Analysis and Synthesis of the Sounds of Impact based on Shape-Invariant Properties of Materials

Eric Krotkov
Robotics Institute
epk@cs.cmu.edu

Roberta Klatzky
Psychology Department
klatzky@cmu.edu

Nina Zumel
Robotics Institute
zumel@cmu.edu

Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

We present a model relating materials to the sounds they make when struck. The model exploits a shape-invariant property of materials: the angle of internal friction. We demonstrate the utility of this model in both analyzing and synthesizing sounds generated by impulsive excitation. For the analysis of impact sounds, we present an active approach for discriminating different materials by impulsively contacting (hitting) them, and sensing and interpreting the resulting sounds. Experimental results suggest that shape-invariance may be encoded in the functional form of the relation between the angle of internal friction and frequency. For the recreation of impact sounds, we identify two key problems—the impact sound reconstruction problem, and the impact sound synthesis problem—and discuss preliminary considerations in developing solutions.

1 Introduction

Computer vision and computer graphics both exploit models of geometry and radiometry. The vision and graphics fields have evolved to view those models from differing perspectives, one from the point of view of image analysis, and the other from the point of view of image synthesis. Despite the difference in perspective, the fields rely on a similar set of underlying models.

Computer audition and computer sound generation need an analogous set of models. For the particular case of speech, these models have already been developed in the speech recognition and speech synthesis fields. However, models for complex sounds are significantly less developed.

In this paper we develop a model relating materials to the sounds they make when struck, and demonstrate its utility in both analysis and synthesis of impact sounds. We restrict our attention to the sounds of impact—sounds generated by impulsive excitation of a material—because such sounds are fundamental, in the sense that many complex sounds are composed from a number of elementary impact sounds.

In the next section on impact sound analysis, we identify related research, present the theoretical framework for a shape-invariant acoustic measure, describe our novel approach to estimating material type, and present and discuss experimental results. In the following brief section on impact sound synthesis, we formulate and state two specific problems in generating sounds of impact. We conclude the paper with a discussion of future research directions and applications.

2 Analysis of Impact Sounds

Consider a metal rod and a wooden rod of the same length and diameter. When you strike the metal rod with your knuckle, it rings; when you strike the wooden rod, it produces a much shorter "thud" sound. This difference in the sound despite the same excitation is due to the difference in the way that the materials vibrate, which in turn is due to stress/strain properties. The rods sound different because they have fundamentally different material properties, so sound waves travel through them quite differently.

Now consider two metal rods that are identical except that one is twice as long as the other. Given the same excitation, the shorter rod will "ring" at a higher frequency than the longer rod. The rods sound different because the waves travel different distances inside them.

How can these differences in the way things sound, one due to material and one due to shape, be resolved? Or, in other words, *What acoustic information is diagnostic of material, but invariant over object shape?*

This fundamental question, and the example that led up to it, concern the sensory modality of audition. And that will be the central topic of this paper. However, we are developing a general approach, applicable to all sensing

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

modalities. Before plunging into the central topic of the paper, we first describe the more general approach.

By definition, a material property is independent of the size and shape of a particular sample. Although there are visual cues to material properties (for example, surface luminance is a cue to the coefficient of friction), reliable determination of the material composition of an unknown object generally requires contact with it. Humans who wish to determine material properties show stereotypical patterns of manual exploration; they press, poke, tap, heft, squeeze, shake, rub, and strike, according to the type of information desired [6]. We are developing a robotic approach analogous to these patterns of human behavior.

In our approach, materials are disambiguated by actively contacting and probing them and by sensing the resulting forces, displacements, and sounds. One can visualize this capability by imagining a game of non-verbal "Twenty Questions," in which one player is the robot and the other player is any object placed in the robot workspace. The robot probes (presses, pokes, taps, etc) the object, in effect asking questions about the object stiffness, density, and other material properties. At the end of the game the robot announces its decision about the material composition of the object.

2.1 Related Research

Compared to the perception of shape or position, perception of material properties is a field in its early infancy. In our previous work, Durst and Krotkov [2] develop a decision-map classifier that achieves good performance for a collection of known objects and materials, but does not address or achieve shape-invariance. Krotkov [5] reports preliminary investigations using a variety of sensory modalities, including force, vision, touch, and audition.

The artificial intelligence, robotics, civil engineering, mechanical engineering, and materials literature documents two families of techniques to estimate mechanical and mass properties, one employing non-contact sensing, the other employing contact sensing. However, no strongly related work appears in this literature, with the exception of Wildes and Richards [10], which is discussed in detail below.

The auditory perception literature typically deals with problems such as the "cocktail party" (attend to one conversation and then switch) and "Prince Shotoku" (attend to several conversations simultaneously) effects [4, 12]. One exception is the work of Warren and Verbrugge [9], which addresses the psychophysics of the perception of breaking and bouncing events. The psycho-acoustics literature dwells only briefly on the topic of perceiving materials from impact sounds [1]. The physics literature [7, 11] addresses the relation between acoustic waves and material properties, but the level of analysis is not appropriate for our effort.

2.2 Theory

Wildes and Richards [10] have advanced a theoretical approach to recovering the material type of an object from the sound generated when it is struck. They restrict their attention to anelastic solids, and study the modulus of compliance as the key to understanding the vibration of the struck object. Following classical analysis they relate the modulus of compliance to the angle of internal friction. This is a shape-invariant property of a given material.

They propose two methods for determining the angle of internal friction of an unknown sample: one that impulsively excites the sample and then measures the acoustic decay rate; another that periodically (say, sinusoidally) excites the sample and then identifies the bandwidth of the acoustic signals. They did not experimentally verify either method, although they did cite supporting evidence from earlier empirical studies [3].

Let us consider the decay rate approach first, because experimentally it is simpler to provide an impulsive excitation than a periodic one. In this method, the angle of internal friction ϕ is determined by the time t_e it takes the amplitude of vibration to decay to $\frac{1}{e}$ of its original value after the material sample is struck. According to Wildes and Richards,

$$\tan \phi = \frac{1}{\pi f \, t_e}, \tag{1}$$

where f is the observed frequency associated with the amplitude. Thus, the problem of determining the angle of internal friction ϕ reduces to determining t_e.

Let θ be a retention parameter representing the proportion of the amplitude present at time t_i that is still present at t_{i+1}. For an exponential process, θ is constant for all i, and the amplitude $A(t)$ at time t is then given by

$$A(t) = A_0 \theta^t ,$$

where A_0 is the initial amplitude.

If the amplitude has decayed to a proportion of $\frac{1}{e}$ of the initial value, then $A(t) = \frac{A_0}{e}$. Note that at this point, $t = t_e$, by definition. So, $\frac{A_0}{e} = A_0 \theta^{t_e}$. Canceling A_0, taking the natural logarithm of both sides, and rearranging leads to

$$t_e = -\frac{1}{\log \theta} . \tag{2}$$

Thus, the problem of determining t_e reduces to determining $\log \theta$.

Assuming an exponential decay process,

$$A(t) = A_0 \theta^t = A_0 e^{-t \log \theta} .$$

Taking logarithms yields $\log A(t) = \log A_0 - t \log \theta$. Thus, the plot of log amplitude against time will be linear with slope equal to $\log\theta$. So, we can determine $\log\theta$ by finding the envelope of the signal waveform as a one-dimensional curve, plotting this curve on a logarithmic scale (this will be linear), and identifying the slope of the plotted line.

Summarizing, in theory we can determine $\log\theta$ from the original waveform, and then determine t_e using (2), and finally determine $\tan\phi$ from (1).

2.3 Approach

We analyze the discrete digital signal $x[n]$ in four main steps: (1) Compute the signal spectrogram. (2) Determine where the contact transient ends, and where the "signal" begins. (3) Find bands of concentrated signal energy; (4) For each band, determine the angle of internal friction.

Spectrogram The spectrogram of a signal describes the distribution of the signal energy in the time-frequency plane. The spectrogram is a popular representation in fields such as speech recognition and acoustic analysis. Formally, the spectrogram $S[l,k]$ is the squared modulus of $X[l,k]$, where

$$X[l,k] = \sum_{n=-\infty}^{+\infty} x[n]g[n-l]e^{-j\frac{2\pi k}{n}l} \qquad (3)$$

is the discrete-time Fourier transform of a windowed version $x[n]g[n-l]$ of the original signal $x[n]$.

We compute the spectrogram by performing the following steps: (1) Split the given signal into $N_{overlap}$ overlapping segments; (2) For each segment, establish a Hanning window of size N_{FFT}; (3) Compute the Fourier transform of each windowed segment.

As an example, Figure 1 shows the spectrogram of the signal produced by striking an aluminum rod, computed with $N_{FFT} = 256$ samples and $N_{overlap} = N_{FFT}/2$. The energy is concentrated in four main bands, at approximately 1200, 2500, 4000, and 6000 Hz. Note that these bands are not harmonics of a common fundamental.

Transient Due to the impulsive contact, the early part of the signal contains energy at all frequencies. This transient effect, which sounds like a click, does not convey meaningful modal information, so we desire to exclude this segment of the signal from analysis. In brief, we determine where the transient ends and signal begins by computing correlations between adjacent temporal windows with respect to their amplitude values, and identifying a dip in the correlation value caused by the transition from the click to the residual excitation of the rod.

Bands Once we have discovered when the signal starts, we then identify those frequency bands with a significant concentration of energy. We accomplish this in five steps: (1) To find the initial location of bands, we begin with a thresholding operation, perform a region growing operation and connected component analysis on the surviving magnitudes. (2) To improve this coarse estimate, we seek the smallest sub-bands that contain a given fraction (currently, 99 percent) of the within-band power. (3) We eliminate those bands that contain only a small percentage of the total power. (4) For each remaining band, we identify the start as the point of maximum power, and the end as the point at which the power has declined to 0.2 percent of the maximum power. (5) For each band, we determine the frequency which contains the most power.

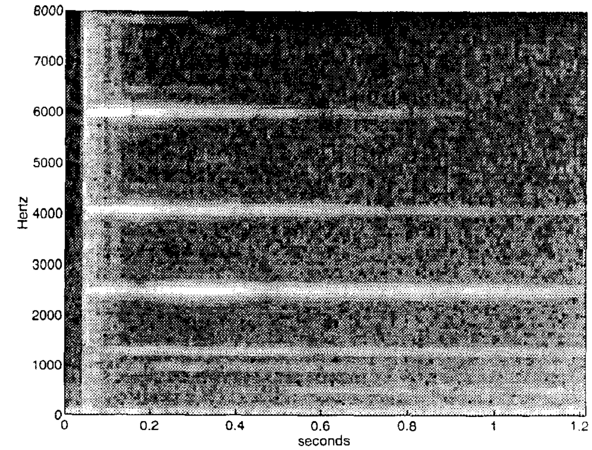

Figure 1. Spectrogram for aluminum rod

In operation, the five processing steps identified three bands when applied to the spectrogram in Figure 1. The band at 1000 Hz was discarded because it contained less than one percent of the total power.

Angle of Internal Friction For each band computed in the previous stage, we fit a line to the within-band log power. At the same time, we compute the goodness (r value) of the linear fit and the length of the line. We filter out those lines with $r < 0.866$ (the fit accounts for 75 percent of the variance) and with length less than 10 time steps. The slope of each line determines the $\log\theta$ term in (2). Now it is possible to determine $\tan\phi$ for each frequency band by substituting (2) into (1).

Figure 2 illustrates the total power associated with the three bands (band 1 is at 6000 Hz, band 2 is at 4000 Hz, and band 3 is 2500 Hz). It also shows the background, defined as the sum of all spectrogram magnitudes that did not pass the threshold test described in the previous section. In addition, the figure shows the lines fit to the three power

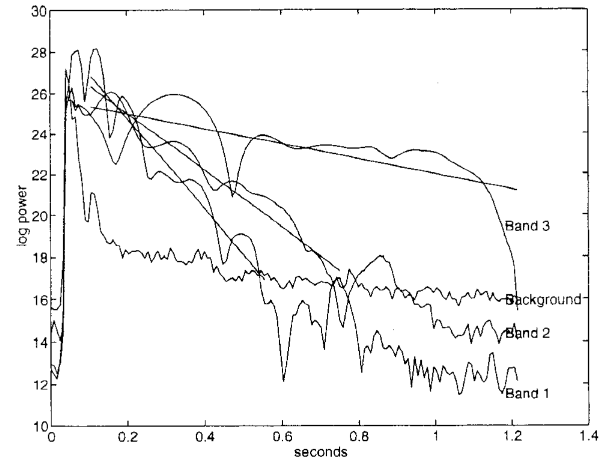

Figure 2. Linear fits to log power within bands

curves. Bands 1 and 2 are reasonably linear (r values above 0.95, that is, the fit accounts for more than 90 percent of the variance), and band 3 is not (r value of 0.70).

Experiments To assess the validity of the decay rate approach to identifying the angle of internal friction, we produced thin rods of wood, brass, aluminum, glass, and plastic. For each material, we produced two rods, one of length $L = 15$ cm and one of length $2L$.

We suspended each rod by string from above, and struck it with a well-damped solid object. We used an electret condenser microphone, and fed the signal to an analog/digital converter installed on a Macintosh workstation, operating at sampling rate of $f_{sample} = 22$ kHz.

The acquired data relate $\pi \tan \phi$ and the frequency of greatest power. Each data point represents a single band; a given trial may provide from 1–4 points. We fit quadratic functions of the form

$$\pi \tan \phi = a_2 f^2 + a_1 f + a_0$$

to the data points for each material, finding the a_i coefficients minimizing the squared error. The fitting procedure combined data from short and long samples, with the exception of brass, where the short rod produced a single frequency with an associated value of $\tan \phi$ that was anomalous with respect to the function obtained with the long rod.

The salient features of the results are the following:

1. The variability associated with a single frequency tends to be reasonably small, relative to the variability across frequencies, suggesting some stability to the estimated $\tan \phi$ at a frequency.

2. The $\tan \phi$ values exhibit clear variation across frequencies. Hence different samples, which produce bands at different frequencies, will give different distributions of $\tan \phi$.

3. With the exception of brass, it appears that the relation between $\tan \phi$ and frequency can feasibly be fit by a single quadratic, with r^2 ranging from 0.35 to 0.84 in the present data.

4. Whether the quadratic is concave upward or downward varies with the material.

Discussion On the basis of the theory presented in Section 2.2, we expected the value of $\tan \phi$ to vary with material type and to remain constant despite changes in the length of the material sample. We find that the angle of internal friction is not invariant over the frequency of the sound spectrum for which it is estimated. We are surprised to observe the frequency dependence of the $\tan \phi$ values, since the theory does not predict it. This frequency dependence suggests that the shape invariance may be encoded in the functional form of the relation between $\tan \phi$ and frequency. Further, the surprising results indicate the need for a new theory.

3 Synthesis of Impact Sounds

In this section we formulate and state two specific problems in creating an acoustic signal representing a realistic impact sound. Just as computer graphics is the dual of computer vision, so is this general problem the dual of the analysis problem studied in Section 2. We defer solutions to these problems to a future paper.

The literature on audio signal synthesis is dominated by work on speech signals. Our literature review did not reveal prior work addressing synthesis of contact sounds. However, the speech literature has dealt to some extent with the problem of synthesis from limited spectral information. Shannon et al. [8] observe nearly perfect speech recognition under conditions of greatly reduced spectral information, namely, three bands of modulated noise. They conclude that the presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech. We observe that if such limited information suffices for the recognition of speech, it seems plausible that it may also suffice for the recognition of impact sounds.

3.1 Impact Sound Reconstruction Problem

The *impact sound reconstruction problem* can be stated as follows: Given the spectrogram of a signal representing an impact sound, such as that shown in Figure 1; Reconstruct the impact sound signal.

Before stating our solution to this problem, we recall that expression (3) computes X, the discrete-time Fourier transform of a windowed version of the original signal x. If there are N samples in the original time series signal, then there are $N_c = N/N_{FFT}$ samples (columns) in its spectrogram. Given X, an algorithm for reconstructing the acoustic signal can be stated as follows. For each column of the spectrogram ($c = 1, 2, \ldots, N_c$), perform the three following steps:

1. Take the inverse Fourier transform $y[n] = \mathcal{F}^{-1}(X[l, k])$.

2. Compensate the resulting time series for window effects. If the forward Fourier transform uses the Hanning window of length N_{FFT}

$$H[k] = \frac{1}{2}\left[1 - cos\left(\frac{2\pi k}{N_{FFT} - 1}\right)\right] ,$$

then "un-window" the time series by

$$y^*[k] = \frac{y[k]}{H[k]} , \; k = 1, 2, \ldots, N_{FFT} .$$

3. Accumulate data segments. Append the time series of length N_{FFT} created from the c-th column of the spectrogram to the reconstructed time series y' by

$$y'[N_{FFT} \times (c - 1) + k] = y^*[k] \;\; , \;\; k = 1, 2, \ldots, N_{FFT} .$$

By construction, this accumulates $N_{FFT} \times N_c = N$ samples in y', as required.

We verified this approach to reconstructing impact sounds by applying steps 1–3 to the spectrograms computed from recorded impact sounds. We found that the correlation between the original and reconstructed signals approached unity, and conclude that there is virtually no difference between the two signals.

3.2 Impact Sound Synthesis Problem

The *impact sound synthesis problem* can be stated as follows: Given (1) a set $\{(\tan\phi_i, f_i, \epsilon_i)\}$ specifying triples of internal friction parameters, center frequencies, and band widths, (2) the shape of a specimen, and (3) an impact force model; Synthesize a realistic impact sound.

This problem differs from the impact sound reconstruction problem in starting with no spectrogram data. Instead, it starts with the angle of internal friction. This quantity determines the amplitude decay over time, but does not directly provide phase information. Without phase, it is not possible to generate a unique acoustic signal. Thus, one of the key challenges is to identify the signal phase.

4 Future Work

We hope to apply the insights gained by this investigation in at least two fields of application. In the field of non-destructive evaluation, we will develop new materials testing procedures based on quantitative measurement of the angle of internal friction. In the field of virtual reality, we will develop low-bandwidth representations of real-world sounds that can be used to create multimodal events.

A great deal of work remains before such applications are feasible. Future work will concentrate on implementing methodological insights in a new experimental setup including more repeatable striking mechanisms and faster sampling devices. Future work will also expand the inquiry to encompass more thorough study of shape effects, beginning with variable length rods and extending to plates, solids, and irregularly formed objects.

In the more distant future, alternative measures and approaches need to be pursued before the new field of material-based acoustic modeling emerges from its early infancy, and contributes productively to our understanding of what sounds tell us about the real world.

References

[1] A. S. Bregman. *Auditory Scene Analysis*. MIT Press, Cambridge, Massachusetts, 1990.

[2] R. Durst and E. Krotkov. Object classification from analysis of impact acoustics. In *Proc. Intl. Conf. Intelligent Robots and Systems (IROS)*, pages 90–95, Pittsburgh, Pennsylvania, August 1995.

[3] A. Gemant and W. Jackson. The measurement of internal friction in some solid materials. *Philosophical Magazine*, 157:960–983, 1937.

[4] S. Handel. *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, Cambridge, Massachusetts, 1989.

[5] E. Krotkov. Perception of material properties by robotic probing: Preliminary investigations. In *Proc. IJCAI*, pages 88–94, Montreal, August 1995.

[6] S. Lederman and R. Klatzky. Hand movements: A window into haptic object recognition. *Cognitive Psychology*, 19:342–368, 1987.

[7] T. Rossing and D. Russell. Laboratory observation of elastic waves in solids. *Am. J. Physics*, 58(12):1153–1162, December 1990.

[8] R. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. *Science*, 270:303–304, October 13 1995.

[9] W. Warren and R. Verbrugge. Auditory perception of breaking and bouncing events: Psychophysics. In W. Richards, editor, *Natural Computation*. MIT Press, Cambridge, Massachusetts, 1988.

[10] R. Wildes and W. Richards. Recovering material properties from sound. In W. Richards, editor, *Natural Computation*, pages 357–363. MIT Press, Cambridge, Massachusetts, 1988.

[11] J. Wolfe. Acoustic wavefronts in crystalline solids. *Physics Today*, pages 34–40, September 1995.

[12] W. Yost and C. Watson, editors. *Auditory Processing of Complex Sounds*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1987.