Robotic Measurement and Modeling of Contact Sounds

Joshua L. Richmond and **Dinesh K. Pai**
Department of Computer Science
University of British Columbia
2366 Main Mall
Vancouver, BC V6T 1Z4, Canada
+1 604 822 8197
{jlrichmo|pai}@cs.ubc.ca

ABSTRACT

In this paper we describe a robotic system for automatically acquiring sound measurements of everyday objects. These sound measurements are used to estimate the parameters of a model for contact sounds. The hardware comprising the system is described, and a general procedure for acquiring sound measurements is introduced. The results of a typical data collection are discussed briefly.

Keywords

Robotics, acoustic measurement, sonification

INTRODUCTION

Sound is a critical component of our everyday interactions. Contact sounds, in particular, provide us with important cues including event timing, material composition, force of interaction, and location of contact. By a "contact sound", we simply mean those sounds caused by forces produced by contact between two or more objects. The sound of a mug being set on a coffee table, or that of a pen rolling across a desk are examples of everyday contact sounds. Percussion instruments also produce contact sounds.

Such contact sounds provide valuable feedback to users of virtual environments. Coupled with haptic force feedback or on its own, sound effectively conveys properties of virtual objects and the interactions between them. A sufficient representation of an object's sound characteristics enables the synthesis of appropriate contact sounds for an arbitrary force applied at any point on the object's surface. A model that parameterizes this representation over the object's surface can be viewed as an acoustic map of the object, analogous to a texture map in graphics.

Much research has been conducted in the field of modeling the sound of objects, particularly musical instruments (see [1, 2, 4, 5, 7, 8]). Most of these models can be broadly classified into two approaches: analytical solutions and empirical solutions. By these classifications, we are referring to the method of calculating the model parameters. Our principal motivation is to model "everyday" objects (i.e., non-musical). Since these everyday objects usually have complex material composition and geometry, an empirical parameter-estimation approach is more practical.

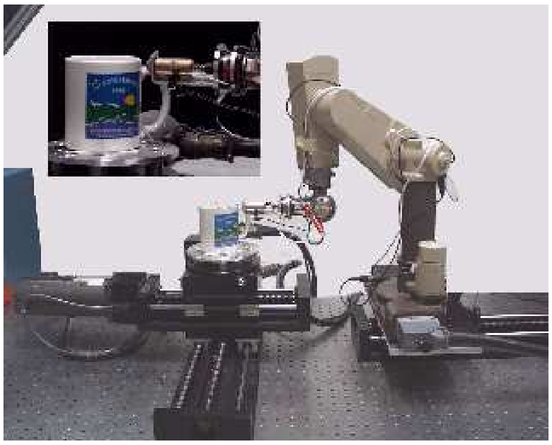
The parameters of empirical models are estimated from recorded samples of an object's acoustic impulse response. Typically these impulse responses are elicited at different locations by striking the object with a force hammer. The sound produced by the impact is recorded, and the model parameters are estimated from the recording. For examples of such models and methods see [2, 3, 8].

In this paper we describe a novel robotic system for automatically acquiring these sound samples. This sound measurement system is part of a larger facility: the Active Measurement facility (ACME) at the University of British Columbia. As such, the sound measurements are registered with the other models collected by ACME, such as surface and deformation models.

This paper begins with a description of robotic sound measurement. We then present a brief overview of the ACME facility, with emphasis placed on the sound measurement hardware. A general procedure for measuring contact sounds using ACME is outlined, and the results of a typical data collection are presented and analyzed. Plans for future work and conclusions are included in the last section.

ROBOTIC SOUND MEASUREMENT

The number of sound measurements required to populate even a simple virtual world with sonified objects is very large. Such sound measurement could be performed manually with a microphone and hammer as in [2, 8]. This would, however, become very tedious for many objects. In building a complete model of an object, these sound measurements should be registered to



(a) ACME Measurement Setup

(b) Detail of Sound Effector

Figure 1: The ACME Facility

the object's surface. Again, this can be performed manually with a ruler or 3-D positioning device, but it is difficult to achieve high precision and repeatability for objects of complex shape.

Robotic measurement is therefore an attractive alternative. Such a robotic system should be able to acquire a model of the object's shape, locate a dense set of points 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 the sound measurement system we designed to meet these requirements. This system is a part of the UBC Active Measurement facility (ACME) [6].

The Active Measurement Facility (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 1a) provide computer control over the motion of both a 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 two 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.

Sound Effector for ACME

The acoustic impulse response is elicited from an object by striking it with an impulsive force at a point on its surface. Applying this force is the task of the sound effector. 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 [8].

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.

Our design of the sound effector 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 1b). 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.

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

Two sensors are connected to the solenoid's mounting bracket. 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. The second sensor is an Optimus^{TM2} condenser microphone with a flat frequency response from 70 to 16 000 Hz. The microphone is positioned 35 mm from the surface of the object being struck. An identical microphone is located on the ACME gantry, and can be moved around the object being sampled.

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

SOUND SAMPLING PROCEDURE

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. The robot arm then 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 the following section.

The selection and distribution of sample locations on the object's surface is computed manually. Currently, a naive algorithm is used: sampling is performed on a regular linear grid in Cartesian space. A more efficient sampling algorithm would adaptively adjust the granularity of the sampling grid. This is one area of future research.

A TYPICAL DATA COLLECTION

For purposes of illustration, we describe a data collection conducted using the ACME sound measurement system. The object from which sound measurements were collected was the coffee mug pictured in Figure 1a. Two samples were recorded at each of six locations on the surface of the mug. These samples were then used to estimate the parameters of a sound model [8]. A surface model was not used for this experiment.

The six locations of sound measurement were spaced at 10 mm intervals along a vertical line in Cartesian space. While this sampling grid does not represent the most efficient selection of sampling locations, it is adequate for this simple collection. Typically, a sampling grid would be generated from a surface map produced by the Triclops stereo-vision camera.

By recording multiple samples at each surface location, a better estimate of the parameters can be obtained. The estimation technique used for the following analysis is developed in [8]. This technique does not capitalize on the availability of these multiple samples, so for this analysis only one measurement per location was used to create the model.

Results

Using the parameter estimation technique described in [8], a model of sound pressure (p) was generated for each of the samples (Eq. 1). Each model consists of the forty most dominant frequency modes of the sample. The amplitude (a_i), exponential damping coefficient (d_i) and frequency (f_i) of each mode (i) were estimated.

$$p(t) = \sum_{i=1}^{40} a_i e^{-d_i t} \sin(2\pi f_i t) \tag{1}$$

The estimation algorithm identifies the forty frequencies with maximum amplitude in each window of the signal's spectrogram. Each window thus casts votes for forty frequencies. Once every window has been examined, the forty frequencies with the most votes over all the windows are selected as the dominant frequency modes of the model. The initial amplitude and damping coefficients of each frequency are computed by performing a least-squares line fit to log amplitudes of the spectrogram.

Once the model parameters were estimated, a one-second sound sample was generated using the model. The amplitude spectra in Figure 2 compare the recorded sound (on the left), and the sound synthesized using the model (on the right). The estimation technique has correctly identified most of the dominant frequencies. There is some quantization error due to the size of the frequency voting bins. The estimate obtained for the initial amplitudes of each mode is not as good as the frequency estimate, but perceptually, the two sounds are very similar.

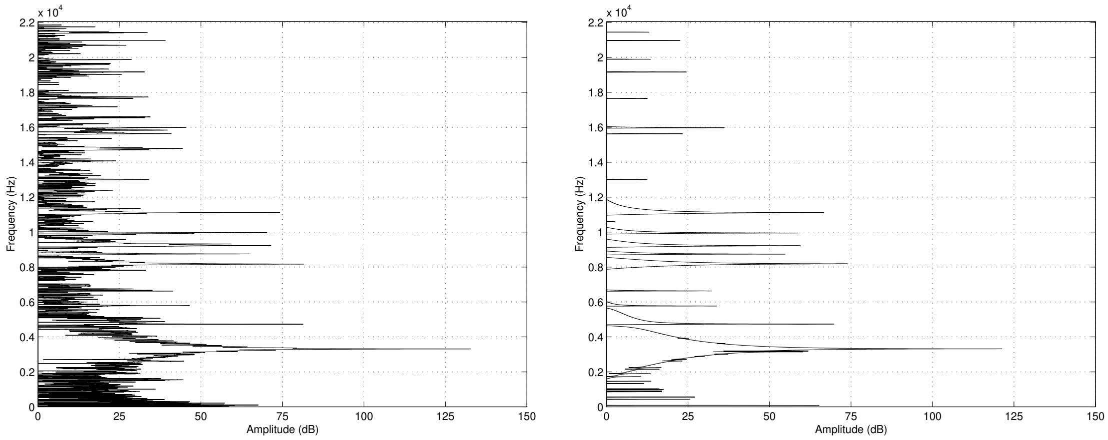
A video segment of the data collection and sound samples is available on our web site: *http://www.cs.ubc.ca/nest/lci/acme/demos/SoundMeasurement.html*. An extended video will be played at the presentation of the paper to allow listeners to compare the recorded and synthesized sounds.

---

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

[3] Sound Blaster 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.



(a) Recorded (b) Synthesized (40 modes)

Figure 2: Comparison of Recorded and Synthesized Amplitude Spectra

FUTURE WORK AND CONCLUSIONS

These preliminary results have led us to focus our future efforts in three areas: developing adaptive sampling algorithms, exploring alternative estimation techniques, and recording the force profile for each actuation of the sound effector,

We are currently investigating the use an adaptive algorithm for selecting sampling locations. An adaptive sampling algorithm will provide a more efficient measurement of objects. The most obvious candidate for an adaptive sampling algorithm is one that adjusts the granularity of the sampling grid based on the differences between sounds measured at neighboring locations. Of course, such a sampling algorithm requires a perceptually relevant metric for comparing two sounds.

In the near future we hope to explore other parameter estimation techniques that will capitalize upon the availability of multiple measurements at each surface location. Using multiple measurements should provide a more robust estimation of the model parameters and greater noise rejection.

To correct the effects of a non-impulsive force, we must record the force profile of each impact. Additional hardware is being considered for the recording of this information.

In this paper, we have presented a system for the automatic measurement of contact sounds. Using this system, a sound model for six locations on the surface of a coffee mug was produced. A spectral comparison of the synthesized sound and the recorded sound was presented as evidence that the sound measurements acquired using the system are adequate for empirical parameter estimation.

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 Perry R. Cook and Dan Trueman. A Database of Measured Musical Instrument Body Radiation Impulse Responses, and Computer Applications for Exploring and Utilizing the Measured Filter Functions. Available Online: *http://www.cs.princeton.edu/ prc/ism98fin.pdf*

3 Robert S. Durst and Eric P. Krotkov. Object Classification from Analysis of Impact Acoustics. In *Proceedings of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pages 90–95, Pittsburgh, 1995

4 W. W. Gaver. *Everyday listening and auditory icons*. PhD thesis, University of California in San Diego, 1988.

5 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.

6 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.

7 Julius O. Smith III. Physical Modeling Synthesis Update *Computer Music Journal*, 20(2):44–56, 1996.

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