

How Do We Hear in the World?: Explorations in Ecological Acoustics

William W. Gaver
*Rank Xerox EuroPARC and
Technische Universiteit Delft*

Everyday listening is the experience of hearing events in the world rather than sounds per se. In this article, I explore the acoustic basis of everyday listening as a start toward understanding how sounds near the ear can indicate remote physical events. Information for sound-producing events and their dimensions is investigated using physical analyses of events to inform acoustic analyses of the sounds they produce. The result is a variety of algorithms which enable the synthesis of sounds made by basic-level events such as impacts, scraping, and dripping, as well as by more complex events such as bouncing, breaking, spilling, and machinery. These algorithms may serve as instantiated hypotheses about the acoustic information for events. Analysis and synthesis work together in their development: Just as analyses of the physics and acoustics of sound-producing events may inform synthesis, so listening to the results of synthesis may inform analysis. This raises several issues concerning evaluation, specification, and the tension between formal and informal physical analyses. In the end, however, the fundamental test of these algorithms is in the sounds they produce: I describe them in enough detail here that readers may implement, test, and extend them.

Imagine that you are a participant in a psychology experiment on the perception of complex sounds. Your task is to listen to a series of sounds and write down a brief description of what you hear.

"A single-engine propeller plane flying past," you write in response to the first sound, pleased with yourself for providing so much detail.

The experimenter, on the other hand, is not pleased. He says, with some irritation, "No, no no. Write down what you *hear*, not what you think it is."

"But I heard a propeller plane fly past," you object. "I didn't think about it; that's what I heard."

"You may not have thought about it consciously," he retorts, "but you didn't hear an airplane, you heard a quasi-harmonic tone lasting approximately 3 seconds with smooth variations in the fundamental frequency and the overall amplitude. That's what I want you to tell me about."

"I don't understand," you persist, though a little hesitantly. "I didn't hear whatever it is you said. I heard a propeller plane."

The experimenter sighs and explains patiently, "No, you *interpreted* the sound as a propeller plane by matching the incoming stimulus with representations stored in memory. I'm not interested in how people interpret sounds; that's a job for cognitive science. I'm interested in how you hear the sound itself. Now try again . . ."

Are you wrong, in this scenario, to insist that you hear a passing plane? Or is the experimenter wrong in pressing you to describe the sound and ignore its source? Neither: The misunderstanding arises because you are focusing on different ways of experiencing the same sound. Your experience of hearing "a single-engine propeller plane flying past" is an example of *everyday listening*, the perception of sound-producing events. The experience of "a quasi-harmonic tone lasting approximately 3 s with smooth variations in the fundamental frequency and the overall amplitude," on the other hand, is an example of *musical listening*, the experience of the sounds themselves. It's not that you have mistaken one sort of sound for another—it is possible to hear any sound in terms of its source (everyday listening) or in terms of its sensory qualities (musical listening). And it is not that as the subject in this experiment, you are somehow more naive than the experimenter—both ways of experiencing the sound are valid. It's simply that you are listening to the sound differently from the experimenter, and the descriptions appropriate for the two experiences are incompatible with one another.

Fortunately, you are unlikely to find yourself in this situation. Experienced psychoacousticians would be unlikely to present such natural sounds, or such an open-ended task, precisely because their subjects would be likely to respond to the sounds in terms of the events that caused them. Everyday listening is rarely addressed in psychology; most research on sound and hearing has focused on musical listening. In part this is because the acoustic richness of natural sounds

makes understanding their perception a daunting task. Beyond this, however, many psychologists assume that sounds do not usually convey sufficient information to specify their sources. From this perspective, supplemental information about sound-producing events must be provided by memory, unconscious inferences, or problem solving, and thus everyday listening must reflect higher level processes to a greater degree than musical listening. This line of reasoning suggests that hearing a passing plane is more arbitrary than hearing the pitch or loudness of the sound it makes, because the information provided from memory or experience depends on the individual, and may simply be wrong (see J. Gibson & E. Gibson, 1955/1982). The result is a focus on the general cognitive mechanisms assumed to underlie perception of the world, with the specifics of everyday listening falling in the cracks between psychoacoustics and cognition.

But people *do* hear events as well as sounds, and the information they pick up about events is not as arbitrary as the experimenter in the example might have us believe. People can reliably hear the material of a struck bar (Gaver, 1988), the hardness of the mallet striking it (Freed, 1990), whether a bottle has bounced or broken (Warren & Verbrugge, 1984), even the configuration of clapping hands (Repp, 1987). These results suggest that sound conveys more information than proponents of the cognitive approach give it credit for, information sufficient to specify many features and dimensions of events. Thus everyday listening can be studied in its own right, not just as an instance of general-purpose pattern matching. Instead, an approach can be pursued that focuses on the consistent structure of the world that allows sounds to relate reliably to their sources. This is both the premise and the goal of this article.

STUDYING EVERYDAY LISTENING

Studies of everyday listening may serve as the foundation of a new framework for understanding sound and hearing. Such an account complements the one offered by traditional psychoacoustics, allowing us to understand listening and manipulate sounds along dimensions of sources rather than sounds. So, for example, we might explore what people reliably hear about a passing airplane and seek to understand the acoustic correlates of their experiences. Such studies lead in turn to a number of practical applications involving the recognition or synthesis of sound-producing events.

What should an account of everyday listening be like? It must answer two fundamental questions. The first question is "*What do we hear?*" In seeking to expand accounts of sound and hearing beyond the traditional preoccupation with sensations such as pitch and loudness, we must develop a framework for describing ecologically relevant perceptual entities: the dimensions and features of events that we can actually hear. The second question is "*How do we hear it?*" As we expand our account to include dimensions and features of events, we

must also develop an *ecological acoustics*, one that describes the acoustic properties of sounds conveying information about the events we hear. The first question deals with the content of everyday listening, whereas the second deals with the acoustic structures enabling that content to be heard. By exploring both questions together, we can start to build a psychophysics of auditory event perception, in which the relations among perceptual and acoustic dimensions of events are examined.

In a companion article (Gaver, 1993), I explored the first of these questions, using qualitative physical analyses combined with the results of protocol studies to build a map of sound-producing events. According to this scheme, sounds are caused by and convey information about the interaction of materials at a location in an environment. Sources of sounds—interactions of materials—can be categorized according to whether they involve solids, liquids, or gasses. Within each of these categories, *basic-level events* (e.g., impacts, pouring, explosions) are those that involve a single interaction of objects. More complex sounds can be described in terms of these simple basic-level events, with complexity coming from temporal patterning of a single basic-level event, combinations of more than one event, or a hybrid of events from different material classes. This framework represents an early step in understanding the structure of everyday listening: It is primitive, but useful in interpreting existing research and guiding new explorations.

The purpose of this article is to start addressing how we hear events by examining the acoustic patterns that correspond to a variety of sound-producing events and their attributes. This may seem a strange approach to the question of how we hear the world. Traditional accounts, after all, would emphasize the internal processes that mediate between the arrival of a sound at the ear and the experience of its source, focusing on transduction, pattern-recognition, and the like. But from the ecological perspective, understanding the ways that structured sounds might specify their sources is a fundamental step in understanding how listeners can hear events in the world. If the input for perception is inadequate to specify events, then processing mechanisms must be complex to compensate. But if the input is rich with information, the job of registering it becomes simpler (see Runeson, 1977). This trade-off between the structure in the world and structuring in the mind is, I believe, at the heart of J. Gibson's (1979/1986) claim that perception is direct. It involves both a claim about perception and a metatheoretical argument about where the burden of explanation is to be placed and how research is to proceed. From this point of view, a major component of explaining how we hear events in the world involves understanding how acoustic patterns near the head can indicate potentially remote physical events.

My explorations of acoustic information have been guided by the framework described in Gaver (1993). They focus on basic-level events and complex sounds that may be described in terms of their basic-level components. In addition, they have been guided by more pragmatic considerations such as the tractability of

physical analyses, possibilities for efficient synthesis, and value for practical applications. The result of these explorations is a variety of algorithms permitting everyday sounds to be synthesized and controlled along dimensions of their sources—for example, by specifying that a long bar of metal is to be struck by a soft mallet. These algorithms may be seen as instantiating theories about the acoustic information for events. In addition, they may serve as the basis for a variety of applications that use everyday sounds to convey information.

ANALYZING ACOUSTIC INFORMATION FOR EVENTS

Creating algorithms that enable synthesis of virtual events implies an understanding of the acoustic information for event attributes—how sounds indicate the material or size of an object, for instance. Such attributes often have very complex effects on sounds, effects that must be described as functions of frequencies and amplitudes over time that describe the *partials*, or frequency components, that make up a sound. If these functions are understood, source attributes can be specified directly, instead of via separate controls over partial frequencies, amplitudes, and durations. The algorithms thus constitute hypotheses about the acoustic information for these attributes. But how can we determine what these functions are?

Analysis and Synthesis of Sounds

Computer musicians have addressed a similar problem: One of the goals for some (but not all) computer musicians is to find efficient means to capture the sounds of traditional acoustic instruments. Such sounds can be completely described by the output of a time-varying Fourier analysis, but this description is likely to be huge. A method known as *analysis and synthesis* (see, e.g., Risset & Wessel, 1982) aims at understanding how to reduce the data from such analyses so that only the information necessary for recreating a perceptually identical sound is retained.

Analysis and synthesis, as the name implies, involves an iterative process of analyzing a recorded instrument sound and then synthesizing a duplicate on the basis of the analysis (Figure 1). The analysis data can be systematically reduced, for instance by using straight line-segments to approximate complex time-varying attributes of the sound, and the resulting synthesized sound compared to the original. If there are no perceptible differences between the original and the synthesized versions (or if the differences are acceptable) then the discarded data are deemed irrelevant for perception. In this way, an understanding of the aspects of a given sound crucial for music may be obtained.

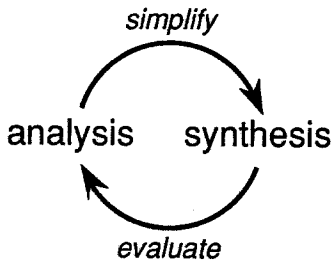


FIGURE 1 Traditional analysis and synthesis.

Analysis and Synthesis of Events

Acoustic information for events can be studied in a similar way. The sounds made by actual events can be analyzed, the data reduced, and synthesized sounds compared to the originals. For instance, if one supposes that the temporal features of a sound are responsible for the perception of some event, but that its frequency makeup is irrelevant, one might use the amplitude contour from the original sound to modify a noise burst. The hypothesis can then be assessed simply by listening to the result.

Note that the criterion in this case is not that the synthesized version should be perceptibly identical to the original sound, but instead that it should convey the same information about a given aspect of the event. Analysis and synthesis as used by computer musicians is judged with respect to musical listening, but analysis and synthesis used to study information for events must be judged with respect to everyday listening. Thus while computer musicians aim at creating a simpler but perceptibly identical sound, students of everyday listening must find the mapping between the physics of the event and the attributes of the resulting sound that serve as information to a listener. They must relate three levels of analysis, understanding—at some level of detail—(a) the physics of the event, (b) how that is reflected by the acoustics of the sound, and finally (c) how that gives rise to the perception of the event. Accurate reproduction is thus a validational criterion, not an end in itself.

Physical analyses of events. Simple analysis and synthesis explores the acoustic information for different sound-producing events by comparing analyses of the sounds they make (see, e.g., Figure 3). In practice, however, it is often difficult to identify the acoustic information for events in the mass of data produced by acoustic analyses. Thus, it is useful to supplement acoustic analyses with physical analyses of the event itself (Figure 2). Studying the physics of sound-producing events is useful both in suggesting relevant source attributes that might be heard and in indicating the acoustic information for them. Resynthesis, then, can be driven by the resulting physical simulations of the event.

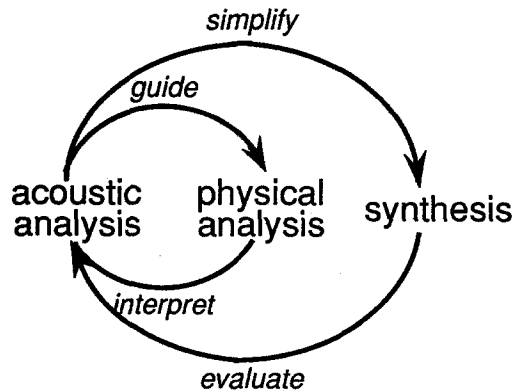


FIGURE 2 Analyzing and synthesizing events requires physical as well as acoustic analyses.

Computer musicians, too, have increasingly turned to physical modeling to guide synthesis (e.g., Borin, De Poli, & Sarti, 1993; Cremer, 1981/1984; Jaffe & Smith, 1983; McIntyre, Schumacher, & Woodhouse, 1983), and their work offers a great deal of insight into analytical, modeling, and synthesis techniques. However, although their work is relevant to an understanding of everyday listening, it is a clearly distinct endeavor. The domain of sounds they have studied—primarily those produced by musical instruments—is relatively limited. The aim of their models is not to explicate acoustic information for events, but instead to recreate the sounds of traditional musical instruments for musical purposes. Thus the focus is on developing means for synthesizing sounds that are musically expressive (and experienced in terms of musical listening), rather than on explicating how the acoustic structure thus provides information about sources. Finally, acoustic instruments are often mechanically very complex. Not only does this imply that physically modeling them may be very difficult, but that the physical dimensions necessary for musically accurate modeling may not correspond to those that are phenomenally perceptible. For these reasons, I consider a variety of simpler, everyday sound-producing events in trying to understand the acoustic basis of everyday listening.

One of the strengths of analysis and synthesis as a methodology is the ability to evaluate synthesis algorithms simply by listening to the sounds they produce. A good example of this is the development of an algorithm for creating dripping sounds, described later, in which informal listening was sufficient for ruling out an otherwise plausible model. But it should be noted that the algorithms I present in this article have not been subjected to experimental evaluation. Unless otherwise indicated, their ability to capture relevant information has only been assessed by myself, colleagues, and audiences of talks in which I have presented this work, simply by listening to the sounds the algorithms create and judging informally whether they “sound right.” Clearly, more formal tests are possible and desirable; potential methods are considered in the Conclusions

section of this article. Nonetheless, the efficacy of analysis and synthesis in allowing a quick iteration between modeling and listening should not be underestimated.

In the following sections, I discuss several case studies of events that have been studied using analysis and synthesis of sounds and events and describe the algorithms that have resulted. The algorithms are described in order of their complexity. I start with the sounds made by three basic-level events: impacts, scraping, and dripping. Impacts are an example of a basic-level event that has been well studied by mechanical physics; my analysis focuses on the object properties that are perceptible. Algorithms for scraping and dripping sounds rely on more qualitative analyses; these analyses emphasize perceptible properties of the interactions causing the sounds. This style of analysis becomes important in considering more complex sound-producing events. Thus, I next expand on Warren and Verbrugge's (1984) analysis of a complex sound-producing event, developing an algorithm that produces bouncing, breaking, and spilling sounds depending on the temporal patterning and materials involved in a series of impact events. Finally, I describe an algorithm for producing machinelike sounds, showing that high-level attributes of complex events may be captured directly.

IMPACT SOUNDS

Many of the sounds we hear in the everyday world involve one solid impacting against another. Tapping on an object, placing it against another, letting it fall—all these involve impact sounds. Impacts are a basic-level event in the sense that they are produced by a simple interaction of objects; combinations of impacts may produce more complex events such as footsteps, hammering, or bouncing noises. Because they are basic-level events, understanding the information they convey is useful in understanding a great many more complicated events.

Information for Length and Material

I studied the acoustic information available for length and material of struck wooden and metal bars and people's abilities to perceive these attributes (Gaver, 1988). To understand the acoustic information for length and material, I recorded and analyzed the sounds made by wood and metal bars of several different lengths (see Figure 3) in an attempt to identify the acoustic correlates of these dimensions. As suggested earlier, however, it was difficult to identify the information for material and length using these analyses alone. Clearly, the sounds made by metal bars last longer than those made by wooden ones, but what else might distinguish wood and metal? What are the differences due to length? Which of these apparent differences generalize, and which are idiosyncratic to the specific examples that were analyzed? To approach these questions,

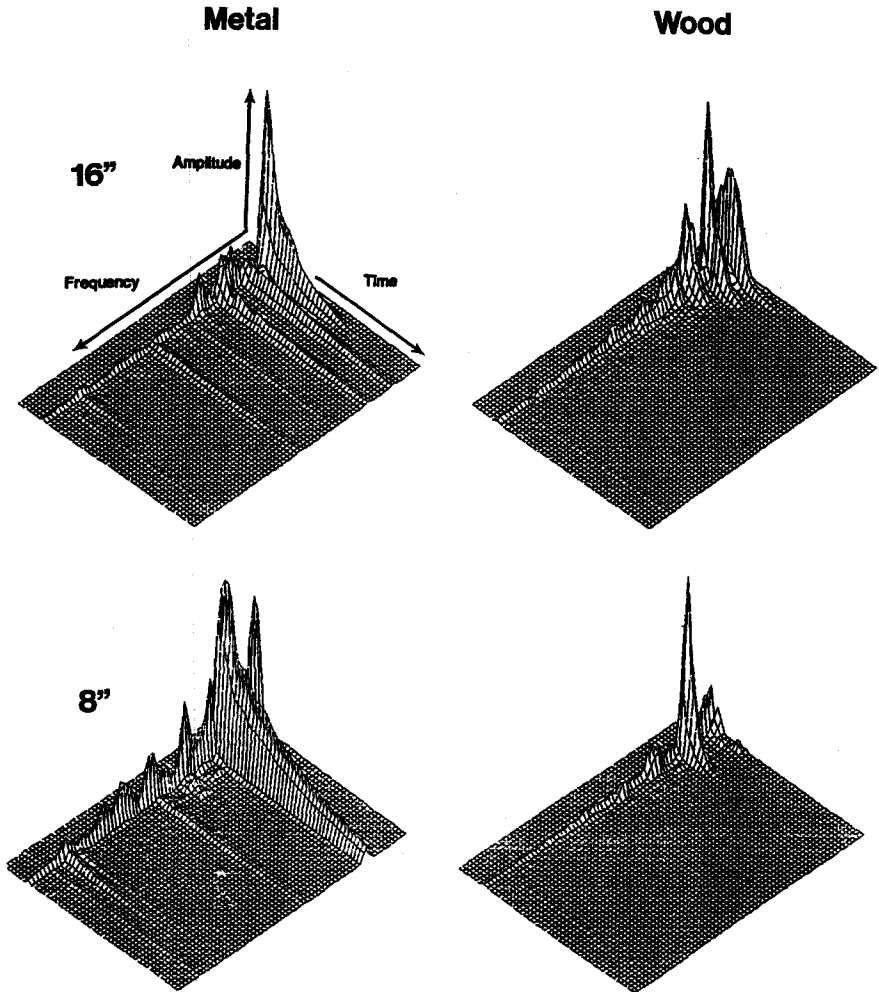


FIGURE 3 Spectrograms of impact sounds from long and short, wood and metal bars (after Gaver, 1988).

I developed a model of the physics of the impacts that combined analytical solutions to the wave equation for transverse vibrations in a bar (see, e.g., Lamb, 1960) with empirical measurements of damping and resonance amplitudes. This model was used both to aid interpretation of acoustic analyses of the sounds and to synthesize new tokens.

The material of the bars affected the sounds they made in several ways. Perhaps most important, materials have different characteristic frequency-dependent damping functions: The sounds made by vibrating wood decay

quickly, with low frequencies lasting longer than high ones, whereas the sounds made by vibrating metal decay slowly, with high-frequency components showing less damping than low ones. In addition, metal sounds have partials with well-defined frequency peaks, whereas wooden sound partials are smeared over frequency space. Finally, it appeared that each material supported some frequencies of vibration better than others; this was modeled as an inverted U-shaped function around the material's peak frequency that modified the amplitudes of partials according to their difference from that frequency.

These complex effects of material on impact sounds contrast with the simple effect of length. Changing the length of a bar simply changes the frequencies of the sound it produces when struck, so that short bars make high sounds and long bars make low ones. However, the effects of length interact with the effects of material. For instance, the frequencies of the partials change monotonically with length, but because of the inverted U-shaped function modifying the amplitudes of the partials, the frequency of the partial with the highest amplitude may change nonlinearly with length (Gaver, 1988; see Figure 4). This was reflected by the judgments made by experimental participants: Listeners were quite accurate at judging material based on sounds (including synthesized ones), but their length judgments were less accurate. Although a brief training session (listeners judged two examples of sounds from each bar length with feedback) improved performance greatly, discontinuities in judgments remained, apparently because of the nonlinear relation between the loudest partial and the length of the bars. Consultations with physicists revealed, however, that the inverted U-shaped amplitude function is probably not accurate for modeling material, but rather an artifact of the way the bars were supported. Eliminating this function from the model produces sounds that more clearly indicate length.

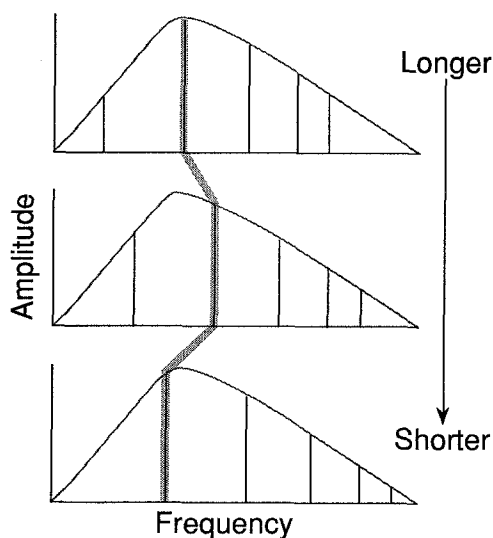


FIGURE 4 The partials of impact sounds rise in frequency as shorter bars are struck, but a stationary amplitude function causes nonlinear changes in the frequency of the highest amplitude partial.

Internal friction and material. Wildes and Richards (1988) independently studied material identification based on impact sounds from an analytic point of view. They start by noting that whereas there is no time dependence between stress and strain for ideal elastic materials, there is for any real solid. This means that an object's return to equilibrium after the removal of an impacting force lags compared to the impact. This characteristic is embodied in their model of a *standard anelastic linear solid*. The dynamic behavior described by this model depends on an intrinsic parameter of material called *internal friction*, which determines both the bandwidth of a vibrating object's partials and the rate of their decay. Thus, Wildes and Richards suggested that hearing material may depend on assessing internal friction on the basis of peak bandwidth and decay rate. This view is consistent with my finding that wood and metal are differentiated both by their decay rates and by the fact that partials of metal sounds are more sharply defined than those made by wood (Gaver, 1988).

Mallet Hardness

Freed (1990) studied people's perception of the hardness of mallets used to strike objects. He recorded the sounds made by hitting cooking pans with mallets of various hardnesses. The sounds were analyzed using a model of the peripheral auditory system which provides a description similar to that provided by a Fourier analysis, but which is more similar to the output of peripheral auditory processing.

Freed described the results of this analysis in the form of four "summarizing parameters" which were meant to capture the information for mallet hardness in these sounds. The first two, *spectral level* and *spectral-level slope* are measures of overall loudness and change of loudness with time, respectively; the second two, *spectral centroid* and *spectral-centroid time-weighted average* (TWA) are measures of the ratio of high- to low-frequency energy in the sounds and its change, respectively. These parameters were used to predict hardness judgments in a multiple regression; most useful were the measures of spectral centroid and spectral-centroid TWA. To a good approximation, then, mallet hardness is conveyed by the relative presence of high- and low-frequency energy.

A Synthesis Algorithm for Impact Sounds

The results of the studies reviewed in the previous sections may be captured in a synthesis model that uses frequency and amplitude functions to constrain a formula for describing exponentially decaying sounds. Equation 1 describes a complex wave created by adding together a number of sine waves with independent initial amplitudes and exponential decay rates¹:

¹I am grateful to David Woodhouse (personal communication, 1991) for showing me Equation 1 and its efficient implementation as described in the Appendix.

$$G(t) = \sum_n \Phi_n e^{-\delta_n t} \cos \omega_n t \quad (1)$$

where $G(t)$ describes the waveform over time, Φ_n is the initial amplitude, δ_n the damping constant, and ω_n the frequency of partial n .

This equation has two properties that make it a useful foundation for synthesizing impact sounds. First, its components map well to event attributes. Second, it can be made computationally efficient using trigonometric identities (see Appendix).

Mapping synthesis parameters to source attributes. By constraining the values used in Equation 1, useful parameters can be defined which correspond well to the attributes of impact sounds discussed earlier. Equation 1 involves three basic components: the initial amplitudes of the partials Φ_n , their damping $e^{-\delta_n t}$, and their frequencies $\cos \omega_n t$. These can be set separately for each partial. However, these three components also correspond to information for mallet hardness and impact force, material, and size and shape respectively (see Table 1). Thus it is more useful to define patterns of behavior over the partials for each component.

For example, the partial frequencies ω_n can be constrained to patterns typical of various object configurations. The sounds made by struck or plucked strings, for instance, are harmonic, so that $\omega_n = n\omega_1$. The sounds made by solid plates, in contrast, are inharmonic and can be approximated by random frequency shifts made to a harmonic pattern. The sounds made by solid bars can be approximated by the formula $\omega_n = (2n + 1)^2/9$. Finally, the sounds made by rectangular resonators are given by the formula $\omega_{(p,q,r)} = c/2[p^2/t^2 + q^2/w^2 + r^2/h^2]^{1/2}$ where c is the velocity of sound; l , w , and h are the length, width and height of the box, respectively; and p , q , and r are indexed from 0 (Richards, 1988). An algorithm based on Equation 1 can thus be constrained so that one of these patterns is used to control the partial frequencies ω_n . In addition, ω_1 can be specified such that $\omega_1 \propto (1/\text{size})$ to reflect the size of the object (this affects all the other partial frequencies).

The initial amplitude of the partials, Φ_n , can be controlled by a single parameter corresponding to mallet hardness. Recalling that Freed's (1990) results identified the ratio of high- to low-frequency energy as a predictor of perceived mallet hardness, we might maintain a linear relationship among the

TABLE 1
Mapping Parameters to Events

Term	Effect	Event Attribute
Φ_n	Initial amplitudes	Mallet hardness; force or proximity
$e^{-\delta_n t}$	Damping	Material
$\cos \omega_n t$	Partial frequencies	Size, configuration

partials' initial amplitudes, and use the slope from Φ_1 to control perceived hardness. Thus $\Phi_n = \Phi_1 + h(\omega_n - \omega_1)$, where h is the slope—note that h should often be negative, so that higher partials have less amplitude than low ones; thus a useful range of amplitude slopes might range from about -0.005 to 0.005 . Φ_1 (and thus all the amplitudes) may also be changed to indicate impact force or proximity.

Finally, the damping constants for each partial (δ_n) can be controlled by a parameter corresponding to material. A useful heuristic is to set $\delta_n = \omega_n D$, with D ranging between about .001 for metal and about .5 for a highly damped material such as plastic. This means that high harmonics will die out relatively quickly for highly damped materials and last longer for less damped materials (e.g., metal, which has low damping, tends to ring; wood, which is highly damped, tends to thunk). This strategy is suggested both by Wildes and Richards (1988) and by my own research (Gaver 1988; see Figure 3).

In sum, Equation 1 can be controlled by parameters that make effects corresponding to attributes of impact events. The pattern of partial frequencies corresponds to an object's configuration, whereas the overall frequency of the sound corresponds to the object's size. The pattern of partial amplitudes corresponds to mallet hardness, whereas the overall initial amplitude corresponds to the force or proximity of the impact. Finally, the degree of damping corresponds well to the virtual object's material. By controlling these five parameters, then, a wide range of sounds can be created which vary along perceptible dimensions of this basic-level sound-producing event; the mapping between these dimensions and the acoustic effects they produce embodies a hypothesis about ecological acoustics.

SCRAPING: SEPARATING OBJECTS AND INTERACTIONS

In the last section, I provided a fairly detailed description of impacts and the acoustic information they convey that benefited from the results of a relatively rigorous, quantitative physical analysis. In this section, I take a different approach to analyzing and synthesizing events, one reflecting the fact that the interactions and materials that make up sound-producing events are relatively independent from one another (Gaver, 1993). That is, it is possible to hit, crush, scrape, and perhaps roll the same object; conversely, a given interaction (e.g., scraping) may apply to any of a number of objects. This suggests that we may analyze and synthesize interactions separately from materials. Algorithms that take this approach allow the synthesis of a number of basic-level events involving solid objects.

The algorithm discussed in this section is based on a more qualitative analysis of impacts and scraping. Figure 5 shows an object being scraped across a rough

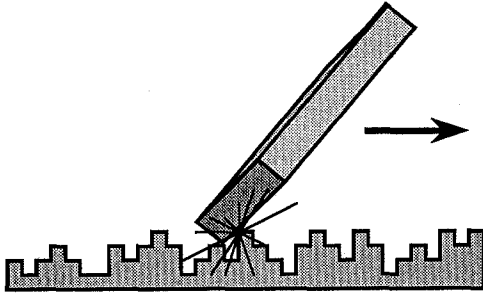


FIGURE 5 Scraping involves multiple impacts of an object against a textured surface; the waveform describing the force input into the object should be approximated by band-limited noise.

surface. This interaction may be thought of as a series of impacts as the object falls into slight depressions or hits raised elements, each of which will cause an impact sound like those discussed earlier. But because scraping is likely to involve many individual impacts occurring in a relatively short time, the series of impacts may be modeled as a continuous waveform describing the forces on the object. Thus objects may be described in terms of a characteristic set of resonant modes, and interactions by the waveform characterizing the forces they introduce to the object. This suggests that the acoustic specification of an object may be invariant over the interactions that cause it to make sound and, conversely, that there may be information for interactions that is invariant over objects.

The spectrogram presented in Figure 6 supports this hypothesis. This shows the sound made by a ceramic plate that has been hit and then scraped across a rough surface. The sounds made by hitting and scraping show very different temporal patterns, but the resonant frequencies of the object correspond well to those excited by hitting the object.

Because the effects of interactions and objects are separable, they can be modeled independently. The resonant modes of a virtual object may be modeled as a bank of band-pass filters (which only pass energy within a range of frequencies). Interactions, then, can be modeled in terms of energy passed

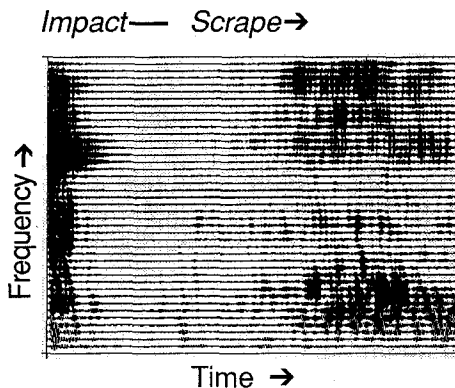


FIGURE 6 Spectrogram of the sound made by a ceramic plate as it is set on a hard surface and then scraped along it. Note that the resonant frequencies of the object (indicated by long-lasting dark regions on the spectrogram) are the same despite the different interactions.

through the filter bank. This strategy is similar to one applied commonly to speech synthesis and recognition, and has been applied elsewhere in studies related to ecological acoustics (Li, Logan, & Pastore, 1991; Richards, 1988).

Modeling Objects as Filter Banks

A simple formula for a bank of band-pass filters is:

$$y_n = \sum_m \Phi_m (c1_m x_n + c2_m y_{n-1} - c3_m y_{n-2}) \quad (2)$$

where Φ_m is an amplitude scalar for partial m , y_n is the n th output, x_n is the n th input, and:

$$\begin{aligned} c1_m &= (1 - c3_m)[1 - (c2_m^2/4c3_m)]^{1/2}, \\ c2_m &= (4c3_m \cos 2\pi f_m/S)/(c3_m + 1), \text{ and} \\ c3_m &= e^{-2\pi b_m/S}, \end{aligned} \quad (3)$$

with f_m the peak frequency and b_m the peak bandwidth of partial m , and S the sampling rate used for synthesis.

The same sorts of parameters can be used to control the peak frequencies and amplitudes of this sort of filter bank as are used to control the impact algorithm described in the last section. The pattern of peak frequencies reflects the object's configuration, the overall frequency indicates its size, and the pattern of amplitudes Φ_m can indicate mallet or surface hardness.

Material damping can be controlled via manipulations of the peak bandwidths of the filters. Bandwidth b is proportional to damping: The narrower the resonance peak of the filter, the longer the resonant response to excitation. This corresponds well to the characteristics of resonating materials: Highly damped materials such as wood have partials that are less well defined than those produced by less damped materials such as metal (Gaver, 1988; Wildes & Richards, 1988). Thus manipulating damping via the bandwidth of the filter actually simulates more information for material than simply manipulating the damping of sine waves, as was used in the previous algorithm.

Simulating Interactions With Input Waveforms

The characteristics of the filter bank described in Equation 2 defines a virtual object. The waveform passed through the filters models the interaction that causes the object to sound. For instance, the pattern of force generated by an impact is characterized by a short impulse such as one of those shown in Figure 7. Note that these impulses vary in duration and sharpness (or changes in amplitude). Thus the energy they contain is spread out over many frequencies: Low-frequency energy contributes to a given pulse's width, and high-frequency

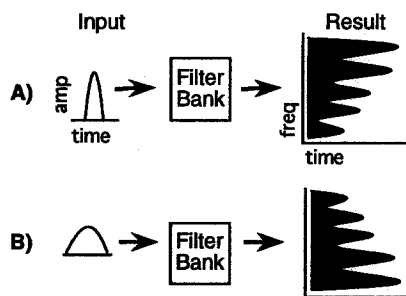


FIGURE 7 Sample impulse waveforms characterizing different impacts. A sharp impulse produces a greater proportion of high-frequency energy (A), corresponding to hard mallets. Rounded impulses produce more low-frequency energy (B), corresponding to soft mallets.

energy to its angularity. This corresponds to the physical effects of mallet hardness. Hard mallets introduce force suddenly to an object, deforming it quickly, and thus introduce a relatively high proportion of high-frequency energy to the resonant object. Soft mallets, on the other hand, introduce energy relatively slowly (because they deform as they hit the object), and thus the corresponding impulses are characterized by a relatively high proportion of low-frequency energy.

Scraping: A second basic-level event. Because the filter bank models an object independently from the interactions that cause it to make sound, this approach potentially allows the synthesis of many different sound-producing events. One that I have explored involves the sounds made by scraping. Relatively little is known about the physical attributes that are important for scraping sounds. Thus, the informal analysis of the physics of scraping presented earlier was the main basis for generating hypotheses that might be tested via synthesis.

As described earlier, as an object is scraped over a textured surface, a waveform describing the forces applied to the object is formed from the myriad of impacts involved. The nature of this waveform will depend on the texture of the surface, with regular surfaces producing repetitive waveforms (involving only a few frequencies of excitation) and irregular surfaces producing noisy waveforms. Most surfaces are likely to fall between these extremes, producing band-limited noise waveforms with a prominent center frequency depending on the predominant size of texture elements and the dragging speed, and a bandwidth that depends on the irregularity of the texture (Figure 8). These

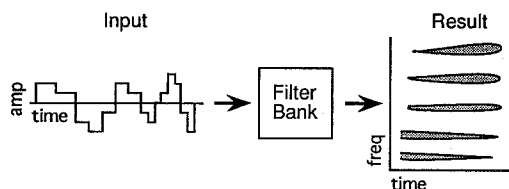


FIGURE 8 A sample force waveform characterizing a scrape with increasing speed. As the center frequency of the waveform increases, the proportion of high-frequency energy produced also increases.

parameters are approximate and less firmly grounded in previous physical or psychological research than those used to model impacts. Nonetheless, they allow the production of a wide variety of realistic scraping noises.

In sum, this algorithm is based on a physically plausible model of sound-producing events. By separating the parts of the model that specify the object from those specifying the interaction, a wide range of virtual sound-producing events can be simulated. The model can create any of the impact sounds that the algorithm described in the last section can. In addition, it can also be used to create a variety of scraping sounds (and, potentially, any other sound involving solid objects).

An important implication of this algorithm is that it allows us to generate the sounds of the same object being caused to sound by different interactions. This instantiates the claim that there is invariant information about objects over changes in interaction; and, conversely, that there is invariant information for interactions over changes in objects.

DRIPPING: A BASIC-LEVEL EVENT INVOLVING LIQUIDS

Impacts and scraping are both basic-level events involving solids. In this section, I consider dripping, a basic-level event involving liquids.

The physics of fluid dynamics is difficult, and analytic solutions for the physics of dripping are not as readily found in basic texts as analyses of mechanical vibrations. Nonetheless, the sounds made by dripping may be understood and instantiated in synthesis algorithms using the qualitative analysis described in Gaver (1993; see also Bragg, 1921).

Figure 9 shows an object falling into a liquid, producing a dripping sound. As the object hits the liquid, it pushes it aside, forming a cavity that resonates to a

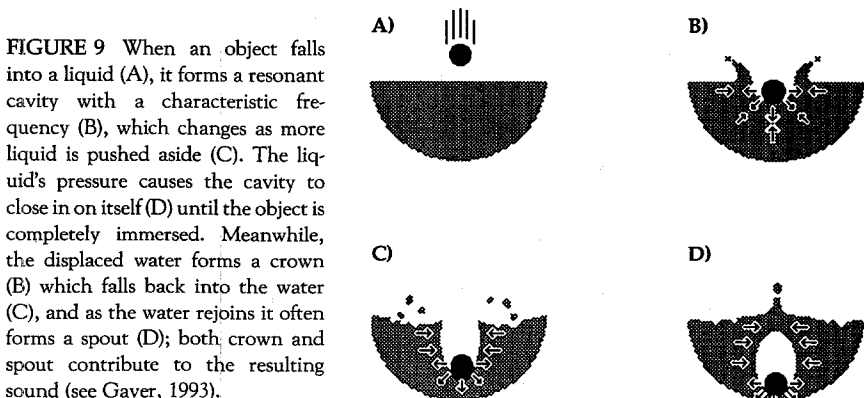


FIGURE 9 When an object falls into a liquid (A), it forms a resonant cavity with a characteristic frequency (B), which changes as more liquid is pushed aside (C). The liquid's pressure causes the cavity to close in on itself (D) until the object is completely immersed. Meanwhile, the displaced water forms a crown (B) which falls back into the water (C), and as the water rejoins it often forms a spout (D); both crown and spout contribute to the resulting sound (see Gaver, 1993).

characteristic frequency, amplifying and modifying the pressure wave formed by the impact itself. The cavity grows as the object displaces more liquid, and thus the resonant frequency decreases. But the liquid's pressure causes it to close in on the cavity, decreasing the size of the resonant cavity, until the object is ultimately immersed and the sound stops.

On the basis of this qualitative analysis, a synthesis algorithm was developed to model dripping sounds. The algorithm depended on an analogy between the cavity formed behind a dropping object and a simple Helmholtz resonator, an enclosed body of air joined to the atmosphere by an aperture (Figure 10; see Elmore & Heald, 1969). Such resonators have a single resonant frequency which is inversely proportional to the volume of air enclosed, as long as the dimensions of the resonator are small compared with the wavelength associated with that frequency. Thus one might expect that a dripping sound could be modeled simply as a short-lived, pulsed sine wave at some appropriate frequency.

Unfortunately, experience shows this strategy to be inadequate. The sounds produced by the algorithm could be heard as dripping sounds, but they could also be confused with sounds made by other events (e.g., simple impacts). Adding smooth changes to the frequency of the wave to reflect the changing size of the cavity did not improve the result. Clearly, these algorithms did not capture sufficient information, and thus a more detailed description of dripping was sought.

Closer examination revealed that as the object enters the body of liquid, a "crown" often forms from the displaced liquid around the cavity which disintegrates into smaller droplets each of which creates a small resonating cavity as it falls. Similarly, a "spout" may form in the middle of the closing cavity as the object is immersed, which again produces a sound as it falls.

Based on this analysis, I have developed an algorithm for synthesizing dripping sounds that works by creating a short-lived pitched impulse followed quickly by several higher pitched, even shorter lived ones. The details of such sounds are likely to be influenced by many factors, particularly the mass, size, and speed of the object and the viscosity of the liquid, all of which influence the evolution of the resonating cavity and the formation and separation of the

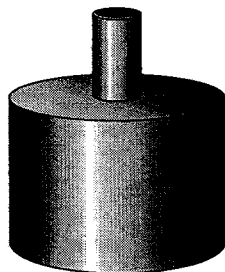


FIGURE 10 A Helmholtz resonator, an enclosed body of air joined to the atmosphere by an aperture, was used as the basis for an initial analysis of dripping sounds.

crown and spout. But even without such a detailed quantitative analysis of the physics of this event, the algorithm produces realistic dripping sounds.

The development of this algorithm illustrates the complementary roles analysis and synthesis may play in studies of everyday listening. The first algorithm was guided by a physical analysis, but this failed to capture an important aspect of the event. This inadequacy was discovered by listening to the results of the algorithm. Synthesis thus motivated another iteration in the attempt to understand the relevant physics, one resulting in an algorithm that succeeds in producing dripping sounds. Its success suggests that such qualitative analyses, when coupled to synthesis, may be sufficient for early exploration of such events.

TEMPORALLY COMPLEX EVENTS: BREAKING, BOUNCING, AND SPILLING

In the companion article to this one (Gaver, 1993), I suggested that complex sound-producing events can be understood in terms of their basic-level components. Thus the algorithm for creating impact sounds described earlier may serve as the basis for algorithms modeling more complex sounds involving impacts. In this section, I describe an algorithm for creating bouncing, breaking, and spilling sounds based on the work of Warren and Verbrugge (1984). This algorithm may be seen as an example of the relation between basic-level and more complex events.

Warren and Verbrugge's (1984) study of breaking and bouncing sounds is an early example of analysis and synthesis of everyday sounds. In this study, they used physical and acoustic analyses to examine the auditory patterns that characterize breaking and bouncing, and verified their results by testing listeners on synthetic sounds.

Consider the mechanics of a bottle bouncing on a surface (see Figure 11, Panel

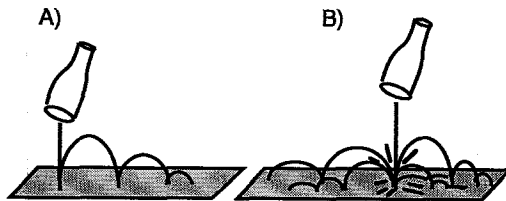


FIGURE 11 Bouncing (A) and breaking (B) sounds are characterized by the temporal patterning of a series of impacts.

Note. From "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics" by W. Warren and R. Verbrugge, 1984, *Journal of Experimental Psychology: Human Perception and Performance*, 10, p. 705. Copyright 1984 by the American Psychological Association. Adapted by permission.

A). Each time the bottle hits the surface, the impact causes the bottle to vibrate in a characteristic way depending on its shape, size, and material (as discussed in the earlier section on impact sounds). Energy is dissipated with each bounce so that, in general, the time between bounces and the force of each impact decreases (some irregularities in the pattern are likely to occur due to the bottle's asymmetry). Thus, bouncing sounds may be expected to be characterized by a repetitive series of impact sounds with decreasing period and amplitude. When a bottle breaks, on the other hand, it divides into many separate pieces with various sizes and shapes (see Figure 11, Panel B). Thus, a breaking sound should be characterized by an initial impact sound followed by several different, overlapping bouncing sounds, each with its own frequency makeup and period. The differences between breaking and bouncing, then, should be conveyed largely by the temporal patterning of the sounds.

This qualitative physical analysis is born out by acoustic analyses of natural tokens of breaking and bouncing sounds. Spectrograms of recorded bouncing sounds show a series of impacts, each with identical frequency components, which repeat at increasing rate. Spectrograms of breaking sounds, on the other hand, show a more complex pattern; individual bouncing patterns of the pieces are overlapped but still may be distinguished. Spectral components may play a role in distinguishing different bouncing patterns, but temporal patterning seems the most salient distinguishing feature between the sounds.

Warren and Verbrugge (1984) created artificial tokens of breaking and bouncing sounds by combining natural tokens of single impacts in various temporal patterns. The sounds made by four individual pieces of a broken bottle were recorded separately. To create an artificial bouncing sound, the individual sound tokens were synchronized to the timing of a real bouncing bottle, so that all four played simultaneously. To create an artificial breaking sound, each of the four component sounds was synchronized to a different bouncing pattern (taken from a natural bouncing-bottle sound), so that they were not in phase. Thus, the spectral components of the artificial breaking and bouncing sounds were identical, and they could only be distinguished by their temporal patterning.

To validate this analysis, listeners categorized both natural and artificial bouncing and breaking sounds as *bouncing*, *breaking*, or *don't know*. Listeners were about 99% correct for natural bouncing and breaking sounds, 93% correct for artificial bouncing, and 87% correct for artificial breaking sounds. Clearly, Warren and Verbrugge's (1984) characterization captures sufficient information to distinguish bouncing and breaking.

Two things should be noted about these results. First, the constructed breaking sounds were quite simple: Natural breaking sounds are often characterized by many more than four overlapping bouncing sounds. Second, the task was constrained, offering participants a limited choice of hypotheses about what they might have heard. Listeners were offered a third category (*don't know*) to try

to reduce this constraint. Still, they might have considered a particular sound to be more representative of breaking than bouncing, for instance, without it necessarily sounding like breaking. Nonetheless, that these sounds allowed constrained identification suggests that Warren and Verbrugge (1984) did capture important aspects of the information available about these events.

Synthesized Breaking, Bouncing, and Spilling

The impact algorithm described in the first section can be used to explore the results of Warren and Verbrugge's (1984) study. To create bouncing sounds, the impact algorithm is embedded in another that calls it at exponentially decaying intervals. To create breaking sounds, the bouncing algorithm is embedded in another algorithm that calls it with parameters specifying sources of different sizes at times corresponding to several exponentially decaying time series.

Several new event parameters become relevant for these algorithms: The initial height of the virtual object is indicated by the time between the first and second bounce, its elasticity by the percentage difference of delays between bounces, and the severity of breaking by the number of pieces produced. In addition, the asymmetry of the perceived object can be varied by adding randomness to the overall temporal pattern.

Sounds synthesized using this algorithm make it clear that although Warren and Verbrugge's (1984) analysis suggested that information for breaking and bouncing depends only on temporal patterning, the perceived event also depends on the materials involved. For instance, if impacts specifying wooden objects are produced in a temporal pattern typical of breaking, the result sounds like spilling several objects rather than breaking a single one. Similarly, if each of several virtual objects has different material properties, spilling is heard again instead of breaking.

In sum, the impact algorithm described earlier can be used not only to generate the sounds made by mallets of different hardnesses striking virtual objects of a wide variety of shapes, sizes, and materials, but it can serve as the basis for more complex sounds such as those made by bouncing, breaking, and spilling. As such, it allows us to explore the space of such sounds to discover the constraints on perceived events placed by materials. Thus, the algorithm not only instantiates information for everyday listening, but serves as a research tool in its own right.

MACHINE SOUNDS

In this section, I illustrate the power of informal physical analyses by describing complex machine sounds. This work explores the possibility that high-level

attributes of such sounds may be captured without modeling the attributes of their basic-level components.

A detailed account of the mechanical physics of even a simple machine seems prohibitively difficult. Even small appliances consist of a myriad of interacting parts, each of which contributes to the overall pattern of sound. Trying to analyze and then synthesize the effects of these components precisely appears to be intractable. However, just as the scraping model already described models the overall parameters of the complex force applied to an object rather than each of the impacts making up that force, so an approximate model of machine sounds might seek to capture some of the high-level characteristics of the sounds they produce. In particular, three aspects of machine sounds seem particularly relevant for modeling: (a) The overall size of the machine is likely to be reflected in the frequencies of sounds it produces, (b) most machines involve a number of rotating parts which can be expected to produce repetitive contributions to the overall sound, and (c) the work done by the machine might be expected to affect the bandwidth of the sound, much as the force of an impact determines the bandwidth of the sound it produces.

Using Frequency Modulation Synthesis to Create Machine Sounds

I have developed an efficient algorithm for creating a variety of machine-like sounds that capture these properties. The basic strategy is to synthesize a sound with one or more complex tones which vary in a repetitive way, indicating cyclical motion. The rate at which the virtual machine is working, then, can be indicated by repetition speed, the size of the virtual machine by the base frequency, and the amount of work by the bandwidth of the sounds (see Figure 12).

This class of sound may be synthesized efficiently using simple frequency modulation (FM) synthesis (Chowning, 1973). FM synthesis involves modulating the frequency of a carrier wave with the output of a modulating wave. This produces a complex tone with a number of frequency components spaced equally around the carrier wave. These components are separated from one another by the modulating frequency; their number (and thus the bandwidth of the tone) depends on the amplitude of the modulating wave (see Figure 13).

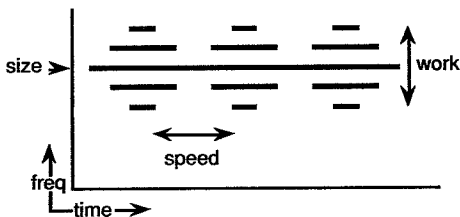
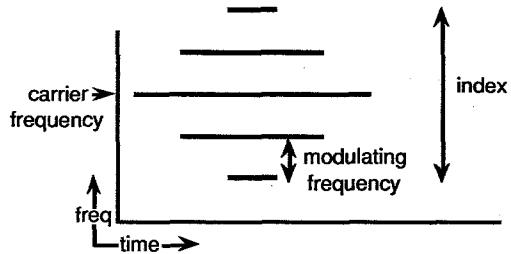


FIGURE 12 Machine sounds can be characterized by a complex wave which varies repetitively over time. Machine size is indicated by the center frequency, speed by repetition time, and work by bandwidth.

FIGURE 13 FM synthesis is well suited for generating machinelike sounds. The central frequency component is determined by the carrier frequency, the spacing between partials by the modulating frequency, and the bandwidth by the index.



Thus, the machine sound shown in Figure 12 can be created simply by associating the carrier frequency with the size of the virtual machine, setting the maximum amplitude of the modulator to the amount of work done by the virtual machine, and modulating the amplitude of the modulator according to the speed of the virtual machine, as shown in this pseudocode:

```
/* Pseudocode for generating machinelike sounds */
```

```
mod.wave.amp = work * sin(speed * time/samprate);
mod.wave.sample = mod.wave.amp * sin(mod.freq * time/samprate);
```

```
output = amp * sin((size + mod.wave.sample) * time/samprate);
```

The resulting sounds are pitched humming noises which pulse at the speed of the virtual machine. When “work” is low, the throbbing is subtle; when it is high, it becomes quite pronounced. Moreover, the quality of the sounds can be varied by changing the ratio of modulating to carrier frequency: When the two are an integral multiple of one another, the resulting sound is harmonic; when they are not, the sound is inharmonic or noisy. Using this algorithm, a wide variety of machinelike sounds can be produced, varying from low-pitched, subtle hums to higher, more insistent roars.

This algorithm does not take into account attributes of the basic-level events that constitute a running machine. This is reflected in the sounds it produces: Whereas the other algorithms produce sounds that clearly indicate details of their virtual sources (such as the material and size of the objects involved), the sounds produced by this algorithm only hint at some of the high-level properties of their supposed sources. Nonetheless, the sounds produced by this algorithm do seem to capture some of the information conveyed by machinery. Insofar as detailed information is left out, they may be thought of as “cartoon sounds,” sounds that caricature some aspects of the events while omitting others. This makes them an interesting example to consider here, suggesting the possibility of modeling some complex attributes of events without capturing supposedly simpler details.

CONCLUSIONS

The algorithms described in this article instantiate hypotheses about how we hear events and their attributes. The claim is that, in offering the synthesis of a variety of everyday sounds specified in terms of the events that cause them, the algorithms capture the information these sounds convey about events. The basic test of this claim is in the sounds the algorithms produce: I have described the algorithms here in enough detail that readers can implement them and hear the results themselves.

Several questions concerning evaluation, specification, and the tension between formal and informal physical analyses remain open, however. As pointed out earlier, using analysis and synthesis of events as a methodology for exploring acoustic information has allowed informal listening to guide the development of these algorithms. Nonetheless, more formal empirical tests of these algorithms are clearly desirable. On the face of it, such tests seem to be straightforward, but closer examination shows that the design of experimental methodologies raises issues concerning specification and physical analyses. Preconceptions about specification are reflected in experimental methodologies; conversely, the experimental methods used may restrict what can be said about how sounds specify their sources and about the justification of the underlying physical analyses.

First, it is important to distinguish between simulating an event to produce a pattern of sound and characterizing the sound pattern allowing that event to be recognized. An algorithm for producing a sound based on a description of an event cannot always be converted easily into one meant to recognize the event based on a description of the sound. For many of the algorithms presented here, this conversion is relatively straightforward. But one, at least, is problematic: The scraping algorithm uses a bank of filters to model the resonant modes of an object, and a band-limited noise to model the forces exerted on that object when it is scraped over a surface. Because the scraping sounds emerge from the interaction of these two components, the acoustic pattern for scraping itself has not been modeled. The algorithm produces sounds that people hear as scraping, but that does not mean that the acoustic information for their perception has been made explicit. For this, it will be necessary to describe the acoustic patterns themselves rather than the algorithm for producing them.

Second, there is also an important distinction between identifying an acoustic attribute that provides information for a particular event and one that *specifies* it. Specification implies a one-to-one mapping between an event and a pattern of sound. Saying that some acoustic pattern specifies an event carries an implicit claim that no other event could cause that sound—a claim that seems difficult to evaluate, given that it implies rejecting all other events as possible sources of the same sound. Empirical studies, nonetheless, should reflect the possibility that a synthesized sound based on a model of one event might sound as if it had been made by another. From this perspective, tasks requiring listeners to assign

sounds to predefined categories, as used, for example, by Warren and Verbrugge (1984), are problematic because they manipulate listeners' expectations of the events they will hear and constrain their eventual identifications. Free-identification tasks, in which subjects are asked to describe a sound-producing event with no a priori categories being provided, are more likely to indicate that a sound might be produced by several possible sources.

A problem for free-identification studies, however, comes in interpreting situations in which listeners describe a sound in terms of a number of alternative sources. In this case, two possibilities hold. The first—and more worrying from an ecological point of view—is that there is a one-to-many mapping between the sound and the events or parameter under consideration. In this case, the sound may provide information about an event, but it does not specify it. For instance, in the impact algorithm described here, a change in an object's length produces a change of the fundamental frequency of the sound it makes. But other attributes of an object, such as its shape, density, and hardness, also determine its functional frequency. Thus, an impact sound's fundamental frequency does not specify its length. Nonetheless, it may provide information about its length in a more limited sense, either through *constrained specification*, or *specified constraints*.

Constrained specification refers to situations in which perceptible patterns specify events if other conditions are met (these are termed *ecologically incomplete invariants* by Runeson, 1989). For instance, the fundamental frequency of an impact sound does specify an object's length if other attributes of the object are held constant. This might be tested empirically by comparing listeners' judgments of length when other attributes are held constant or changed. Similarly, fundamental frequency might be used by listeners to assess length if it is only affected to a small degree by other attributes varying within their typical values. This could be evaluated both by estimating the common range of frequency shifts caused by other attributes and by investigating whether listeners spontaneously use frequency to judge length.

Specified constraints, on the other hand, apply to situations in which perceptible patterns do not uniquely specify one event or parameter value, but do constrain possibilities in meaningful ways. For instance, the lowest perceptible frequency produced by an object may be higher than the object's fundamental frequency (e.g., if the object is very large, its fundamental frequency will be too low to be heard). This implies that the lowest frequency heard for a particular sound may not specify the size of an object. However, the lowest frequency heard does specify a constraint about the object: Because large objects may produce high frequencies, but small ones cannot produce low ones, the lowest frequency heard specifies a minimum size for the object. Specified constraints such as these seem easily amenable to experimental evaluation.

The second possibility, when listeners describe a sound in terms of several alternative events, is that they are focusing on an inappropriate level of detail.

The possibility of one-to-one mappings between sounds and events depends in part on the way events are characterized. When several traditionally defined, primitive physical properties of an event have the same effect on the sound it produces, it is possible that a single perceptual dimension is heard, one that incorporates all of them. For instance, we may not hear the several attributes that affect fundamental frequency separately, but rather a new dimension (perhaps moment of inertia) which combines all of them. Similarly, material *per se* does not exist as for mechanical physics, but instead is separated into many other dimensions such as density, elasticity, and homogeneity. Nonetheless, people do seem to hear the material of a struck object, rather than these other properties (Gaver, 1993). Treating constructs like material as unitary attributes, as I do in this article, has pragmatic implications—it allows the use of relatively qualitative analyses, though it also makes justification of the models more difficult—but also reflects a move toward an ecological acoustics.

One way to justify the incorporation of several primitive physical parameters into a single, higher level attribute such as material is to use analysis and synthesis together as tools for research. Just as analyses of the physics and acoustics of sound-producing events may inform synthesis, so the results of synthesis may inform physical and acoustic analyses (as it did for the dripping algorithm described earlier). This strategy allows higher level physical attributes to be used for synthesis, if listeners characterize the resulting sounds in terms of those attributes. Analysis and synthesis may thus be seen as a means of bootstrapping one's way from traditional to ecological physics, as listening provides the basis for modifying algorithms based on traditional physical analyses to reflect higher level, ecological analyses.

In the end, experimental evaluations of the claim that these algorithms convey acoustic information about events will also help to justify the move toward an ecological acoustics. As suggested by the previous discussion, a variety of experimental methods will be helpful in evaluating these algorithms. Free-response tasks are useful in seeking evidence that the sounds produced by these algorithms are spontaneously heard in terms of the relevant events. More constrained categorization, comparison, and rating tasks can be used to induce listeners to focus on information for particular attributes at a desired level of detail. Finally, to increase ecological validity and avoid problems associated with conscious verbalization, tasks might be designed in which the sounds are assessed in terms of their effects on action. All these techniques are needed to fully assess these algorithms: Any one may not provide complete information, but all together may converge to a satisfactory account.

Issues of evaluation, specification, and ecological physics are difficult and must remain open. For the time being, though, this work may be judged pragmatically. The algorithms do produce quite realistic sounds, despite variations in the formality of the physical analyses used in their construction. Listeners comment that they have a strong impression of an actual event causing

them, rather than hearing them as synthesized. Insofar as some of the sounds (notably those made by the machine algorithm) do differ from those made by real events, they may be considered as "cartoon sounds" which capture some of the relevant features of their sources and omit others, just as visual caricatures capture some defining features while leaving out incidental ones. Thus, these algorithms represent an important step toward understanding the acoustic information for various sound-producing events.

In addition, these algorithms may be applied to the creation of *auditory icons*, everyday sounds mapped to computer events by analogy with everyday sound-producing events (Gaver, 1989, 1993). Developing effective auditory icons has been difficult because standard synthesis techniques are aimed at creating musical sounds, whereas sampling (digital recording) techniques are constrained in the manipulations of sound they offer. For the purpose of creating auditory icons, these algorithms show great potential, combining flexibility, intuitive controls, efficiency, and relevance. They offer many possibilities for the design of auditory interfaces that we can listen to the way we do to the everyday world.

Of course, there is still much to learn about how we do listen to the world. I have described the algorithms presented here in sufficient detail to guide readers in implementing and exploring them. I do this in the hope that the algorithms will spur further research on ecological acoustics; that readers will implement these algorithms, listen to the results, test, and extend them. For just as the map of everyday sounds presented in Gaver (1993) is only a start toward understanding what we hear, so these algorithms are only a first step toward describing how we hear it: There remains a wealth of opportunity for pursuing an ecological approach to everyday listening.

ACKNOWLEDGMENTS

I am grateful to Dave Woodhouse and Roy Patterson for help with the physical and acoustical analyses presented here and to Anne Schlottman, Don Norman, Michael Beaudouin-Lafon, William Mace, Allan Maclean, Wendy Mackay, and Michael Turvey for valuable discussions about this research. Finally, I thank the reviewers of this article, Gary Kidd and Brad Rakerd, for their thoughtful and constructive critiques.

REFERENCES

- Borin, G., De Poli, G., & Sarti, A. (1993). Algorithms and structures for synthesis using physical models. *Computer Music Journal*, 16, 30-42.
- Bragg, W. (1921). *The world of sound*. London: Bell.
- Chowning, J. (1973). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21, 526-534.

- Cremer, L. (1984). *The physics of the violin* (John S. Allen, Trans.). Cambridge, MA: MIT Press. (Original work published 1981)
- Elmore, W., & Heald, M. (1969). *Physics of waves*. New York: Dover.
- Freed, D. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*, 87, 311-322.
- Gaver, W. (1988). *Everyday listening and auditory icons*. Unpublished doctoral dissertation, University of California, San Diego.
- Gaver, W. (1989). The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4, 67-94.
- Gaver, W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5, 1-29.
- Gibson, J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1979)
- Gibson, J., & Gibson, E. (1982). Perceptual learning: Differentiation or enrichment? In E. Reed & R. Jones (Eds.), *Reasons for realism: Selected essays of James J. Gibson* (pp. 317-332). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1955)
- Jaffe, D., & Smith, J. (1983). Extensions of the Karplus-Strong plucked string algorithm. *Computer Music Journal*, 7, 56-69.
- Lamb, H. (1960). *The dynamical theory of sound* (2nd ed.). New York: Dover.
- Li, X., Logan, R., & Pastore, R. (1991). Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America*, 90, 3036-3049.
- McIntyre, M., Schumacher, R., & Woodhouse, J. (1983). On the oscillations of instruments. *Journal of the Acoustical Society of America*, 74, 1325-1345.
- Repp, B. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America*, 81, 1100-1109.
- Richards, W. (1988). Sound interpretation. In W. Richards (Ed.), *Natural computation* (pp. 301-308). Cambridge, MA: MIT Press.
- Risset, J., & Wessel, D. (1982). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (pp. 25-58). New York: Academic.
- Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, 18, 172-179.
- Runeson, S. (1989, September). A note on the utility of ecologically incomplete invariants. *ISEP Newsletter*, pp. 6-9.
- Warren, W., & Verbrugge, 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., & Richards, W. (1988). Recovering material properties from sound. In W. Richards (Ed.), *Natural computation* (pp. 356-363). Cambridge, MA: MIT Press.

APPENDIX

An Efficient Algorithm for Synthesizing Exponentially Decaying Cosine Waves

Equation 1 (shown earlier) is useful in allowing parameters to be defined in terms of source events. It is also attractive because it can be implemented in a computationally efficient way.

An efficient implementation of this formula relies on Euler's relationship:

$$e^{i\omega t} = \cos\omega t + i\sin\omega t. \quad (\text{A1})$$

This implies that for a single logarithmically decaying cosine wave:

$$e^{-\delta t} \cos\omega t = \text{Re}[e^{-\delta i\omega t}]. \quad (\text{A2})$$

Because $t = nT$, where T is the sampling interval, Equation 1 can be expressed as:

$$S_n = \text{Re}[(e^{-\delta i \omega})^T]^n. \quad (\text{A3})$$

where S_n is the value of the n th sample. Or:

$$S_n = \text{Re}[S_{n-1} * \lambda], \text{ where } \lambda = e^{(-\delta i \omega)T}. \quad (\text{A4})$$

Now, $S_n = a_n + ib_n$, where $S_0 = a_0$ is the initial amplitude, and $b_0 = 0$, and $p + iq = 1$, where $p = e^{-\delta t} \cos \omega t$ and $q = e^{-\delta t} \sin \omega t$. Thus from A3:

$$\begin{aligned} S_n &= \text{Re}[(a_n + ib_n)(p + iq)] \\ &= \text{Re}[(a_n p - b_n q) + i(b_n p + a_n q)]. \end{aligned} \quad (\text{A5})$$

This means that samples can be generated by first calculating p and q and setting a and b to the initial amplitude and 0, respectively, and then iteratively applying Equation A5. The output sample is then the real part of the result, and a and b are updated to the real and imaginary parts of the result, respectively. Consider the following pseudocode:

/* Pseudocode for generating logarithmically decaying cosine wave */

```
p = cos(freq * 1/samplerate) * power(e, -1 * damping.rate * 1/samplerate);
q = sin(freq * 1/samplerate) * power(e, -1 * damping.rate * 1/samplerate);
a = initial.amplitude;
b = 0;
```

repeat for duration.in.secs / samplerate:

```
  anew = a * p - b * q;
  bnew = b * p + a * q;
  a = anew;
  b = bnew;
```

```
  output = anew;
end repeat;
```

Computationally expensive sine and cosines need only be figured once for the calculation of p and q , and only four multiplies, one add, and one subtract are needed for each partial for a given sample. The efficiency of this implementation allows fairly complex impact sounds to be generated in real time on many computers.
