Perception of Material from Contact Sounds

Roberta L. Klatzky[1], Dinesh K. Pai[2], & Eric P. Krotkov[3]

[1] Dept. of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213
klatzky@cmu.edu
[2] Dept. of Computer Science, U. British Columbia, Vancouver, B.C. V6T 1Z4, Canada
pai@cs.ubc.ca
[3] Cytometrics Inc., Philadelphia, PA
krotkov@cytometrics.com

To appear in Presence, The MIT Press, 2000.

Authors' note: The authors wish to thank Constantine Nicolaides and Jesse Kates for their help with data collection and analysis.

Perception of Material from Contact Sounds

Abstract

Contact sounds can provide important perceptual cues in virtual environments. We investigated the relation between material perception and variables that govern synthesis of contact sounds. A shape-invariant auditory-decay parameter was a powerful determinant of the perceived material of an object. Subjects judged the similarity of synthesized sounds with respect to material (Experiment 1 and 2) or length (Experiment 3). The sounds corresponded to modal frequencies of clamped bars struck at an intermediate point. They varied in fundamental frequency and frequency-dependent rate of decay. The latter parameter has been proposed as reflecting a shape-invariant material property: damping. Differences between sounds in both decay and frequency affected similarity judgments (magnitude of similarity and judgment duration), with decay playing a substantially larger role. Experiment 2, which varied the initial sound amplitude, showed that decay rate, rather than total energy or sound duration, was the critical factor in determining similarity. Experiment 3 demonstrated that similarity judgments in the first two studies were specific to instructions to judge material. Experiment 4, in which subjects assigned the sounds to one of four material categories, showed an influence of frequency and decay but confirmed the greater importance of decay. Decay parameters associated with each category were estimated and found to correlate with physical measures of damping. The results support use of a simplified model of material in virtual auditory environments.

The fundamental goal of VR technology is to provide the feeling of presence within an interactive environment. Full presence requires a multimodal surrounding, including not just vision but audition and touch. Multimodal information is particularly important for simulated contact with objects, an event that in the real world provides substantial perceptual information through the haptic and auditory modalities. To date, these modalities have been largely neglected relative to vision in the development of simulated enviroments. This circumstance is unfortunate, when one considers that sounds provide important cues to event onset at low latency, complement haptic cues to contact, and can modulate or even override visual cues (Massaro, 1987; McGurk & MacDonald, 1976).

Simulating interactions between objects (one of which may be the hand) requires not just conveying the onset of a contact event, but providing information about the objects and how they interact with one another -- for example, their shape and material, the force of impact, and the location of contact relative to object geometry. In this paper we concentrate on the cues that signal an object's physical material from the sound generated by contact. Our approach is an ecological one, in that we quantitatively manipulate cues inherent in sound and investigate how they map onto perceptual judgments.

One of the important barriers to designing auditory displays that convey contact sounds is our lack of knowledge about how the human auditory system processes such information. Although there has been relevant psychophysical work (see below), we lack explicit prescriptive models that would provide guidelines for synthesizing sounds. We are also unaware of experiments that quantify the relative importance of cues to object interaction. An alternative approach to sound simulation would be to record a large number of impact sounds and provide libraries; however, the number of samples becomes computationally intractable when one considers that sounds vary not only with the object properties but with the forces applied and locations of contact (see van den Doel & Pai, 1998).

This paper provides at least an initial step toward modeling the perception of object material from properties of sound. The results can be used for real-time synthesis of impact events in virtual environments. We seek to identify a shape-invariant parameter of sound that leads to the perception that the sound emanates from impact with a particular material. The parameter we identify is one of a number of potential cues to material, not all of which are shape-invariant. We next describe those properties and the one on which we focus.

Potential auditory cues to material. In general, material properties influence sound in complex ways. From a physical point of view, perhaps the most important material property is elasticity (or stiffness). For isotropic linear materials, this is characterized by Young's modulus and Poisson's ratio. Elasticity is directly related to the speed of sound in the material and therefore influences all aspects of sound production, including frequency. In general, the modal frequencies of an object depend on its elasticity, density, shape, and boundary conditions. If other factors are fixed, frequency increases with stiffness.

Another, lesser known, property of material is damping, or the internal coefficient of friction. This influences how the sounds decay over time. It is the shape-invariant parameter of interest here.

The state of stress of a material is another influence on the sound, as is clear to anyone who has tuned a guitar. Many other aspects of material could influence sound, such as grain pattern, method of fabrication, porosity, or age.

Turning to the present focus, material damping, as determined by the internal coefficient of friction, is an intrinsic property of material that measures its anelasticity. The parameter varies from one material to another; damping values for some materials can be found in Wildes and Richards (1980). Specifically, the sound field at some distance to a struck object can be approximated, under fairly general assumptions, as a combination of decaying sinusoids (complex exponentials), corresponding to the eigenmodes of the vibrating object. The amplitude of each component is proportional to

$$\exp(- b(f) * t) * \sin(2 * \pi * f * t)$$

where f is the frequency of the mode, b(f) is a frequency-dependent damping component, and t is time. The nature of b depends on the material model.

Wildes and Richards (1980) proposed, based on the work of Zener (1948) and others on anelastic behavior of metals and crystalline solids, that b(f) depends linearly on the frequency.

This model was used by Takala and Hahn (1992) and van den Doel and Pai (1998) to synthesize decaying sounds of impact, with results that informally appeared to convey realistic differences in material, particularly with ringing sounds. As shown by Chaigne and Doutaut (1997), using a model viscoelastic dissipation adds an additional quadratic term. This was also used by Gaver (1988), who empirically modeled observed damping with a general quadratic polynomial. Finally, a constant term in b(f) indicates an overall, frequency-independent decay, for instance due to fluid damping; however, the contribution of this term is usually small. (In Gaver's experiments, objects were placed on a carpet, which may have contributed additional damping.)

In this work we use the linear model, in part because of the physical relevance to metals and crystalline solids, and in part because a linear dependence on frequency is the simplest possible model. In fact, the experimental data reported by Gaver (1988) and by Chaigne and Doutaut (1997), with the exception of two or three data points at the highest frequencies, appear to fit a linear model very well. Occam's Razor would suggest the use of the simplest possible model, specifically, $b(f) = \pi * \tan(\phi) * f$, where $\tan(\phi)$ is the internal friction coefficient. Wildes and Richards (1980) proposed that this parameter could be recovered from impact sounds, for a single frequency, by measuring the time required for the sound to decay a constant proportion. Since the decay time is dependent on frequency, the decay equation took the following form:

$$t_e = 1/(\pi * f * \tan(\phi)) \qquad \text{(Equation 1)}$$

where t_e is the time for the sound to decay by a proportion 1/e, f is the frequency, and $\tan(\phi)$ is the internal coefficient of friction. We denote by τ_d a frequency-independent parameter, such that:

$$\tau_d = 1/(\pi * \tan(\phi)) \qquad \text{(Equation 2)}$$

It is conjectured that $\tan(\phi)$ or alternatively, τ_d, are shape-independent material properties. Krotkov, Klatzky, and Zumel (1996) provided one experimental test of the constancy of $\tan(\phi)$ across frequencies, but their data should be regarded as preliminary.

This model was used by van den Doel and Pai (1998) to synthesize decaying sounds of impact. They computed the modes of vibration for an object of a specific shape that is struck at a particular impact site, and used the above equations, with an assumption of the value of τ_d, to implement decay at each mode. Whereas the decay parameter τ_d is assumed by the model to be shape- and strike-point invariant, the modal frequencies and the amplitude spectrum are affected by those factors. This paper provides a test of the perceptual relevance of the decay parameter τ_d and compares the relative contributions to material perception of decay and frequency.

Related work. A wide variety of psychophysical work has demonstrated the sensitivity of human perceivers to attributes of the event that produced a sound. The most extensive body of research, generally covered in introductory texts on hearing, has been on spatial attributes of the event, such as cues that convey the distance and azimuth of the auditory source from the observer (see, e.g., Coleman, 1963; Green, 1976; Yost & Hafter, 1987). Another relevant body of work concerns the multidimensional scaling of real or synthetic musical instruments by their timbre (e.g., Grey, 1977; McAdams, Winsberg, Donnadieu, DeSoete, & Krimphoff, 1995; Miller & Carterette, 1975; Plomp, 1970). Other research concerns the ability to perceive the nature of the interaction between two objects; such as whether a dropped glass bottle broke or bounced (Warren & Verbrugge, 1984). Along these same lines, Repp (1987) found that the configuration of the two hands during clapping was identifiable above a chance level on the basis of spectral properties of the sound; for example, the low frequency peak represented the resonance between the two palms.

A relatively small body of work concerns people's ability to judge the material or geometric attributes of the object(s) that produced a sound. Gaver (1988) was the first to conduct such experiments. His experiments have many features in common with ours. He studied two physical attributes of the object that produced a sound, namely, material and length, using recorded and synthesized sounds of struck bars. Material changes included not only damping but also Young's modulus and density (which affect the pitch of the sounds). Two experiments used a material discrimination task, which produced ceiling-level performance, as it required only discrimination between wood and metal. A multidimensional scaling analysis of synthesized sounds was also conducted, based on judgments of the overall similarity of sounds. Material was found to account for more variance than length, understandably given the discriminability of the materials. Our experiments examine more closely the perception of material and the relative contributions of internal friction and frequency to material categories. As in Gaver's work, both scaling and classification approaches were used. Our scaling task used directed similarity judgments (similarity of material vs. length), and the classification task required a finer partitioning than simply between wood and metal, allowing us to relate the results to the damping parameter associated with various materials.

In other work dealing with the perception of object attributes from sound, Freed (1990) found that the perceived hardness of a mallet striking a metal object was predictable from acoustical parameters of the attack portion of the sound. Lakatos, McAdams, & Caussé (1997) investigated how well listeners could discriminate the geometric shape of the cross sections of two bars. Discrimination performance depended on the difference between the bars' width:height ratio, which in turn was related to their torsional vibrational and transverse bending modes. Carello, Anderson, and Kunkler-Peck (1998) investigated people's ability to judge the length of an object from a recording of the sound when it was dropped onto a hard surface. Judged length was related to physical length but was not reliably related to the acoustic variables of duration, amplitude, and frequency. The authors suggested that rotational inertia could be the basis for judging rod length, but it is not yet apparent what acoustic variable(s) mediate the judgment. In all of these studies, the subjects listened to real events, properties of which were determined by acoustic analysis (and, in the case of Lakatos et al., accelerometer measurements of the bars themselves). Our own work differs in that we began with theoretically relevant properties of sound and synthesized samples that vared along the relevant acoustic dimensions.

Experiments 1-3

Method

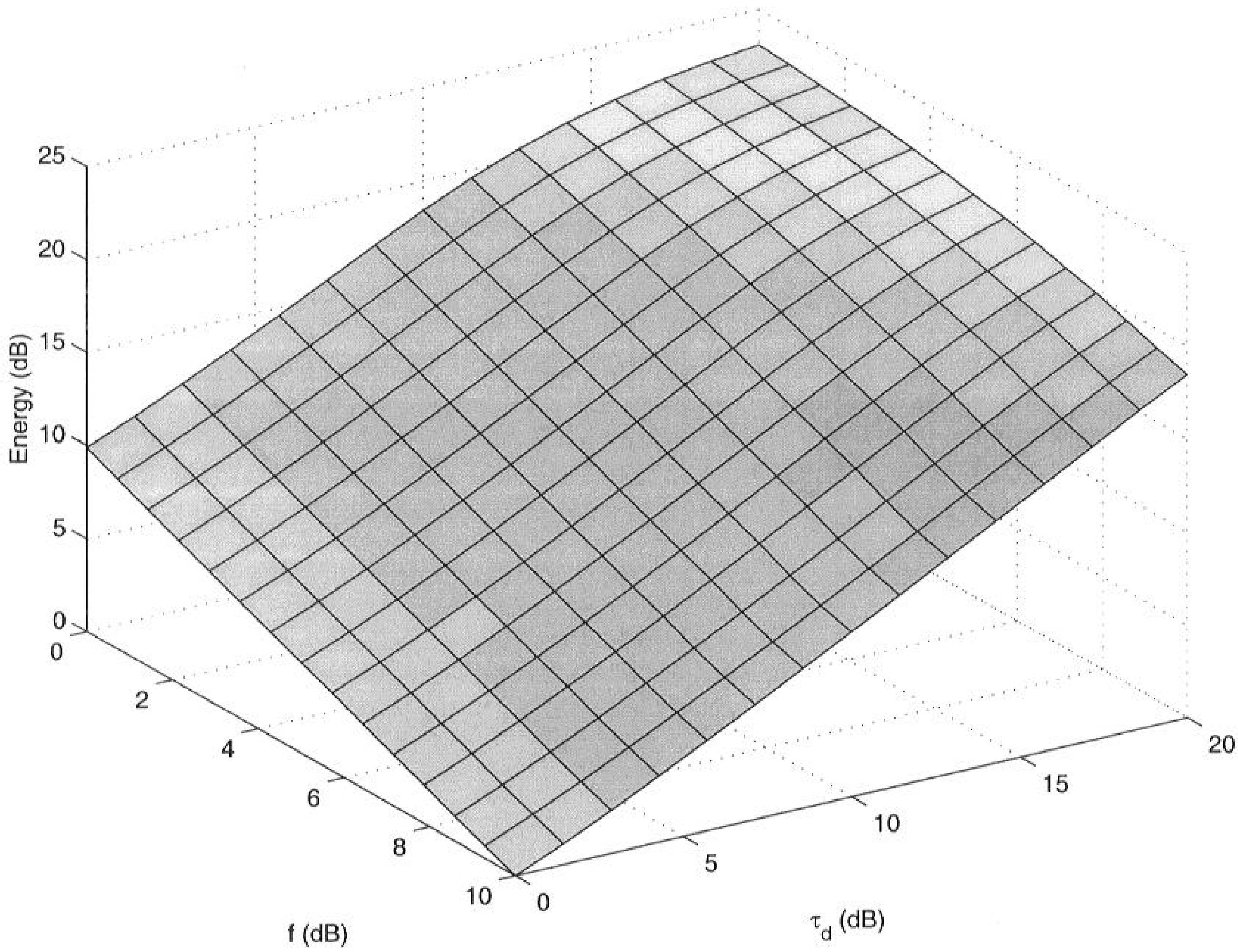
Subjects. University students participated as part of a course requirement or for pay. There were 13 subjects in Experiments 1 and 3 and 14 in Experiment 2; each student participated in one experiment only. All reported normal hearing. Subjects were naive with respect to the purpose of the study and were not selected to have musical experience or experience with virtual auditory technology.

Stimuli. The stimuli were 25 synthesized sounds, combined into all possible pairs (N = 300 pairs). Each sound was generated to correspond to an ideal bar, clamped at both ends, and struck at a point 0.61 of its total length. The bar shape and strike point generate a set of modal vibratory frequencies relative to a fundamental. Five fundamental frequencies at equal log intervals from 100 to 1000 Hz, representing different lengths of bar, were used: approximately 100, 178, 316, 562, and 1000 Hz (roughly the notes G2, F3, E-flat4, D-flat5, and C6). These frequencies are within fairly flat regions of the iso-loudness contour (i.e., perceived intensity is roughly constant for a given dB level). While bar sounds contain several inharmonic frequencies, we will refer to each sound by the fundamental, which we will call "frequency."

As described above, the decay of the sounds, assumed to give rise to the impression of a material, was regulated by an exponential rate parameter, τ_d. As shown in Equation 2, this parameter is theoretically equal to $1/(\pi * \tan(\phi))$. The time for the amplitude of the sound at any one frequency component (f) to decay to 1/e of its current value is τ_d/f, where f is the frequency of a mode of vibration. The value of τ_d was approximately 3, 9, 30, 95, or 300, constituting equal log intervals from 3 to 300.

In Experiments 1 and 2, on each trial, the subjects heard a pair of sounds and rated their similarity, in terms of the strength of their feeling that the sounds could come from the same material, regardless of shape. To clarify the meaning of material, the instructions included, "You probably know that when you hit a string, how high or low a note it produces depends on its length and tension. We are not interested in how high or low the note is, but rather what type of material might produce it." In Experiment 1, each sound was simulated as if the strike force was constant, so that the initial amplitude of the waveform was equalized across sounds. However, sounds with faster decay (τ_d) and higher fundamental frequency had (a) shorter duration and (b) lower total energy. The relation of energy to these parameters is shown in Figure 1. In Experiment 2, in order to control for the covariation of decay and energy, the initial amplitude of a given sound was varied according to its frequency and intensity, so as to constrain the total energy to within a range of 10 dB relative to the minimum value. The initial amplitude simulated for a given frequency and decay was randomly selected within these constraints.

Figure 1. Relation of total energy in a sound to fundamental frequency and decay.

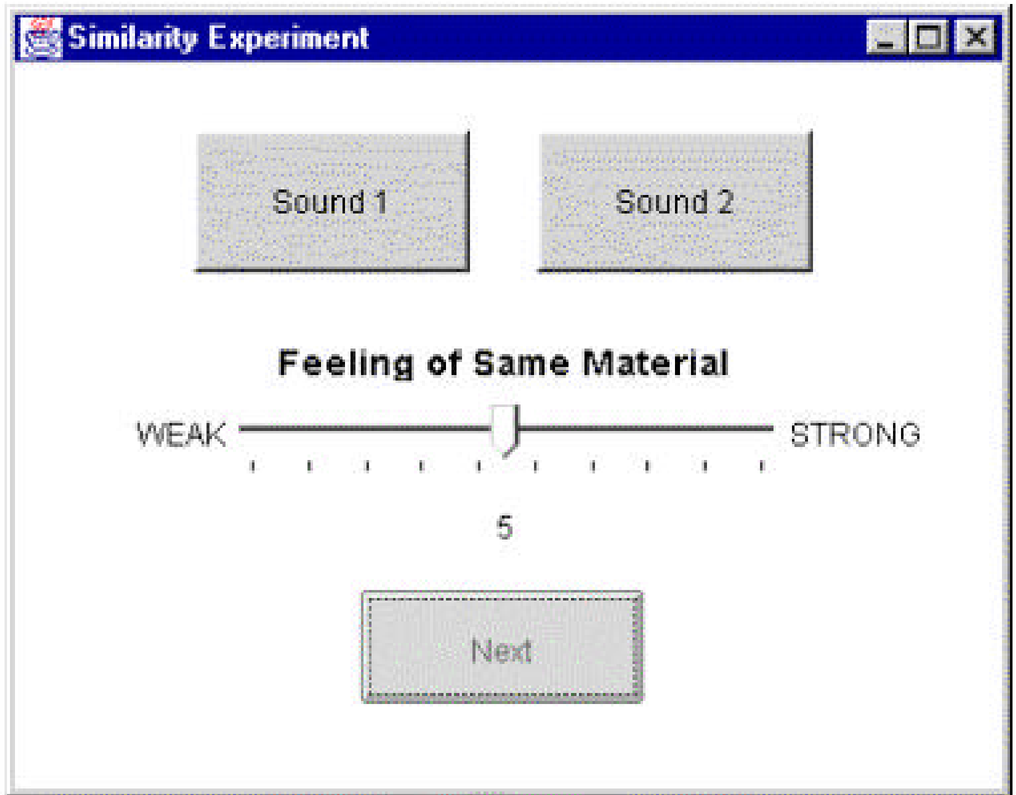


Experiment 3 was conducted as a control, to determine whether subjects were simply responding to the underlying parameters that defined the stimulus set, without trying to judge material. It used the same stimuli and task as Experiment 2, but with instructions that asked subjects to judge the length of the bar. To directly parallel the previous instructions, the subjects were told, "You probably know that when you hit an object of some length, how long the sound lasts depends on its material. We are not interested in how long the sound lasts, but rather what length of bar might produce it." As in Experiment 2, the sound's actual duration depended on fundamental frequency, decay (τ_d), and initial amplitude.

Software environment and Procedure. The test sound samples were computed in Matlab, using the algorithm of van den Doel and Pai (1998) and saved as a Microsoft ".wav" file. The sounds were sampled at 44.1KHz sampling rate, with 16-bit resolution (CD quality).

The experiment was controlled by a Java (JDK 1.1) program running on a Windows95 PC. The program made use of the Java Media Framework API, which allowed rendering of CD quality sounds. The subject interacted with the experiment using a graphical user interface (see Figure 2). Each trial presented one pair of sounds. The subject was allowed to hear each sound in the pair as frequently as desired by pointing and clicking. Volume control was set (constant across all sounds and subjects) so that sound onsets were at a comfortable, audible level. Clicking on a sound location caused the sound to be emitted through headphones (Koss Optimus Pro 4AA) for 500 ms, at which point the sound was truncated (without gradual fading). The subject rated similarity by moving a slide along a scale marked from 1 to 10, with the ends labeled "weak" and "strong". The final slide position was measured by the program in 10ths of marked units, for a total of 100 similarity values. Once the subject began moving the slide, he or she could not resample the sounds. On each trial, the computer measured the response scale position (similarity value) and the number of times each tone was sampled.

Figure 2. Graphical interface; the subject sampled the sounds as often as desired before adjusting the slide to indicate similarity.

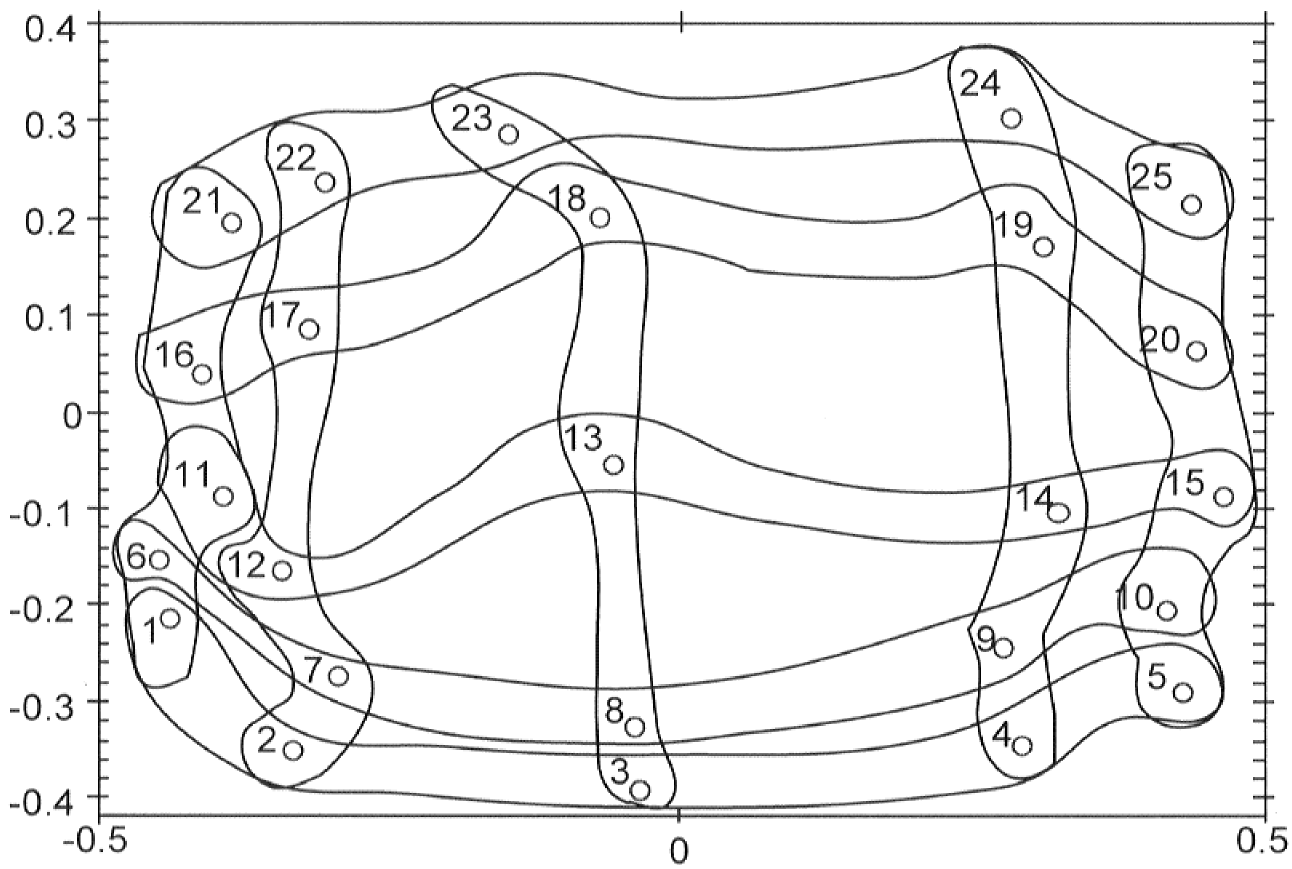


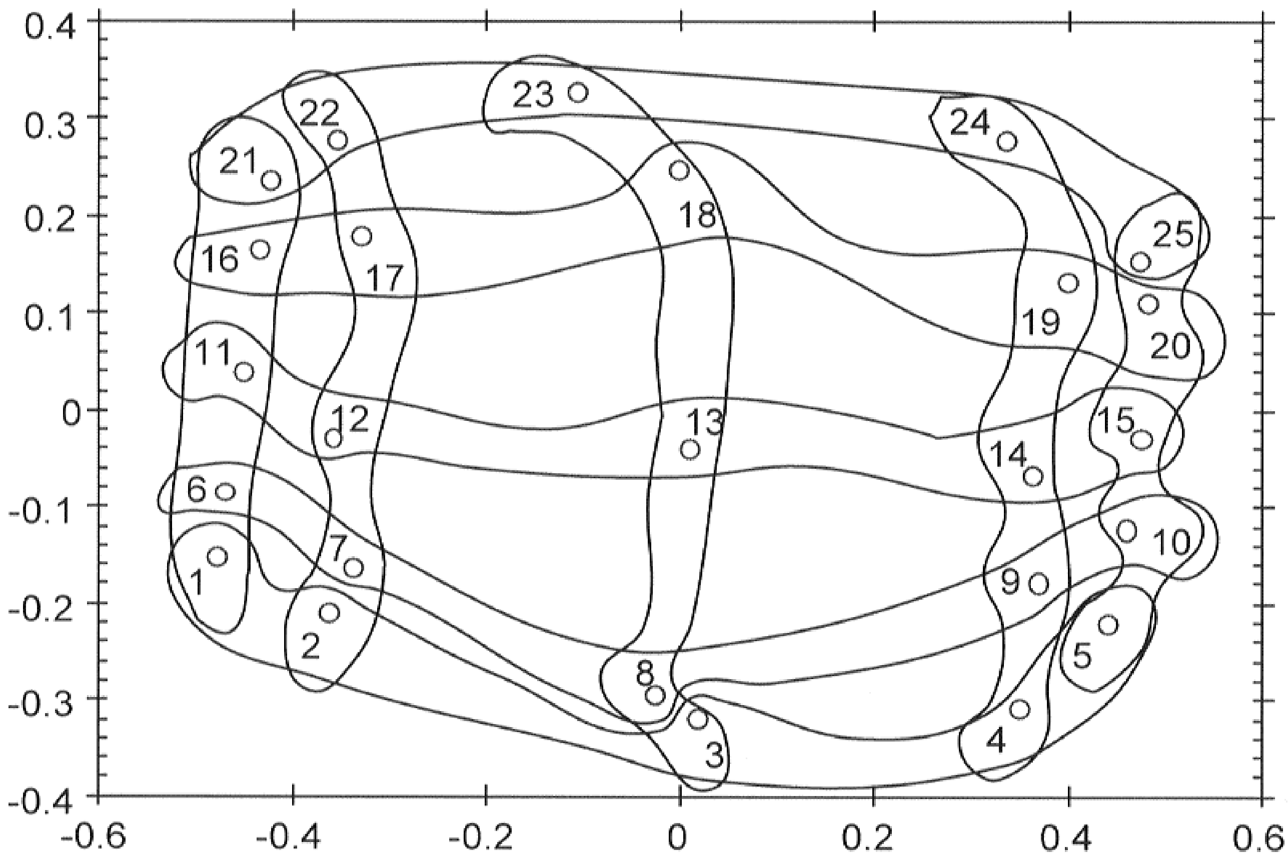
All 300 possible pairs were tested once each, preceded by a test of 25 individual sounds, to verify audibility, and 5 practice trials. The latter used sounds similar but not identical to those used in the experiment: Practice sounds covered two damping (20, 50) and three frequency (175, 300, 600) values and were distributed across the range of experimental sounds without hitting the limits.

Results

Similarity scaling. The initial analyses on similarity treated it as a proximity metric and subjected the data to a two-dimensional multidimensional scaling algorithm (Heady & Lucas, 1997). The solution is a set of x and y coordinates for each of the 25 unique sounds that made up the 300 pairs. Because the solution is rotation invariant in the x-y plane, we constructed two rotated versions, one that optimized the correlation between the x-coordinate and the log decay parameter of the sounds (this correlation is called r_d), and a second that optimized the correlation between the y-coordinate and the log frequency parameter (r_f). For Experiments 1 and 2, each subject's data were initially scaled separately, as one would not want to combine subjects to create an averaged similarity matrix if the scaling solutions were diverse. However, all subjects showed similar patterns -- r_d was significant (and for most, so was r_f), and when r_d was plotted against r_f, the values lay along or below the diagonal (i.e., frequency accounted for less variance than decay). Average solutions were therefore created by combining similarity matrices over the subjects in each experiment. Figure 3 shows these solutions, using the rotation that maximizes r_d. It should be noted that this rotation is also essentially the one that maximizes r_f, since rotations maximzing r_d and r_f differed by only 1° in Experiment 1 and 4° in Experiment 2. This indicates that the contributions of τ_d and f to the similarity judgments were essentially orthogonal. With r_d maximized, the correlation between the x-coordinate and the τ_d values was .98 in both experiments, and the correlation between the y-coordinate and the f values was .96 in Experiment 1 and .95 in Experiment 2. The Kruskal-1 stress values for the scaling solution were .16 and .13 in the two experiments, and Kruskal-2 stress values were .37 and .28. (Stress measures the goodness of fit of the pairwise similarities to intersound distances in a two-dimensional space.)

Figure 3. Multidimensional scaling solutions to similarity data of Experiments 1 (top) and 2 (bottom). The x axis is highly correlated with the decay parameter and the y axis with the frequency parameter (all r values > .95).





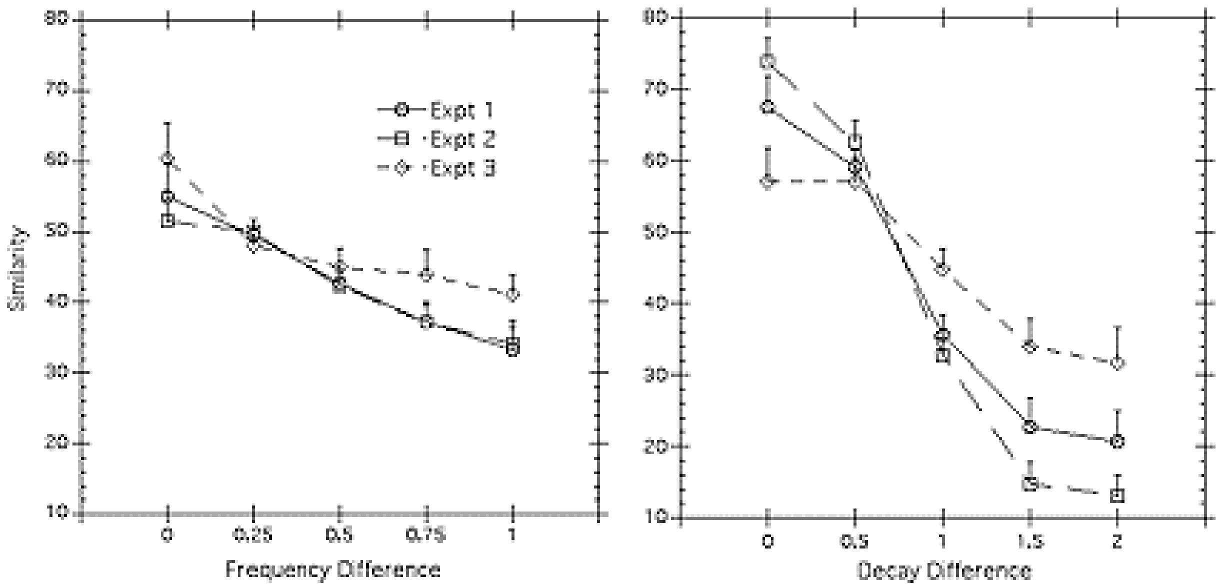
Although the correlations are comparable, the x-axis spread is about 50% greater than the y-axis spread in both studies, indicating a greater contribution of the decay parameter. It should be noted that the range of interstimulus differences (log units) was greater for decay than frequency, which could contribute to decay playing a greater role. However, contrasting scaling results of Experiment 3, using different instructions, and other results reported below indicate the similarities were not governed solely by the range of stimulus parameters.

The bands in Figure 3 indicate sets of stimuli that have equal frequency but vary in decay (horizontal bands) and those that have equal decay but vary in frequency (vertical bands). The stretching of the vertical bands at the middle of the x-axis indicates that the effect of frequency was greatest at intermediate values of decay. This phenomenon is not an artifact of the scaling algorithm; it can be observed as well in the raw similarity data. We have no account for it.

Scaling the data in Experiment 3 was more problematical, because subjects were clearly diverse. Seven of the 13 subjects showed a tendency for similarity to decrease as the difference between the sounds' decay values increased, whereas the remaining 6 showed essentially no effect of the decay parameter. Subjects were more consistent with respect to frequency, with 12 of the 13 subjects showing higher similarity when the sounds had the same frequency than when they differed maximally. Unlike the first two experiments, rotations of the scaling solution to maximize r_f and r_d differed by almost 60°, and Kruskal stress values were approximately twice as high as those found previously. Given these inadequacies, we do not further report a scaling analysis for Experiment 3.

Effects of stimulus differences on similiarity. Additional analyses of similarity were based on the 300 pairs of sounds and used two independent variables, frequency difference and decay difference. The first is the difference between the sounds in a pair with respect to the log value of the fundamental frequency (f) parameter (i.e., $abs\{log(f_1/f_{2)}\}$, where the subscripts index the members of the pair). The second is the difference between the two sounds in a pair with respect to the log value of the decay (τ_d) parameter (i.e., $abs\{log(\tau_{d1}/\tau_{d2})\}$). Each difference variable had 5 levels. Figure 4 shows how similarity decreased as the frequency and decay difference increased in each experiment. It indicates that the rate at which similarity decreased with increases in the difference variable was substantially greater for decay than frequency.

Figure 4. Relation of similarity to frequency difference and decay difference in Experiments 1-3; bars indicate one standard error.



Analyses of variance on similarity were used to assess these overall difference effects. Because frequency difference and decay difference did not vary factorially within the set of 300 pairs, separate ANOVAs were conducted on each. Treatment of these variables separately is consistent with the essentially independent effects of frequency and decay in the scaling analyses reported above. Each ANOVA used subjects as the unit of observation; within each subject, the similarity values for all pairs with the same difference in frequency or decay were averaged to produce a 5-level difference factor. An overall ANOVA included factors of experiment (between-subject) and level of difference (within-subject). The frequency-difference ANOVA yielded only a significant main effect of level of difference, $F(4, 148) = 26.67$, $p < .0001$. The decay-difference ANOVA yielded effects of level of difference, $F(4, 148) = 109.20$, $p < .0001$, and an experiment by difference interaction, $F(8,148) = 5.80$, $p < .01$. Supplementary analyses indicated that Experiments 1 and 2 did not differ with respect to the overall mean or the effect of decay difference, but there were experiment by decay-difference interactions when Experiment 3 was compared to Experiment 1, $F(4,96) = 3.42$, $p < .025$, and Experiment 2, $F(4,100) = 12.51$, p <

.0001. Thus the frequency effect was constant across experiments, but the instructions to judge length similarity in Experiment 3 altered the effect of decay relative to Experiments 1 and 2, where subjects judged material similarity.

Regression analyses predicting similarity, with the 300 pairs as units of observation, were also performed. The predictor variables were frequency difference, decay difference, and the product of these two differences. Table 1 shows the regression results for each experiment. In all three studies, the decay difference had the greatest contribution, with a smaller but significant effect of frequency difference that was essentially constant across experiments. However, the contribution of decay was substantially smaller in Experiment 3 (length judgments) than in the other two studies (material judgments), as is consistent with the ANOVA results. The interaction term (frequency difference times decay difference) was also significant in Experiments 1 and 2, although in both experiments, it accounted for only 2% of variance.

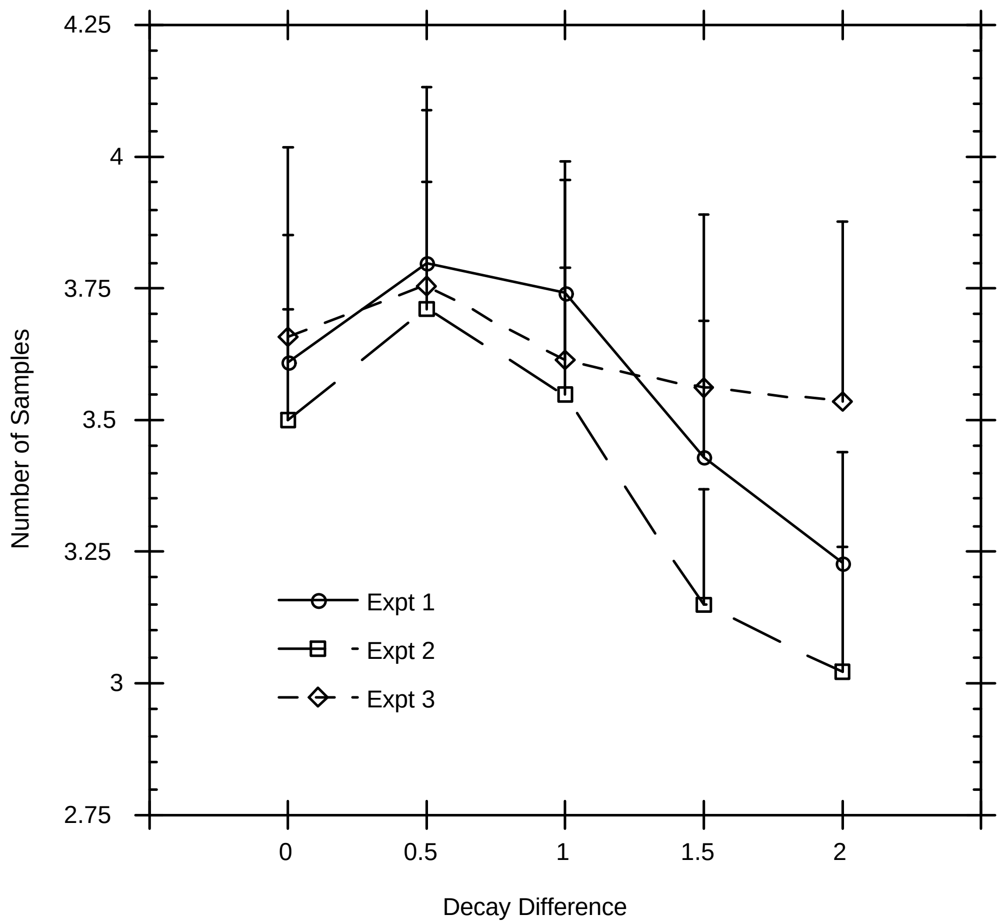
Table 1. Regression of similarity on frequency and decay difference, by experiment. Shown are the intercept and the standardized regression coefficients for frequency difference, decay difference and their product. Coefficients are significantly different from zero by 2-tail test, except where noted. Variance accounted for by regression is also given.

Experiment	Intercept	Freq. Diff. Coeff.	Decay Diff. Coeff.	Product Coeff.	R^2
1	88.7	-0.56	-1.02	0.30	.76
2	96.2	-.50	-1.06	0.31	.80
3	72.2	-0.48	-0.75	-0.13[a]	.56

[a]p > .05

Exploration. Figure 5 shows the total number of times the two sounds were sampled on a trial before the response was made, as a function of the decay difference. The asymmetric inverted-U pattern indicates that fewest samples were needed when the difference in log τ_d was large (i.e., similarity was low), and the greatest sampling occurred when the difference was intermediate. This effect is not large in magnitude (the largest mean difference is within one sample), but it was significant in an ANOVA on number of samples with decay difference as one factor, $F(4, 148) = 14.38$, $p < .001$, and experiment as a second factor. Although the interaction did not reach significance, the decay-difference effect was significant in Experiments 1 and 2, $F(4, 48) = 9.38$, and $F(4, 52) = 7.93$, $ps < .0001$, but not in Experiment 3. The corresponding analysis of the effect of frequency difference on number of samples was not significant overall or in any experiment individually, all $ps > .30$.

Figure 5. Number of times two sounds were sampled on each trial before response, as a function of decay difference, in Experiments 1-3; bars indicate one standard error.



Discussion of Experiments 1-3

The results of these studies indicate that people can use the sound of impact with an object as a cue to its material substance. Both of the auditory parameters that were manipulated here, decay and frequency, affected people's judgments of material. The regression analysis and scaling analyses of Experiments 1 and 2 indicate, in addition, that the contribution of decay was greater by a factor of approximately 2:1 and was the only parameter sensitive to the instructions to judge material vs. length. The regression analysis provides an explicit metric for the perceived similarity of the material of two objects, based on impact sounds. Approximately 80% of the variance in the present similarity judgments was accounted for by just two parameters, frequency and decaydifference. That subjects were judging material rather than arbitrarily comparing the sounds is shown by Experiment 3, in which instructions to judge length significantly altered the contribution of the decay parameter.

The number of times people sampled the sounds before responding was affected only by decay, and not frequency. Relatively little sampling (within the modest range of variation that was observed) was needed to judge the sounds as dissimilar when they had markedly different decay values, which underscores the importance of decay.

Experiment 4

In Experiment 4, we directly asked subjects to indicate what material a sound might emanate from. The assignment of sounds to four material categories was used to determine the relative dependence of each category on frequency and decay. Both parameters are related to material, as was reviewed in the introduction.

Method

Fifty students in a classroom listened to the 25 sounds from Experiment 2 (i.e., varying in initial amplitude) in a common random order, preceded by 2 practice sounds. Each sound was played three times in succession for 500 ms each, followed by a 5 s interval in which the subject responded by marking an answer sheet to indicate which of four classes of material seemed to best correspond to the source of sound. They were told the sounds did not necessarily come from a real material. The four response classes were rubber, wood, glass and steel; across subjects, two different orders of these classes were used on the answer sheets.

Results and Discussion

The principal results are shown in Figure 6 and Table 2. Figure 6 shows the number of subjects, out of 50, who assigned each object to a given material category, as a function of the object's frequency and decay parameters. The intersubject agreement (proximity of response proportion to 1 or 0) indicates how plausible or implausible the sound was as an exemplar of the particular category. As the value of τ_d increased, the tendency to respond with glass and steel increased, and the tendency to respond with rubber and wood decreased. These findings are consistent with data reported in Wildes and Richards (1980), indicating that the parameter t_e (τ_d scaled by a frequency factor) was in an increasing order for rubber, wood, glass and steel. Frequency can also be seen to affect the current responses.

The relative contribution of frequency and decay (individual stimulus values, not difference scores as in Experiments 1-3) was assessed with multiple regressions, predicting the number of respondents choosing a particular category for a stimulus from its values of log frequency and log decay. There was one regression for each material category; Table 2 reports the results by category. For all four material categories, the decay parameter was a significant predictor (by 2-tailed t test, alpha set at .05) and had a higher coefficient than frequency. Frequency was a significant predictor for glass and rubber -- the tendency to label a sound “glass” increased with frequency, and the tendency to label it “rubber” decreased. These contributions of frequency likely reflect people’s everyday experience with contact sounds produced by glass and rubber objects.

Figure 6. Proportion of subjects who assigned each sound to a given material category, as a function of the sound's frequency and decay parameters (log values). Darker shading represents higher values; blanks indicate zero values.

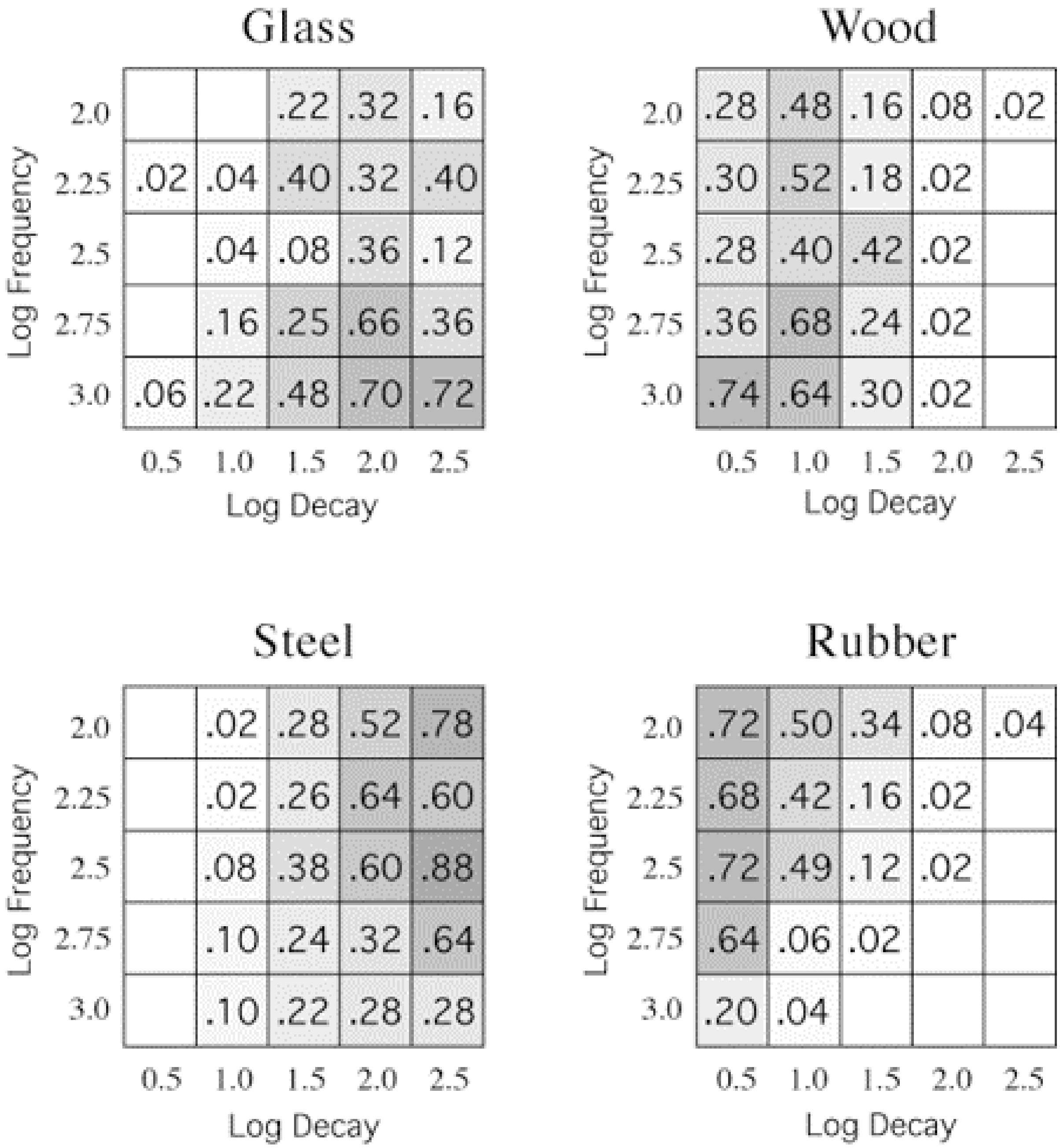


Table 2. Regression of number of responses assigning sound to a material category, on the log frequency and log decay parameters. Standardized regression coefficients and variance accounted for by regression are given. Coefficients are significantly different from zero by 2-tail test, except where noted.

Material	Intercept	Freq Coeff.	Decay Coeff.	R^2
rubber	66.6	-0.38	-0.79	0.77
wood	15.2	0.20[a]	-0.79	0.66
steel	5.9	-0.18[a]	0.89	0.82
glass	-37.7	0.43	0.66	0.62

[a] $p > .05$

A final analysis was used to derive an empirical value of decay associated with each category. Specifically, a simple linear regression was used to predict the number of respondents choosing a particular category for a stimulus from its value of log decay. From the resulting slope and intercept, we determined the value of an object's log τ_d that would lead to half the subjects assigning it to the given category. This critical value of log τ_d was .46, .50, 2.10, and 2.65 for rubber, wood, glass, and steel. These values follow the same ordering as the physical t_e values reported by Wildes and Richards (1980).

General Discussion

To summarize our findings, using multiple measures (similarities, samples taken, and categorical assignment frequencies), we found that a shape-invariant decay parameter was a powerful determinant of the perceived material of an object. Frequency had a lesser, but significant, contribution. These tasks enabled us to predict a substantial proportion of variance in similarity judgments and categorical assignments from the two variables of interest. Moreover, a derived estimate of the decay parameter associated with each material in the categorical assignment task was directly related to reported measures of the internal coefficient of friction for the same material categories.

Our conclusion that decay plays a larger role than frequency in material perception is based on the differential contribution of the two factors to judgments of material similarity, but not length similarity, and on the greater sensitivity of exploration and classification measures to the decay parameter. However, while the present manipulations clearly led to a greater impact of decay, caution about the generality of this finding is warranted: We have noted that the range of interstimulus differences manipulated here was greater for decay. Moreover, even when the two manipulations are compared while taking the stimulus range into account (as in comparing the slopes in Figure 4), equal log-unit changes in two such different variables do not necessarily have equal subjective or objective impact (when considered, e.g., relative to the typical range of each variable, the cost of manipulating a variable during synthesis, etc.). Further research is needed to address these issues.

Frequency and decay were found to be essentially independent determinants of similarity judgments, according to the scaling and regression analyses. On the other hand, the two factors interacted in their influence on subjects' categorization of the sounds into categories. These findings suggest that people can consider frequency and decay independently when comparing sounds of impact. However, when assigning the sounds to categories, they take both parameters into account and define at least some categories by the frequency/decay combination. For example, sounds taken to be glass are high-frequency and also decay slowly.

An important question is what acoustic concomitants of decay underlie judgments of material. We considered whether subjects actually use the rate of intensity loss or whether they use other mediating variables affected by the decay rate, such as the total time that a sound is audible or the total energy during its audible duration. Experiment 2, which controlled for those particular variables by presenting sounds that varied in initial amplitude, suggests that subjects can use the decay rate itself.

Decay affects not only the rate of intensity loss but the mixture of frequency spectra in a sound. Consider Equations 1 and 2, which together indicate that the time for a sound to decay to

1/e of its current value, t_e, is equal to τ_d/f. That is, time to decay is frequency dependent, with higher-frequency modes being lost faster than lower-frequency modes. A high decay parameter τ_d will lead to the rapid dominance of the lower frequencies. This effect of decay on the mixture of frequency spectra appears to be critical to our perception that the sound emanates from a particular material. In an informal test, we synthesized sounds using the modes of a bar, but holding the decay rate constant across frequencies. The resulting sounds did not appear to emanate from a simple object of uniform material.

The present results have direct implications for the design of virtual environments in which interactions with objects have auditory as well as visual consequences. First, they show that a simple, one-parameter model of material is sufficient for many tasks such as detecting a change of material or classification into a material category. This is important because even though real sounds have a complex dependence on material properties, the amount of computation allocated to compute this dependence within a virtual interactive environment is typically very small, in part to achieve low latencies, and in part due to competing demands from graphics rendering. For instance, in computer games it is typical for the "audio" budget to be approximately 3% of the CPU cycles. Related work in the domain of musical sounds has shown some simplified descriptions to be useful for synthesizing sounds that can be discriminated by timbre (McAdams, Beauchamp, & Meneguzzi, 1999).

Second, our results provide an empirical metric of the perceived similarity of sounds, as a function of damping (decay) and fundamental frequency. This is very useful for efficiently sampling sound synthesis parameters of objects. Specifically, since the sound made by a typical inhomogeneous object depends on the contact location on the object, one can associate values of the critical sound parameters, damping and frequency, with each point on the object. One can think of this as a "sound mapping," analogous to the texture mapping used in computer graphics. There has been little guidance as to how to sample this map efficiently -- too fine a sampling leads to large storage requirements, while too coarse a sampling could miss interesting auditory details (such as are caused by a nail or stud inside a wall). The present work takes an important first step towards sampling based on perceptual similarity of contact sounds on the surface of an object, allowing us to pick finer grids only in areas where nearby spatial locations have salient differences in their auditory consequences.

References

Carello, C., Anderson, K. L. & Kunkler-Peck, A. J. (1998). Perception of object length by sound. Psychological Science, 9, 211-214.

Chaigne, A., & Doutaut, V.(1997). Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars. Journal of the Acoustical Society of America, 101(1), 539-557.

Coleman, P. D. (1963). An Analysis of Cues to Auditory Depth Perception in Free Space. Psych. Bull., 60, 302-315.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. Journal of the Acoustical Society of America, 87, 311-322.

Gaver, W. W. (1988). Everyday listening and auditory icons. Unpublished doctoral dissertation, University of California in San Diego.

Green, D. M. (1976). An Introduction to Hearing. Hillsdale, NJ: Erlbaum.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 61, 1270-1277.

Heady, R. B., & Lucas, J. L. (1997). PERMAP: An interactive program for making perceptual maps. Behavioral Research Methods, Instruments, & Computers, 29, 450-455.

Krotkov, E. Klatzky, R. & Zumel. N. (1996). Robotic perception of material: Experiments with shape-invariant acoustic measures of material type. In O. Khatib & K. Salisbury, Experimental robotics IV: Lecture notes in control and information sciences 223, (pp. 204-211). New York: Springer-Verlag.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. Perception & Psychophysics, 59, 1180-1190.

Massaro, D. W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Mahwah, NJ: Erlbaum.

McAdams, S., Beauchamp, J. W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. Journal of the Acoustical Society of America, 105, 882-897.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58, 177-192.

McGurk, H., & MacDonald, T. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Miller, J. R. & Carterette, E. C. (1975). Perceptual space for musical structures. Journal of the Acoustical Society of America, 58, 711-720.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), Frequency analysis and periodicity detection in hearing (pp. 397-414). Leiden: Sijthoff.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. Journal of the Acoustical Society of America, 81, 1100-1110.

Takala, T. & Hahn, J. (1992). Sound rendering. Proceedings of the SIGGRAPH'92 Conference. ACM Computer Graphics, 26(2), 211--220.

van den Doel, K., & Pai, D. K. (1998). The sounds of physical shapes. Presence, 7 (4), 382-395.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. Journal of Experimental Psychology: Human Perception and Performance, 10, 704-712.

Wildes, R. P., & Richards, W. A. (1988). Recovering material properties from sound. In W. Richards (Ed.), Natural Computation. Cambridge, MA: MIT Press.

Yost, W. A., & Hafter, E. R. (1987). Lateralization. In W. A. Yost & G. Gourevitch (Eds.), Directional Hearing. Berlin: Springer.

Zener, C. (1948). Elasticity and anelasticity of metals. Chicago: University of Chicago Press.