. The first is an Optimus^{TM}[2] condenser microphone with a flat frequency response from 70 to 16 000 Hz. The

---

[1] Java is a trademark of Sun Microsystems Inc., MountainView, Ca., USA

[2] Optimus is a trademark of Tandy Corporation, Fort Worth, TX, USA

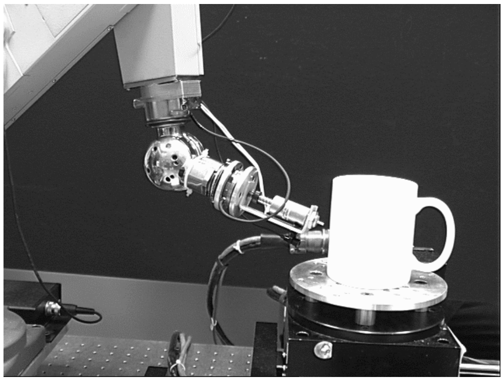


Figure 2: Detail of Sound Effector

microphone is positioned 10 mm from the surface of the object being struck. Mounted between the bracket and the tip of the robot arm is an ATI Mini 40 6-axis force/torque sensor. This sensor is used for guarded moves and impedance control when positioning the robot arm.

A second microphone (indentical to the first) is positioned 60 cm away from the test object. This microphone is used to simultaneously capture the far field acoustic response.

Sound is typically captured using a PC sound card. Currently a Creative Sound Blaster Live!^{R3} is used to record the sound.

3.3 Sound Measurement Software

Sound samples are recorded digitally using the DirectX API for Windows. The program that collects the data runs as a separate process, on a different machine than the main ACME server. This process is controlled remotely by the ACME system, and streams the recorded data back to the user. This design permits real-time recording at 44.1 kHz. A typical sound sample is a 16-bit two second recording.

Real-time performance is achieved using data buffering at two levels. A DirectX buffer collects incoming data from the sound card. Data is then transfered to a buffer in a Java process which analyzes each frame in order to trigger the start and end of recording. Data that is to be recorded is streamed back to the ACME experiment. Our modular design facilitates the incorporation of custom triggering events (e.g., record after a specified amplitude threshold is exceeded).

3.4 Sound Sampling

The procedure for acquiring sound samples using the ACME facility is straightforward. The object to be modeled is fixed on the ACME test station rigidly enough that it will not move from the force of impact.[4] For each surface point that is to be struck, the robot arm is first positioned at approximately 20 mm from the object. The arm performs a slow guarded move toward the surface of the object, monitoring the force sensor until contact is made. Then the robot arm retreats 5 mm. From this position, the sound effector is actuated and the sound sample is recorded to file. Typically the solenoid will be actuated ten or more times at the same location to record multiple samples at the same surface point. These samples are then used to produce the complete sound model of the object. An example of the sound model acquisition process is discussed in Section 4.

The selection and distribution of sample locations on the object's surface is computed manually. Currently, sampling is performed on a regular linear grid in Cartesian space. A more efficient sampling algorithm would use the surface model or a difference threshold applied to sound models produced at surrounding locations. This is one area of future research.

4 Experimental Results

In this section we discuss an experiment in sound sample acquisition with our system. Ninety samples were collected over nine locations on the surface of a brass vase (Figure 3). In Section 4.1 we discuss the details of the experiment, and in Section 4.2 we analyze the data collected using the sound models they produced.

4.1 An Experiment

To illustrate the effectiveness of the system, a simple experiment was conducted. The test object, a brass vase, was secured at the center of the ACME test station. Sound samples (two seconds each) were recorded using the sound effector at nine different locations on the vase's surface. Sound samples were also recorded at a position 10 mm above the vase to record the noise of the solenoid. The locations on the vase's surface were evenly spaced 10 mm apart in a vertical line (see Figure 3). Ten samples were recorded at each location to estimate the variability of the system.

The samples recorded during the experiment were processed using the algorithm described in [10] to produce a

---

[3] Sound Blaster Live! is a registered trademark of Creative Technology Ltd., Milpitas, CA, USA

[4] We note that fixturing introduces boundary conditions on the vibration of the object. Therefore care should be taken to attach the fixture to areas which are to have fixed (Dirichlet) boundary conditions.

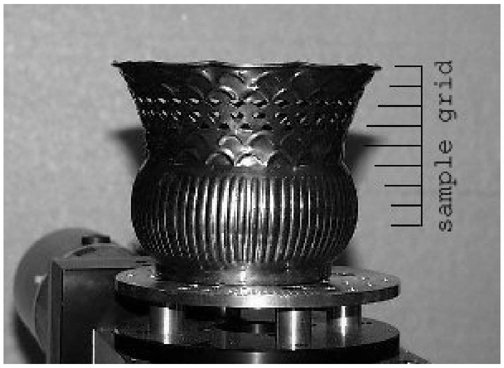


Figure 3: Test Object used for Experiment

sound model of each sample's forty dominant frequency modes. In total, ninety models were produced for the nine locations on the vase's surface.

A video of the experiment and samples of the recorded and reconstructed sounds are available online:

http://www.cs.ubc.ca/nest/lci/acme/ demos/SoundMeasurement.html

4.2 Analysis and Discussion

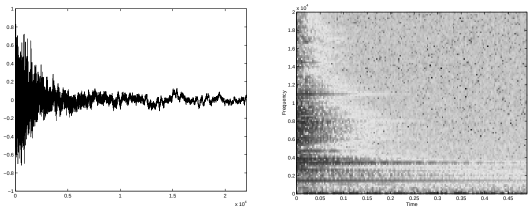
Examining the sample plots in Figure 4 reveals variation in tone, amplitude and decay over the height of the vase. Because the mouth of the vase (Z = 110 mm) is relatively free to vibrate, impact sounds recorded here have more low frequency modes whereas sounds produced near the bottom of the vase (Z = 40 mm) are higher in frequency. Sounds produced at the middle of the vase (Z = 70 mm) are longer in duration than at the other two locations. Audio samples can be compared at the web site mentioned previously.

An example of sounds reconstructed using this model is displayed in Figures 5 and 6. The reconstruction is quite good and appears to capture the significant modes well. Reconstructed sounds can be heard at the above web site.

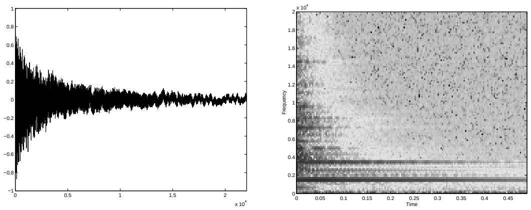
Figure 7 illustrates the consistency of the recorded samples. The amplitude plots and spectrograms are shown for three samples recorded at the same surface position (Z = 70 mm). Both visually and audibly, there is little difference.

5 Summary and Conclusions

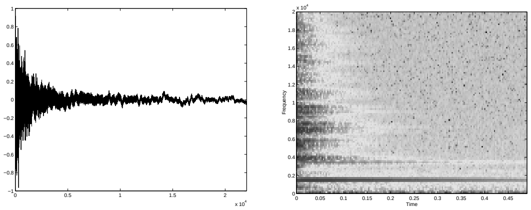
In this paper we have presented a system which can be used to automatically measure sound samples of objects. We proposed that these sound samples may be used to create impulse-response sound models. Applications of these sound models include virtual reality, games, movies and object identification. A simple experiment illustrated the effectiveness of the proposed system.



Z = 110 mm



Z = 70 mm



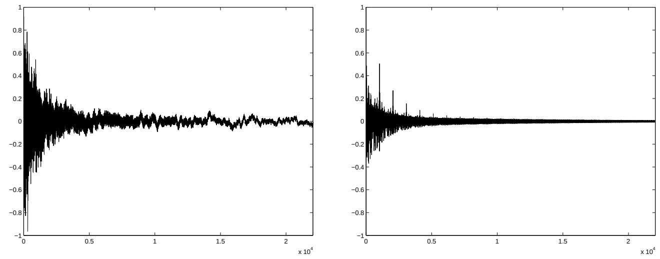
Z = 40 mm

Figure 4: Amplitude and Spectrogram Plots of Recorded Samples

Immediate work will focus on an adaptive sampling algorithm from which sample locations on the surface of arbitrarily-shaped objects may be determined on-line. Additionally, the trinocular stereo vision system will be incorporated into the planning algorithm. This input will facilitate sampling algorithms which adapt by using either the shape of an object, or the difference between sounds that have already been recorded. Our long term goal is to have the system automatically determine an optimal sampling algorithm using a surface which is reconstructed using the stereo vision system.

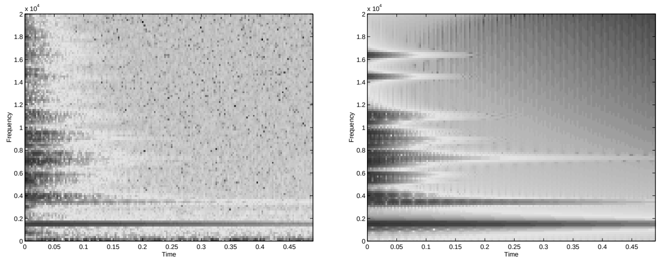
Future work will also be conducted into finding optimal ways of combining multiple samples from the same surface location. Extra information provided by the multiple samples will produce a more robust estimation of the model parameters.

These early results show that our approach to active sound measurement is promising. We believe the system is suitable for automatically measuring the sound response at precise surface locations. In combination with the rest of the



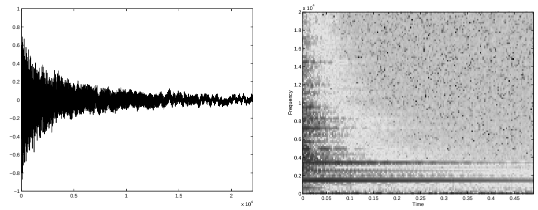
Sample Model

Figure 5: Plot of Sound Waves (Z = 40 mm)

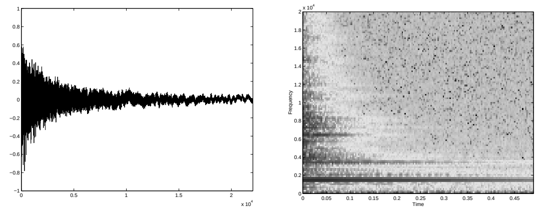


Sample Model

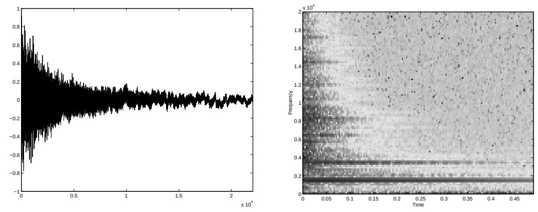
Figure 6: Spectrogram of Sound (Z = 40 mm)
Note: the trailing edges of the spectrogram peaks in the recorded sample are somewhat obscured by background noise in these printouts.



Sample #1 at Z = 70 mm



Sample #2 at Z = 70 mm



Sample #3 at Z = 70 mm

Figure 7: Comparison of Amplitudes and Spectrogram of Three Samples at Z = 70 mm

ACME facility, construction of realistic models becomes a simple and automated task.

Acknowledgments

Many people contributed significantly to the ACME system; for this we gratefully acknowledge the work of R. Barman, S. Kingdon, J. Lang and J. Lloyd. We also thank L. Kolby for his assistance building the mounting bracket and K. van den Doel for his work on the sound model. This work was supported in part by grants from the IRIS NCE, and NSERC. Richmond is supported by an NSERC scholarship and in part by the Advanced Systems Institute of British Columbia.

References

[1] Antoine Chaigne and Vincent Doutaut. Numerical simulations of xylophones. i. time domain modeling of the vibrating bars. *J. Acoust. Soc. Am.*, 101(1):539–557, 1997.

[2] W. W. Gaver. Synthesizing auditory icons. In *Proceedings of the ACM INTERCHI'93*, pages 228–235, 1993.

[3] W. W. Gaver. Synthesizing auditory icons. In *Proceedings of the ACM INTERCHI 1993*, pages 228–235, 1993.

[4] Matti Karjalainen and Julius Smith. Body modeling techniques for string instrument synthesis. In *Proceedings of the International Computer Music Conference*, pages 232–239, Hong Kong, 1996.

[5] Roberta Klatzky, Dinesh K. Pai, and Eric Krotkov. Perception of material from contact sounds. submitted to *Presence*, 1999.

[6] Eric Krotkov, Roberta Klatzky, and Nina Zumel. Robotic perception of material: Experiments with shape-invariant acoustic measures of material type. In *Experimental robotics IV: Lecture notes in control and information sciences 223, (pp. 204-211)*, New York: Springer-Verlag, 1996.

[7] A. P. Morse. The behavior of a function on its critical set. *Annals of Math.*, 40:62–70, 1939.

[8] Dinesh K. Pai, Jochen Lang, John E. Lloyd, and Robert J. Woodham. ACME, a telerobotic active measurement facility. In *Proceedings of the Sixth Intl. Symp. on Experimental Robotics*, 1999.

[9] Tapio Takala and James Hahn. Sound rendering. *Proc. SIGGRAPH'92, ACM Computer Graphics*, 26(2):211–220, 1992.

[10] K. van den Doel. *Sound Synthesis for Virtual Reality and Computer Games.* PhD thesis, U. British Columbia, Dept. of Computer Science, May 1999.

[11] K. van den Doel and D. K. Pai. Synthesis of shape dependent sounds with physical modeling. In *International Conference on Auditory Display (ICAD 96)*, Xerox PARC, Palo Alto, November 1996.

[12] Kees van den Doel and Dinesh K. Pai. The sounds of physical shapes. *Presence*, 7(4):382–395, 1998.

[13] Richard P. Wildes and Whitman A. Richards. Recovering material properties from sound. In Whitman Richards, editor, *Natural Computation*, Cambridge, Massachusetts, 1988. The MIT Press.

[14] C. Zener. *Elasticity and anelasticity of metals*. University of Chicago Press, 1948.