Robotic Perception of Material: Experiments with Shape-Invariant Acoustic Measures of Material Type

Eric Krotkov   Roberta Klatzky   Nina Zumel

Robotics Institute   Psychology Department   Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, Pennsylvania   15213

Abstract

We present an active approach for discriminating different materials by impulsively contacting (hitting) them, and sensing and interpreting the resulting sounds. In theory, the angle of internal friction is diagnostic of material, but invariant over object shape. In our experiments, we observe that the angle of internal friction is not invariant over the frequency of the sound spectrum for which it is estimated. Hence, samples of different shapes, which exhibit power concentration at different frequencies, exhibit different values of the angle of internal friction. However, the results suggest that shape-invariance may be encoded in the functional form of the relation between the angle of internal friction and frequency.

1. Introduction

Consider a metal rod and a wooden rod of the same length. 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 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 [3]. 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.

The capability to perceive material has many potential applications. In general, knowledge of material properties and classes can improve performance of many tasks involving physical interaction. In many real-world scenarios, such knowledge is not given in advance; instead, it must be determined at a worksite or in the field, without jigs or fixtures. Specific applications include grasping, non-destructive evaluation and inspection, reasoning about functionality, handling hazardous waste, recycling, excavating, and traversing natural terrain.

In this paper, we describe a shape-invariant measure of material type, derived through acoustic sensing, and present results of experiments that confirm theoretical predications that the measure is diagnostic of material type. We begin by briefly referring to related research. We present the theoretical framework for the shape-invariant acoustic measure in Section 3. Next, we describe our novel approach to estimating material type in Section 4, and we present experimental results in Section 5. We conclude the paper with a critical discussion of progress to date.

2. Related Research

Compared to the perception of shape or position, perception of material properties is a field in its early infancy.

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. We explore this literature in detail elsewhere [2].

3. Theory

Wildes and Richards [4] 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 [1].

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).

4. 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.

4.1. 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

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

is the discrete-time Fourier transform of a windowed version $x[n]g[n - l]$ of the original signal $x[n]$.
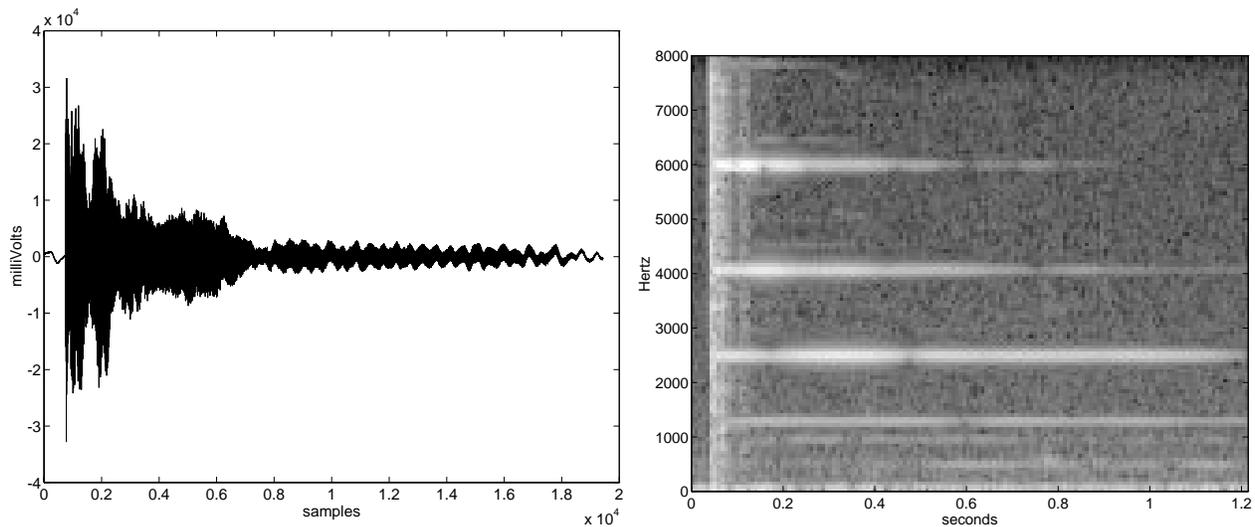
Figure 1. Microphone output after striking aluminum rod (left), spectrogram (right)

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 original signal recorded at $f_{sample} = 16$ kHz after striking an aluminum rod. The figure also shows the spectrogram of that signal, 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.

4.2. 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.

For this, we examine correlations between the spectrogram magnitudes at adjacent temporal windows, computed across frequencies. During the click, the spectrogram magnitudes are highly correlated from instant to instant. Immediately after the click, as energy begins to concentrate in relatively narrow bands, the spectrogram frequencies with high magnitudes are not the same as those in the click. This causes the correlation between adjacent windows to dip. Well after the click, the spectrogram magnitudes are again highly correlated. Based on these observations, we determine where the signal begins by identifying when the correlation coefficients rise from the dip caused by the transition from the click to the residual excitation of the rod.

Let $S[t]$ be a vector of length $N_{FFT}/2$ representing the spectrogram magnitudes at time t. In effect, this is a column in the spectrogram shown in Figure 1. For all pairs of time-adjacent vectors, we compute the correlation coefficient $\rho_t = corr(S[t], S[t+1])$. We search for the time at which ρ_t takes on its globally minimum value. We then search for the time, following that, at which ρ_t takes on a locally maximum value. We treat this as the end of the transient. Figure 2 illustrates the time selected as the start of the signal.

4.3. 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) Find initial location of bands; (2) Refine initial location estimates; (3) Filter out weak bands; (4) For
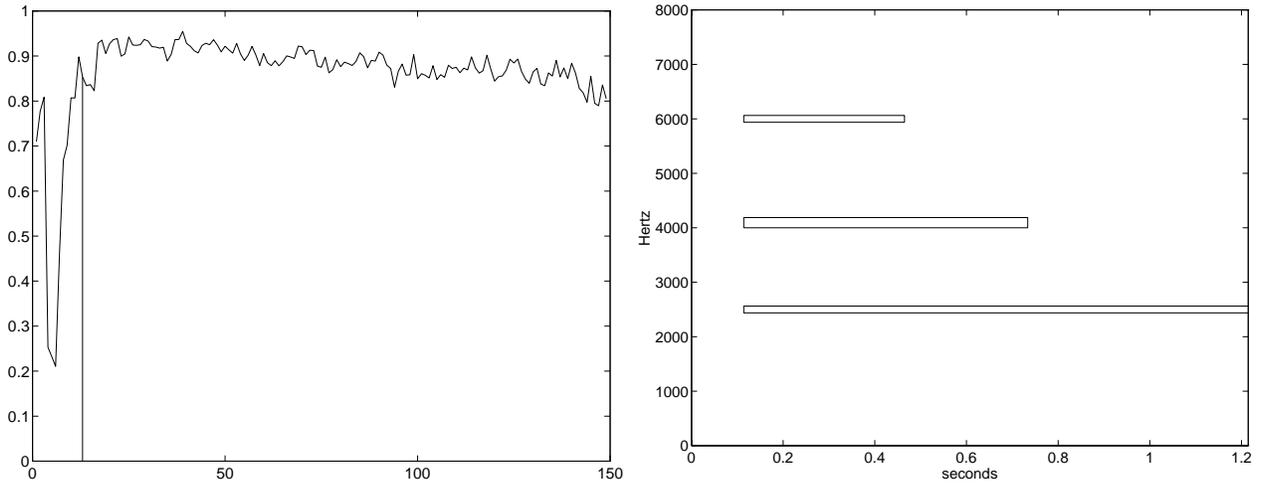
Figure 2. Correlation versus time for data in Figure 1 (left), spectral bands (right)

each remaining band, determine when signal starts and ends; (5) For each remaining signal in a band, determine the frequency with greatest power.

First, to find the initial location of bands, we begin with a thresholding operation: For each time step, we mark as background those spectrogram magnitudes that contribute little to the total power at that time. In the current implementation, the threshold value is set to be the power we would expect if the energy was distributed uniformly across all frequencies. Next, we perform a region growing operation and connected component analysis on the surviving magnitudes. This results in a coarse estimate of the band locations.

Second, to improve this coarse estimate, we seek the smallest sub-bands that contain a given fraction (currently, 99 percent) of the within-band power. This produces bands that are significantly more focussed.

Third, we eliminate those bands which contain significantly less power than the others. In the current implementation, we delete all bands that contribute less than 1 percent of the total power.

Fourth, for each remaining band, we identify when the signal starts and ends. The start is the point of maximum power, and the end is the point at which the power has declined to 0.2 percent of the maximum power.

Finally, for each band, we determine the frequency which contains the most power. We store this value to be used in the computation of $\tan \phi$.

Figure 2 illustrates the result of these five processing steps applied to the spectrogram in Figure 1. The algorithm identified three bands. The band at 1000 Hz was discarded because it contained less than one percent of the total power.

4.4. 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 3 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
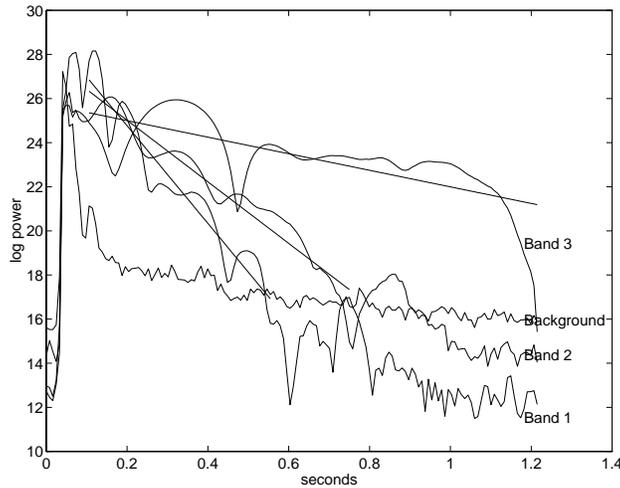
Figure 3. Linear fits to log power within bands

the previous section. In addition, the figure shows the lines fit to the three power 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).

5. 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. We struck the rods with "found" objects, including a soldering pencil stand (selected because its vibrations damped out more rapidly than did the other objects we tried) and the plastic handle of a screwdriver. 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.

Figure 4 plots the observed data relating $\pi \tan \phi$ and the frequency of greatest power (the π was omitted from the figure labels). Each point represents a single band; a given trial may provide from 1–4 points. We fit quadratic functions $\pi \tan \phi = a_2 f^2 + a_1 f + a_0$ 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, for which the single frequency available for the short rod appears anomalous with respect to the function obtained with the long rod.

In summary, the salient features of these graphs 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.

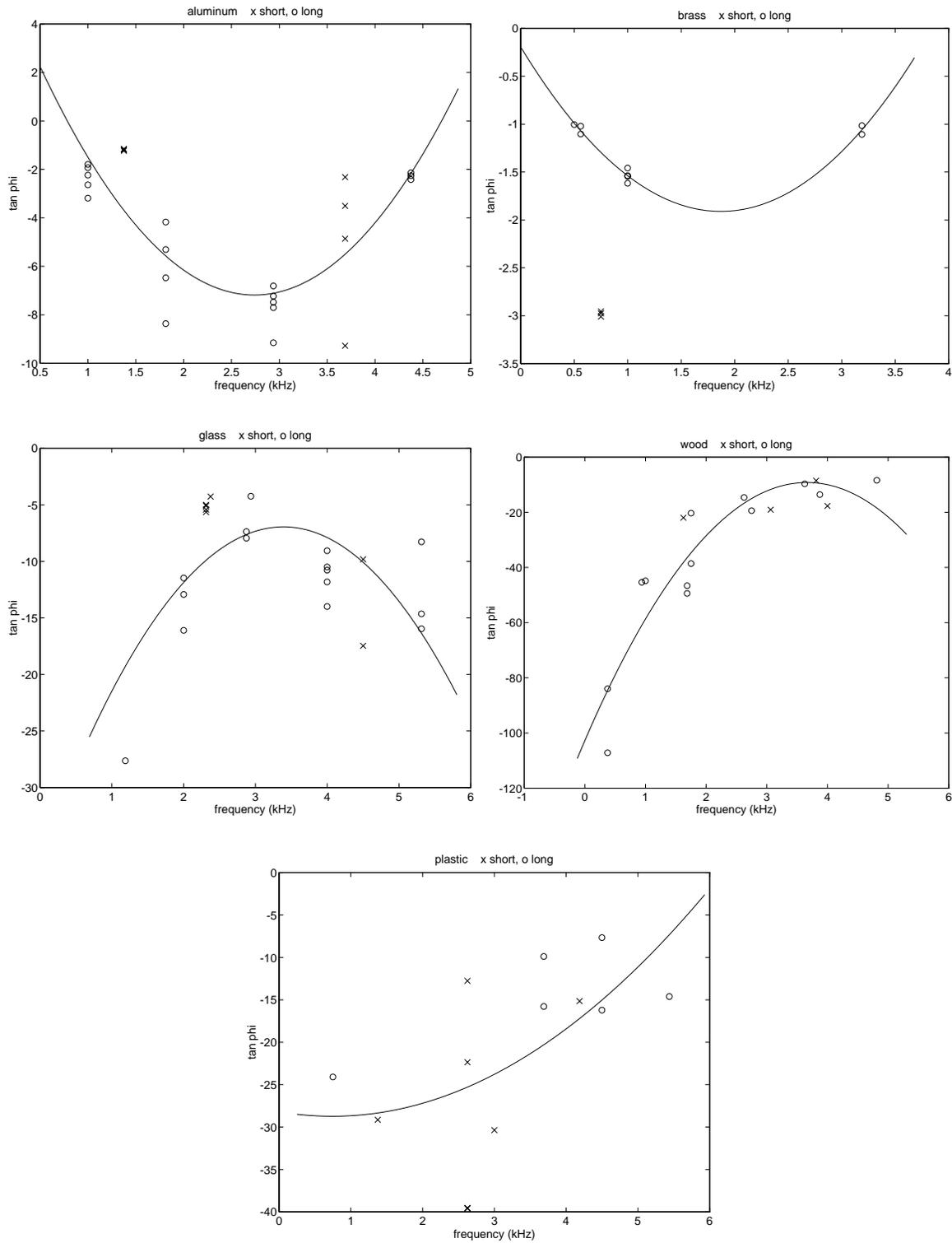Table 1 tabulates the parameters of the quadratic. The R^2 term represents the variance

Figure 4. Experimental results

accounted for by the quadratic function, which we interpret as the goodness of the fit.

6. Discussion

On the basis of the theory presented in Section 3, 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

Material	a_2	a_1	a_0	r^2
Aluminum	1.88	-10.29	6.92	0.61
Brass (long)	0.49	-1.84	-0.19	0.96
Glass	-2.54	17.24	-36.16	0.38
Wood	-6.97	51.09	-102.76	0.84
Plastic	0.97	-1.44	-28.19	0.35

Table 1. Parameters of quadratic function $\pi \tan \phi = a_2 f^2 + a_1 f + a_0$

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.

Our experience during the course of the experiments has yielded a number of methodological insights that will influence future investigations. The first observation is that the experimental method for suspending the sample is more important than we first believed. In particular, the motion of the sample after striking must be taken into account or otherwise "matched." The second observation is that a higher sampling rate is needed to acquire reliable measurements from non-metals. Finally, the spectrogram appears to be a very powerful representation that is well-suited to the discrimination task at hand.

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 these 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 perception of material emerges from its early infancy, and realizes its potential for revolutionizing robotic interaction with the real world.

Acknowledgments

References

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

[2] E. Krotkov. Perception of Material Properties by Robotic Probing: Preliminary Investigations. In *Proc. IJCAI*, pages 88–94, Montreal, August 1995.

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

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