Chapter 5:

EVERYDAY LISTENING

Introduction

Traditional approaches to acoustics and psychoacoustics have provided a number of valuable ways to understand audition, as we have discussed in Chapter X. Moreover, the vocabulary they offer for describing sound and hearing can be applied in explorations of multidimensional data and in creating musical messages, as discussed in Chapters X and X. Nonetheless, such approaches often seem inadequate to describe our everyday experiences of listening to the world around us. In this chapter we introduce an alternative perspective from which listening can be understood. In the next chapter, we show how this approach can be applied, leading to novel methods for using sounds at the interface.

To understand the shortcomings of traditional approaches to audition, listen to a few nonspeech, nonmusical sounds (such as sound examples N - N) and try to describe them in psychoacoustic terms. For each sound, ask yourself: What is its pitch? Its loudness? How would you describe its timbre? Is it discrete, rhythmic or continuous? Do these dimensions even make sense?

Sound examples N - N: Environmental sounds are difficult to describe in psychoacoustical terms; easy to describe in terms of their sources. These sounds -- or, more exactly, the experiences they evoke -- are good examples of everyday listening.

Many sounds prove difficult to describe using the kind of vocabulary suggested by traditional approaches to psychoacoustics. What is the pitch of crumpled paper, or the loudness of a passing airplane? Is the timbre of splashing water rough or smooth? Are the sounds made by a breaking bottle discrete, repetitive or continuous? Many of the dimensions that seem simple to apply to musical sounds become much more troublesome when we try to apply them to the sounds we hear in our everyday lives.

The problem goes beyond ease of description. Certainly one can describe such sounds fully, even if the descriptions may have daunting complexity. Consider, for example, sound example X. We might describe this sound as a repetitive (though not quite regular) series of band-limi1ted noises, each with a fairly sharp attack and quick decay. We might produce a spectrum of the sound, note its temporal progression, perform experiments in which people match its perceived pitch with that of a standard, or ask people to describe it along dimensions such as “sharp - dull” or “smooth - rough.” But is the result of this sort of analysis really an adequate description of the sound?

Now listen to the sounds again, and instead of trying to produce a psychoacoustical description of the sounds, simply ask yourself “what do I hear?” The task should be much easier now -- you hear a plane flying by, a bouncing ball, and so forth. In the case of sound example X, you don’t hear band-limited noises and so forth, you simply hear somebody climbing a flight of concrete steps, turning on the landing, and climbing another flight. The point is a simple one: Psychoacoustics allows us to describe and understand *sound* in enormous detail, but there is more to listening than sounds alone.

Musical and Everyday Listening

Hearing the pitch of a sound or its loudness is an example of *musical listening*. But we often hear *events*, rather than sounds. Listening to airplanes, water, birds and footsteps are examples of *everyday listening*. This is a different sort of experience than that described by traditional psychoacoustics. Instead of being concerned with our ability to perceive attributes of sounds themselves -- their frequency, spectral content, amplitude, etc. -- everyday listening is a matter of listening to the attributes of events in the world -- the speed of a passing automobile, the force of a slammed door, whether a person is walking up or downstairs.

Musical andEveryday Listening are Experiences
Note that the distinction between everyday and musical listening is between experiences, not sounds. It is possible to listen to any sound either in terms of its attributes *per se* or in terms of the event that creates it. For instance, while listening to a string quartet we might be concerned with the interplay of pitches, the juxtoposed timbres, or the intricacies of the rhythm -- the patterns of sensation the sounds themselves evoke. This is an example of musical listening. Alternatively, we might listen to the instruments themselves -- is that the viola playing that line, or the cello? Is the bow frayed on that violin? In this case we are

concerned with identifying the sources of sounds, and properties of those sources. This is an example of everyday listening.

On the other hand, while walking down a city street we are likely to listen to the sources of sounds -- is that car heading our way? How close is that guy walking behind us? Most of our experience of hearing the day-to-day world is one of everyday listening: we are concerned with knowing about the events going on around us, what is important to avoid and what might offer possibilities for action. But occasionally we might listen to the world as we do music -- to the humming pitch of a ventalator punctuated by a syncopated birdcall, to the interplay and harmony of the sounds around us. This may seem an unusual experience to many of us. But hearing the everyday world as music is one way to understand what John Cage (xxx), for instance, is attempting in his compositions. In presenting traffic sounds in a concert setting, he is trying to evoke an experience of musical listening to nonmusical sounds. The distinction between everyday and musical listening is fundamentally one between experiences, not sounds.

Nonetheless, some sounds seem more likely to evoke one experience than another. The tones made by many musical instruments, for instance, convey relatively little information about their source. Instead their most salient features are their pitch, their duration, etc. Hearing such sounds seems to throw the listener into an experience of musical listening. Other sounds, though, are difficult to listen to musically. It may be possible to hear the time-varying pitch of a breaking bottle, for instance, but the source of the sound seems much more compelling. So we may talk about *everyday* and *musical sounds*, but we must be careful. For it is not the type of sound we are interested in, but the type of experience; not the sound's attributes, but whether the attributes of interest are those of the sound or those of the source (this will become important in considering applications of everyday listening).

The Psychology of Everyday Listening

The experience of everyday listening, if taken seriously, has the potential to produce a radically new explanatory framework for understanding sound and listening. Such a framework would allow us to understand listening and manipulate sounds along dimensions of sources rather than sounds. So for example, we might imagine a psychoacoustics concerned with measuring people's ability to hear the forces involved in events, or a synthesizer that allowed us to specify whether a sound source was wood or metal.

Understanding sound in terms of everyday listening complements the account offered by traditional psychoacoustics. Clearly the psychoacoustical phenomena we have discussed are valid whether we are talking about everyday or musical listening: A loud sound will mask a soft one, whether I am concerned with the soft one having a certain pitch or with the size of its source. But a new framework may alter some of the questions about sound and hearing we consider important. It certainly should allow us to refer more directly to attributes of everyday listening than does our present understanding of psychoacoustics.

Remarkably little is known as yet about everyday listening. We know little about how to characterize the fundamental attributes of sources that we hear, or about the acoustic cues that inform us about events in the world. Despite centuries of study of sound and hearing, we don't really know what distinguishes the sounds of somebody going upstairs rather than down.

Why Is So Little Known About Everyday Listening?

There are two reasons for our ignorance, one historical and the other theoretical. Historically, studies of acoustics and psychoacoustics have been guided largely by a concern with understanding music and the sounds produced by musical instruments. From the ancient Greeks' discovery that doubling the length of a vibrating string halves its pitch, through to Helmholtz' studies of the harmonic structure of musical sounds, and even to present day studies of computer music, the major thrust of disciplines concerned with nonspeech audio has been to use musical sounds and to understand musical listening.

But an account of hearing based on the sounds and perceptions of musical instruments often seems biased and difficult to generalize. Musical sounds are not representative of the range of sounds we normally hear. Most musical sounds are harmonic; most everyday sound inharmonic or noisy. Musical sounds tend to have a smooth, relatively simple temporal evolution; everyday sounds tend to be much more complex. Musical sounds seem to reveal little about their sources; while everyday sounds often provide a great deal of information about theirs. Finally, musical instruments afford changes of the sounds along relatively uninformative, musical dimensions such as pitch or loudness, while everyday events involve many more kinds of changes -- changes that are often musically useless but pragmatically important. Our current knowledge about sound and hearing has been deeply influenced by the study of a rather ideosyncratic subset of sounds and sources. It is interesting to turn to a wider variety of sounds and sources in driving a study of everyday listening.

Theoretically, studies of everyday listening have been constrained by the supposed primitives of sound and by sensation-based theories of perception. Physical descriptions of sound are dominated by those suggested by the Fourier transform: frequency, amplitude, duration and so on. Psychoacoustics has traditionally taken these dimensions as the physical primitives that correspond to elemental sensations. The end result, then, is an acoustics and psychoacoustics which emphasizes physical and perceptual dimensions best suited for describing music.

Traditional explanations of psychophysics take these "primitive" physical dimensions as their elemental stimuli and use them to motivate the identification of corresponding "elemental" sensations. From this perspective, more complex perceptions must depend on the integration of elemental sensations -- but often, sensations seem inadequate to specify complex events. Thus traditional approaches argue that often there is a paucity of information in available stimuli, and that veridical perception must depend on representations of the world based largely on memory and "unconscious inference" or problem-solving. The upshot of this approach then, is a strategy in which the first concern is with elemental sensations, the explanation of which can later serve as the building blocks of an integrative account of perception. But sensations, the very objects of perception according to these accounts, play only a small role in understanding our experience of the everyday world. Instead, perception depends on learning, memory and problem-solving: perception isn't perception at all.

The Ecological Approach To Perception

Taking everyday listening seriously as a domain for studies of acoustics and psychoacoustics suggests that we may broaden the range of physical parameters and perceptual experiences to be considered. For instance, we might add new perceptual dimensions such as size or force to the attributes of pscyhoacoustics, and understand them in terms of their acoustic covariates. Such an endeavour implies that traditional psychoacoustics needs to be stretched in two ways: First, the perceptual dimensions we need to study concern those of sources as well as sounds, and second, we must be willing to treat apparently complex acoustic variables as elemental.

This approach is guided and inspired by the ecological approach to perception developed by Gibson (1979. The ecological perspective counters many of the assumptions of traditional accounts of perception. According to the ecological approach, perception is usually of complex events and entities in the everyday world. Moreover, it is *direct*, unmediated by inference or memory. According to this perspective, elemental stimuli for perception do not necessarily correspond to primitive physical dimensions but may instead be specified by complex *invariants* of supposedly primitive features. Thus complex perceptions rely on complex stimuli (or “perceptual information”), not on the integration of sensations. From this point of view, there is rich and varied information in the world, both because our descriptions are no longer limited to primitive physical dimensions and because *exploration* or the world -- as opposed to passive stimulation -- becomes an important component of perception. Thus, according to the ecological account, the study of perception should be aimed at uncovering perceptual information and the ecologically relevant dimensions of perception.

Developing An Ecological Account Of Listening

There has been little development of an ecological account of audition as yet. Gibson did make some preliminary observations about listening (xxx), but did not develop them to a great degree. More recently, there have been several studies of listening based on the ecological perspective (e.g., Warren & Verbrugge, xxx; Vanderveer, xxx; Gaver, xxx); these will be discussed later in this chapter. However, though such studies have proven informative on their own and lent support to the idea that such an approach might be fruitful, a comprehensive account of everyday listening has yet to emerge. One of the purposes of this chapter, then, is to point the way to such a explanatory framework, both to help us understand everyday listening and in order to facilitate the creation of systems which analyze, synthesize, and manipulate sound in this way.

What might such a framework look like? It must answer two simple but fundamental questions. First, in expanding upon traditional accounts of elemental sensations, we must develop an account of ecologically relevant perceptual entities: the dimensions and features of sources that we actually obtain through listening. Thus our first question is *What do we hear?* Similarly, in expanding traditional accounts of the primitive physical features of sound, we must seek to find the acoustic properties of sounds that convey information about the things we hear. This involves the development of an *ecological acoustics*, one which describes the attributes of sounds that both provide information about sources and that are perceptually available. Thus our second question is *How do we hear it?*

In the rest of this chapter, we explore these two questions with the aim of developing an account of everyday listening that will complement and extend traditional psychophysics, and with will allow us to create and manipulate sounds along dimensions relevant for everyday listening.

What Do We Hear?

In trying to characterize what we hear, our concern is to develop a list of dimensions and features of everyday listening that are relatively simple and general. Just as we can capture a great deal of the sensory qualities of sounds with descriptions of their pitch, duration, loudness and so forth, we would like to find an equally simple set of descriptive dimensions and features that characterize everyday listening. And just as qualities such as pitch and loudness apply generally to most musical sounds, we would like our dimensions and features to apply to broad ranges of everyday sounds.

What sort of descriptions will do? Listen again to the sound examples X, and think about how you might describe what you hear. What sorts of dimensions might be useful in describing the sounds and the variations among them? Are there features that seem to apply generally and which have descriptive power?

The vast range of everyday sounds make simple descriptions of them difficult. Consider Figure X, for instance, which shows an exerpt from the table of contents of a sound-effects CD. This is a domain in which descriptions of everyday sounds has developed out of necessity. The first thing to notice is that the number of distinctive sounds listed is quite large -- over 50 in this example, over 100 on the disk, and this disk is only one of many. The world of everyday sounds is immense.

- Figure X goes about here: an extract from the table of contents of a sound-effects CD -

But notice that the sounds in this example are organized to some extent. Most are organized by the context in which they are likely to be heard: at the airport, in the kitchen, etc. This may be useful for finding a desired sound, but it seems a poor basis on which to build a description of what we hear -- the categories are not mutually exclusive; we can easily imagine hearing the same event in an airport and a kitchen. Nor do the category names constrain the kinds of sounds very much. We might expect to hear anything from running water to a small appliance in a kitchen, with the only unifying feature being their supposedly typical environment. Domain-based descriptions of sound-producing events seem unlikely to provide an adequate description of the attributes of everyday listening.

More interesting, however, are the suggestions of a hierarchical description of sounds in this table of contents. For instance, "automobiles" might be a superordinate category, with "sports car" and "Model T" as subordinates. The list also suggests some dimensions (e.g., big door, door closes slowly) and features (metal door, wooden door). A framework based on these sort of entities -- hierarchies, features, and dimensions -- seems a more promising approach than one based on context. Superordinate categories based on types of events (as opposed to contexts) provide useful clues about the sorts of sounds that might be subordinate, while features and dimensions are a useful way of describing the differences among members of a particular category. Such a framework is more likely to be generative, to delineate a space of possible sounds.

We might consider two methods for building a description of everyday sounds that is both hierarchical and based on dimensions and features. The first is to consider the dimensions and features of sound-producing events that cause reliable and audible differences in sounds. The second is simply to ask people what they hear, and to analyze their answers for commonly-used descriptors. We will discuss each of these approaches in turn, and end with an initial set of parameters that seem widely applicable to a number of everyday sounds.

The Physics of Sound-Producing Events

In this section, we describe the physics of sound producing events in a qualitative way. The purpose here is not to provide an exact account of mechanical physics, but instead to provide an initial orientation towards the relevant attributes of sound-producing events such as:

- Closing a door.
- Scraping fingernails over a blackboard.
- Water dropping into a pool.
- Wind whistling through wires.

- An exploding balloon.
- A resonating tuning fork.

It will become clear that these events share a number of common features – the most general being that all are caused by *interacting materials*. But there are substantial differences in the physics of closing a door, for instance, water dropping into a pool, and an exploding balloon. Most fundamentally, these events fall into three categories: those in which sounds are produced by vibrating objects; those in which sounds are produced by changes in the surface of a body of liquid; and those in which sounds are directly introduced into the atmosphere by aerodynamic causes. Here I first describe vibrating objects in some detail, and then use that discussion to illuminate liquid and aerodynamic sounds.

Vibrating Objects

Sound waves are formed when a pressure variation is introduced into the atmosphere. One common source of pressure waves are vibrating objects. This class of events include hitting books and scraping fingernails, as well as footsteps, closing doors, and breaking glass (several vibrating object sounds can be heard in Sound Examples N - N). An example of a vibrating object is shown in Figure X.

Sound examples N - N: Several examples of the sounds made by vibrating objects.

Figure X. *When an object is deformed by an external force, internal restoring forces cause a build up potential energy (A). When the external force is removed, the object's potential energy is transformed to kinetic energy, and it swings through its original position (B). The object continues to vibrate until the initial input of energy is lost to damping (C).*

Objects vibrate when a force is exerted upon and then removed from a system that is otherwise at equilibrium. This input of energy deforms the sounding object from its original configuration; the forces that resist this deformation result in the build up of potential energy in the new configuration. When the deforming force is removed, the object starts to return to its original shape, due to various *restoring forces* acting with the potential energy stored by the deformation. This results in the movement of the object towards its initial configuration. But when it reaches this initial position, the potential energy has been converted to kinetic energy, and the object moves through the resting configuration. Just as a pendulum swings back and forth once it is set in motion, the repeated translation from potential energy to kinetic energy and back again causes the object to move through its resting configuration repeatedly: It vibrates.

If no energy were lost in this translation, vibration would go on forever, and the world would be a very noisy and shaky place. But energy is lost for various reasons, collectively referred to as *damping*. So the object moves until all the initial energy responsible for its deformation has been lost, and it returns to its old (or finds a new) equilibrium configuration.

A number of parameters of the physical system affect how it vibrates. These attributes

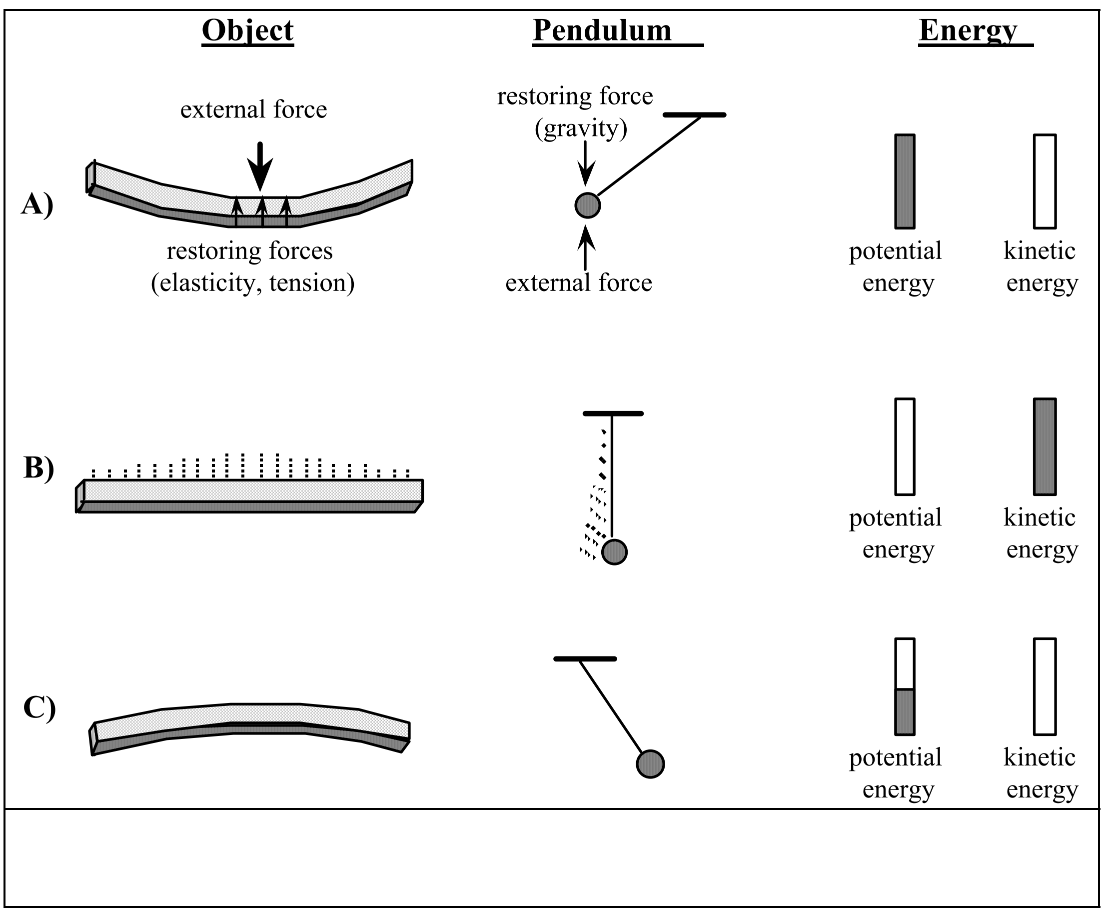

Table 1

Source Attributes	Effects on the Soundwave
Interaction	
Type	Amplitude function; spectrum
Force	Amplitude, bandwidth
Material	
Restoring Force	Frequency
Density	Frequency
Damping	Amplitude functions; also frequency
Homogeneity	Complex effects on amplitude; also frequency
Configuration	
Shape	Frequency, spectral pattern
Size	Frequency, bandwidth
Resonating Cavities	Spectral pattern

are grouped in terms of object properties in Table X. It is also useful to consider two domains of vibration separately in discussing these parameters, and separate those that affect the *frequency domain* from those that affect the *temporal domain*. Some source attributes make substantial effects on the sounding object's initial return from deformation, and thus influence the frequencies of its subsequent vibrations. Other attributes produce effects that become apparent only after repeated cycles of vibration, and thus their influences may be said to exist in the temporal domain. Of course, frequency is the reciprocal of time, so these domains are not physically different, but rather the result of different representations. But the two domains are separable both psychologically – when things repeat fast enough, they are perceived as pitch, otherwise as changes in other attributes – and in terms of the parameters of the event.

Four types of source attributes influence vibrations in the frequency domain. These are the *restoring forces* acting on the object, the object's *density*, the *size and shape* of the object, and the manner in which it is *supported*. For solid objects, restoring forces are either due to elasticity (hardness) or tension. The strength of these forces determines the potential energy resulting from some deformation, and the inertia of the system depends on its density; both together determine how quickly it will return from its deformed state and the frequency of its vibrations. The size and shape of the system constrains its modal vibrational patterns, as does the manner in which it is supported.

Three types of source attributes influence vibrations in the temporal domain. First is the type of *interaction* that causes it to vibrate, second the *damping* that causes it to stop vibrating, and third its *internal structure.* A basic division can be made between interactions that are *discrete* and those that are *continuous.* For instance, hitting is a discrete interaction, while scraping is continuous. The style of interaction usually has very obvious affects on the sounds produced. Damping has various causes, including internal heat transfer, plasticity, and external absorption of energy (including by the air as sound). Finally, the internal structure of the vibrating material makes many complex effects on the sound it produces, particularly in the temporal domain.

Note that, in general, attributes of the object (e.g., the strength of restoring forces, density, size, etc.) tend to influence the sounds in the frequency domain, while attributes of the interaction (e.g., its type and force) tend to influence the temporal domain. While this correspondence is by no means perfect – for instance, the force of interaction can affect a sound's bandwidth in the frequency domain, and the damping of a material is a strong determinant of the sound's temporal behavior – it is good enough to lend some support to Vanderveer’s (1979) hypothesis that *interactions affect the temporal domain* of sounds, and *objects the frequency domain*.

Vibrating objects as described above include many common sources of sounds, such as hitting or dropping books, scraping blackboards, clattering silverware, closing doors, and so on. In addition, this level of description can serve as a foundation for describing other more complicated events. For instance, crumpling paper makes sounds for similar reasons as a hit book. But when the paper is deformed, it doesn't return to its initial configuration but instead bends and creases along lines of stress. The sounds made are a result both of the sudden folds and the vibrating surfaces between them. Though new source attributes may come into play in such an event, those listed in Table X remain important. In general, it may be expected that these attributes are salient in determining the sounds produced by all events involving vibrating objects.

Aerodynamic Sounds

The properties of aerodynamic events are somewhat different than those describing vibrating objects (e.g., Sound Examples N - N). Where vibrating objects introduce pressure waves due to the interaction of a vibrating surface with the atmosphere, aerodynamic sounds are caused by the direct introduction and modification of atmospheric pressure differences from some source.

Sound examples N - N: Several examples of aerodynamic sounds.

The simplest aerodynamic sound is exemplified by an exploding balloon (see Figure X.A.) When a balloon bursts, a mass of high-pressure gas is released into the surrounding atmosphere. This sudden pressure variation propagates as a wave which may be heard if the pressure differences that reach the ear are large enough and if they change at an appropriate rate. In such events, sound is directly caused by sudden pressure variations in the air, not by the effects of a vibrating surface. Most of the information conveyed by explosions seem to be carried by the frequency bandwidth of the sound, and seem likely to concern the size or force of the explosion. High frequency components indicate the suddenness of the pressure change near the source; low frequency components the amount of gas involved (and thus the duration of the initial pressure release). So one can hear large, sudden explosions, or smaller, less abrupt bursting noises.

The sudden change in pressure caused by a bursting balloon or explosive is analogous to a discrete interaction (such as a hit) that causes an object to vibrate. Other aerodynamic sounds are caused by more continuous events, such as the hissing of a leaky pipe or the rush of wind from a fan. These events also make sounds due to the introduction of pressure variations in the atmosphere.

The attributes of the specific sources of pressure variations seem likely to produce the most salient affects on these sounds. That is, the sounds produced by leaky pipes are determined by the pressure within the pipe and variations in this pressure caused by turbulence. The sounds made by wind rushing from a fan are affected by the speed and size of the fan, and thus the volume of air it moves. Although the gases involved also affect the sounds, listeners are likely to be relatively insensitive to this information. In general, we expect aerodynamic sounds to be made by air (or, in more ecological terms, our auditory system is attuned to a world in which air is by far the most common gas). That this is so can be seen by considering why the sound of someone talking after inhaling helium is so humorous. The heightened pitch of the voice is perfectly predictable due to the lower density of the gas, but quite unexpected on the basis of experience.

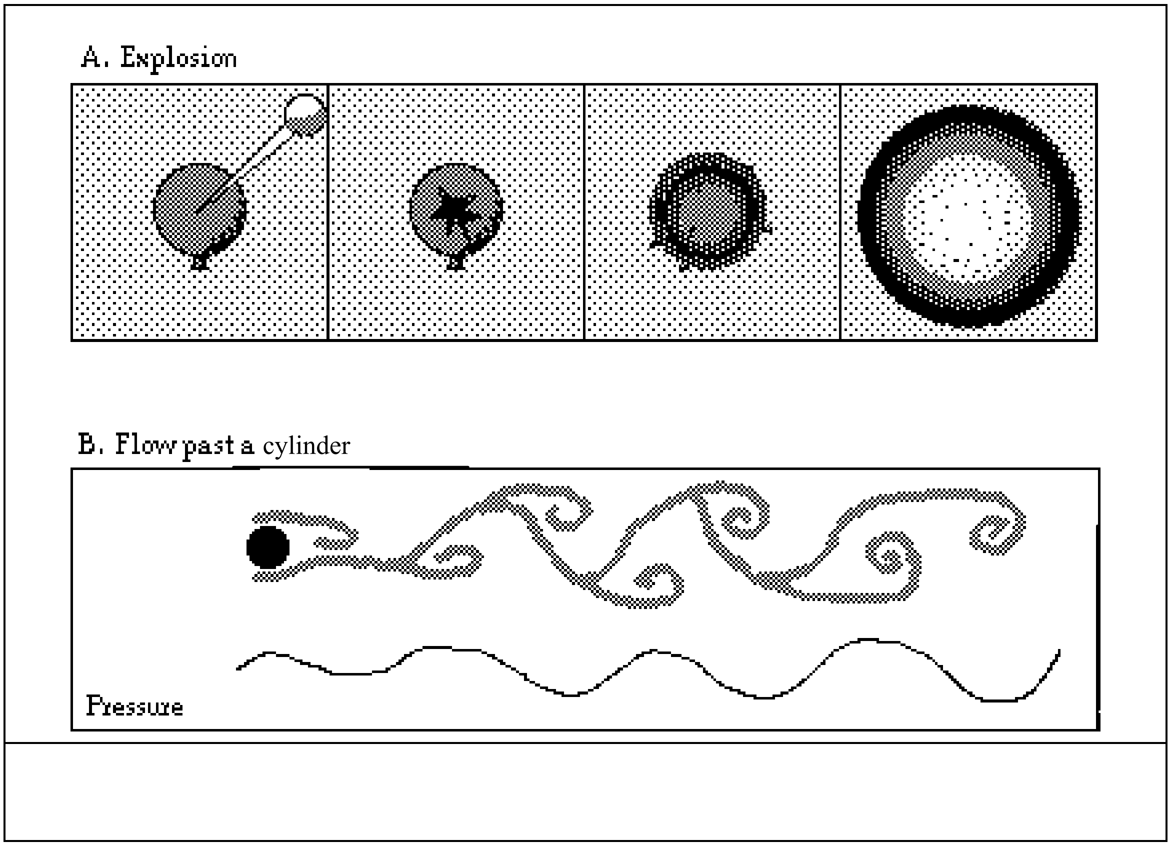

Figure X. Aerodynamic sounds involve abrupt changes of air pressure, as when a balloon explodes (A), or wind rushes past a cylinder (B).

Another sort of aerodynamic event involves situations in which changes in pressure themselves impart energy to objects, causing their vibrations. For example, when wind passes through a wire, eddies form on alternating sides, and the variations of pressure on each side causes the wire to vibrate (Figure X.B). The frequencies of vibration thus produced depend on the windspeed, size, and tension of the wire. This is the principle used in creating Aeolian harps. In addition, sound itself may impart energy to objects, as when its minute pressure variations match the modal frequency of a tuning fork, causing it to ring through sympathetic vibration. Such sounds are not purely aerodynamic, but perhaps better thought of as hybrids of aerodynamic and vibrating object sounds.

Liquid Sounds

Sound-producing events involving liquids (e.g., dripping and splashing) are like those of vibrating objects in that they depend on an initial deformation that is then resisted. But it seems that the resulting vibration of the liquid does not directly affect the air in audible ways. Instead, sounds are affected by the formation and change of resonant cavities in the surface of the liquid.

This can be seen most clearly in considering how an object dropping into liquid makes a sound, as shown in Figure X. As the object hits the liquid, it pushes it aside, forming a resonant cavity with a characteristic frequency. The cavity grows as as the object pushes more liquid aside, and thus the resonant frequency decreases. But the liquid's pressure causes it to close in on the cavity, and ultimately the object is immersed. The sound caused by such an event is 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.

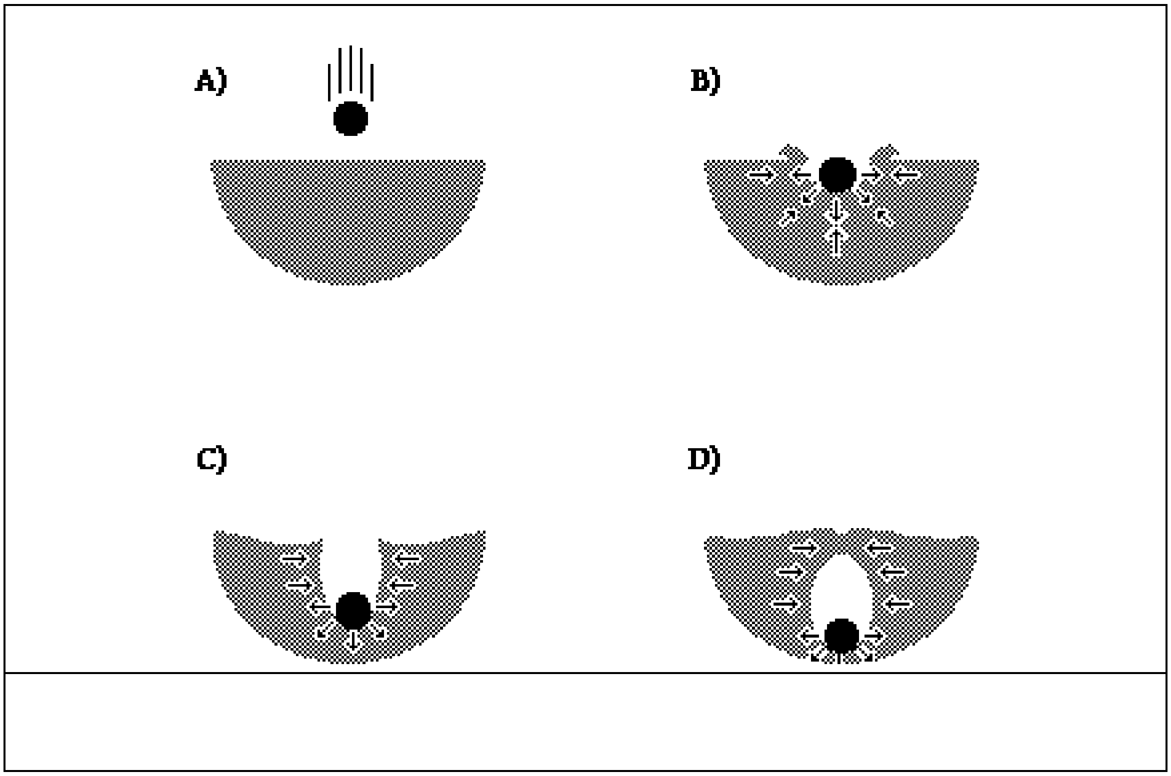

Figure X. When an object falls into a liquid (A), it forms a resonant cavity with a characteristic frequency (B), which changes as more liquid is pushed aside (C). Finally, the liquid's pressure causes the cavity to close in on the cavity (D) until the object is completely immersed.

More complex splashing sounds also seem to produce their sounds as changing cavities are formed which resonate, amplify and modify the sounds made by impacts of liquid on itself and other objects. Again, properties of interacting objects and the liquid itself are likely to affect the sounds. In most cases liquid sounds are probably heard as involving water, just as most aerodynamic sounds are probably heard as involving air. Still, the liquid's viscosity may produce effects on the sounds that are both salient and known by listeners. After all, it seems easy to tell whether a liquid gurgling out of a bottle is water, or a thicker syrup or oil. Such sounds seem to be related to one another as liquid sounds because of common, high-level characteristics of their evolution in time.

Though the physical attributes of aerodynamic and liquid sounds are not the same as those of vibrating objects, they do share common features. Aerodynamic sounds seem to be influenced largely by the interactions that create atmospheric pressure differences, so that explosions, hisses and fan noises all depend to a great degree on their causes. Liquid sounds also depend on properties of their causal interactions, such as the size and speed of an object falling into liquid. They are also influenced by attributes of the liquid, such as its pressure (a restoring force) and viscosity (analogous to density). Most generally, all such sources involve the *interaction of materials*.

Temporally Complex Events

Although all sound-producing events seem to involve vibrating objects, aerodynamic, or liquid interactions, many also depend on complex patternings of the simple events described above. So footsteps consist of temporal patternings of hitting sounds; and door slams require the squeak of scraping hinges, the whoosh of displaced air, and the hitting of the door on the frame. Though the discussion above may point to a useful framework for understanding the attributes of single sounds, it does not address those of more complex events.

Traditional physical accounts of sound-producing events do not address these sorts of complex events, but there are undoubtedly higher-level physical attributes of such events that make reliable effects on their sounds. Some of these involve timing of successive events, so that, for instance, successive footstep sounds probably must occur within a range of rates and regularities to be heard as walking. Others are likely to involve mutual constraints on the objects that participate in related events. For instance, concatenating the squeak of a heavy door slowly closing with the bang of a light door slammed shut would probably sound quite unnatural. These sorts of higher-level attributes of the events are not the sorts of variables that physicists typically study, but they are likely to be quite important for everyday listening. We will consider more complex sound-producing events at some length later in this chapter.

Asking People What They Hear

Considering the physical attributes of sound-producing events is useful in driving intuitions about the sorts of perceptual dimensions and features that might characterize everyday listening. But knowing how the physics of an event determines the sound it makes is not the same as knowing how a sound specifies an event. For instance, several attributes of a vibrating object, including its size, shape, and density, determines the frequencies of sound it produces. When hearing two sounds composed of different frequencies, then, how are we to know which parameters of the source has changed?

There are several possible answers to this problem. First, when several “basic” 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 which incorporates all of them. In other words, we may not hear size, shape, or density separately, but rather a new dimension which combines all of them. On the other hand, it may be that the effects of changing some of the basic variables are much smaller than changing the others -- for instance, changing the density of an object is likely to make a much smaller change in frequency than changing either its size or shape. In this case, the more effective parameters may also be more salient. That is, we may be inclined to hear a change in frequency as a change in size rather than density, just because size changes are the more significant source of frequency changes. Finally, it is quite likely that many parameters that change frequency also change other attributes of a sound. For example, changing the size of an object will change the frequencies of the sound it produces, but not their pattern. Changing the shape, on the other hand, changes both the frequencies and their relationships. These complex patterns of change may serve as information distinguishing the physical parameters responsible.

In any case, we can not base an account of everyday listening on the physics of sound-producing events alone. As Gibson (1979 68? xxx) pointed out, what is simple for physics may not be simple for perception, and vice versa. Instead, it is necessary to build an "ecological" physics, one founded on attributes relevant to listeners. For this reason, several studies have aimed at exploring the kinds of information sound conveys. Experiments of this sort complement analyses of physics: the data from such experiments constrains the sorts of physical attributes we might think we hear, while physical analyses can help in interpreting and organizing experimental data.

One approach to understanding the information people hear is basically experiential, involving introspection and self-observation. For instance, Jenkins (1985) reported that blindfolded students were able to orient themselves within their environment on the basis of auditory information such as "acoustic landmarks," resonances and echoes, and mixtures of near and far sounds. Note that they were not only using their ability to localize and to use reverberation as information about the environment (as discussed in Chapter 3 or so), but that the sounds themselves -- people talking, relatively continuous machine noises and the like -- served as meaningful and relatively stable landmarks (see Chapter X for an example of an application of this observation).

Another approach to understanding what people hear is simply to ask them. For instance, VanDerveer (1979) presented subjects with recorded tokens of 30 everyday sounds such as clapping and tearing paper in a free identification task. Subjects were run in groups, and asked to write a short phrase describing each sound. She found that subjects tended to identify the sounds in terms of the objects and events which caused them, describing their sensory qualities only when they could not identify the source events. In addition, subjects' mistakes tended to be based on the temporal qualities of sounds, so that clapping might be confused with dropping a book, but seldom with tearing paper.

Gaver (1988) ran a similar study in which he played 17 sounds to subjects and asked them to describe what they heard. In contrast to VanDerveer's study, subjects were run individually, and prompted by the experimenter to go into as much detail as they could about what they heard. Like VanDerveer, he found that subjects nearly always described the sounds in terms of their sources. Their accuracy was often impressive. For instance, several subjects could readily distinguish the sounds made by running upstairs from those made by running downstairs; others were substantially correct about the size of objects dropped into water; and most could tell from the sound of pouring liquid that a cup was being filled. Subjects did find that some sounds were extremely difficult to identify (e.g., the sound of a filedrawer being opened and closed), but they were almost always correct about some others (e.g., the sound of writing with chalk on a chalkboard). Often mistakes revealed interesting attributes that were heard. For instance, several people said the filedrawer sounded like a bowling alley, both of which share "rolling" as an important component. In addition, attempts to identify unusual or implausible sounds were equally interesting. For example, the sound of somebody walking across a floor covered with newspaper was described variously as a person walking on snow or gravel or as somebody rhythmically crumpling paper; only one subject guessed correctly, and immediately rejected the correct perception as being too implausible. Finally, their judgments followed the account of physics described above to an impressive degree. For instance, they never confused the sounds made by vibrating objects, liquid, or aerodynamic sources.

Ballas and his colleagues (Ballas, submitted; Ballas & Howard, 1987; Ballas & Sliwinski, 1986) have used free identification tasks to study everyday sounds that are ambiguous as to their sources. They have shown that a measure of the information inherent in a given sound, based on the number of possible sources that subjects propose, can be used to predict reaction times for its identification.

Studies such as these are informative, but sometimes frustrating. The result of asking people what they hear is often a list of events or attributes more akin to the list of sound-effects discussed earlier than to the set of dimensions and features we want. Nonetheless, there are inherent categories in many of these studies which reveal themselve both in correct answers and (perhaps even more often) in confusions. So for instance, the fact that the filedrawer sound in Gaver (1988) was confused with that of bowling suggests that rolling may be a particularly salient event warrenting further study.

Attributes of Everyday Listening

Understanding the physics of sound-producing events is useful in suggesting physical attributes that might be heard, while the sorts of studies described above help to constrain hypotheses about the attributes that actually are heard. Using our knowledge of physics and the results of these studies together, then, we may begin to build up a framework for understanding some of the basic source parameters conveyed by sound. Such a framework may be tentative and speculative, but it is useful both in providing a guide to future research and in suggesting the attributes of sounds available for manipulation in the sorts of applications we describe in the next chapter.

The first part of this framework, shown in Figure X, divides groups first by broad classes of materials and then by the interactions that can cause them to sound. Most generally, sounds indicate that something has happened, that an event has occurred, that there has been an interaction of materials. All sounds, then, convey this information.

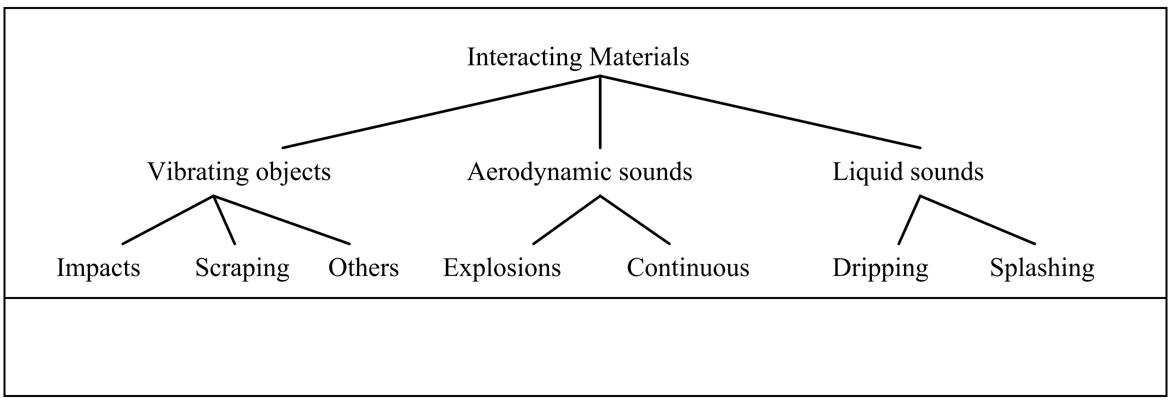

Figure X. *A hierarchical description of simple sonic events.*

At the next level, primitive sounds may be broken into three general categories: those made by vibrating objects, aerodynamics, and liquids. This categorization is supported both by the account of physics outlined at the beginning of this chapter, and by the results of the protocol study described in Gaver (1988). Although subjects often misinterpreted the sources of sounds they heard, no misidentifications crossed these categories: None of the subjects confused the sounds made by vibrating objects, for instance, with those made by water.

Finally, various distinct sorts of sound-producing events are shown at the third level of this hierarchy, defined by sound-producing interactions involving objects, aerodynamics and liquids.. The sounds made by vibrating objects may be caused by impacts, scraping, or other interactions (such as deformation and rolling). Aerodynamic sounds may be made by discrete, sudden changes of pressure (explosions) or more continuous introductions of pressure variations (e.g., fans, leaking pipes). Similarly, liquid sounds may involve discrete drips, or more continuous splashing.

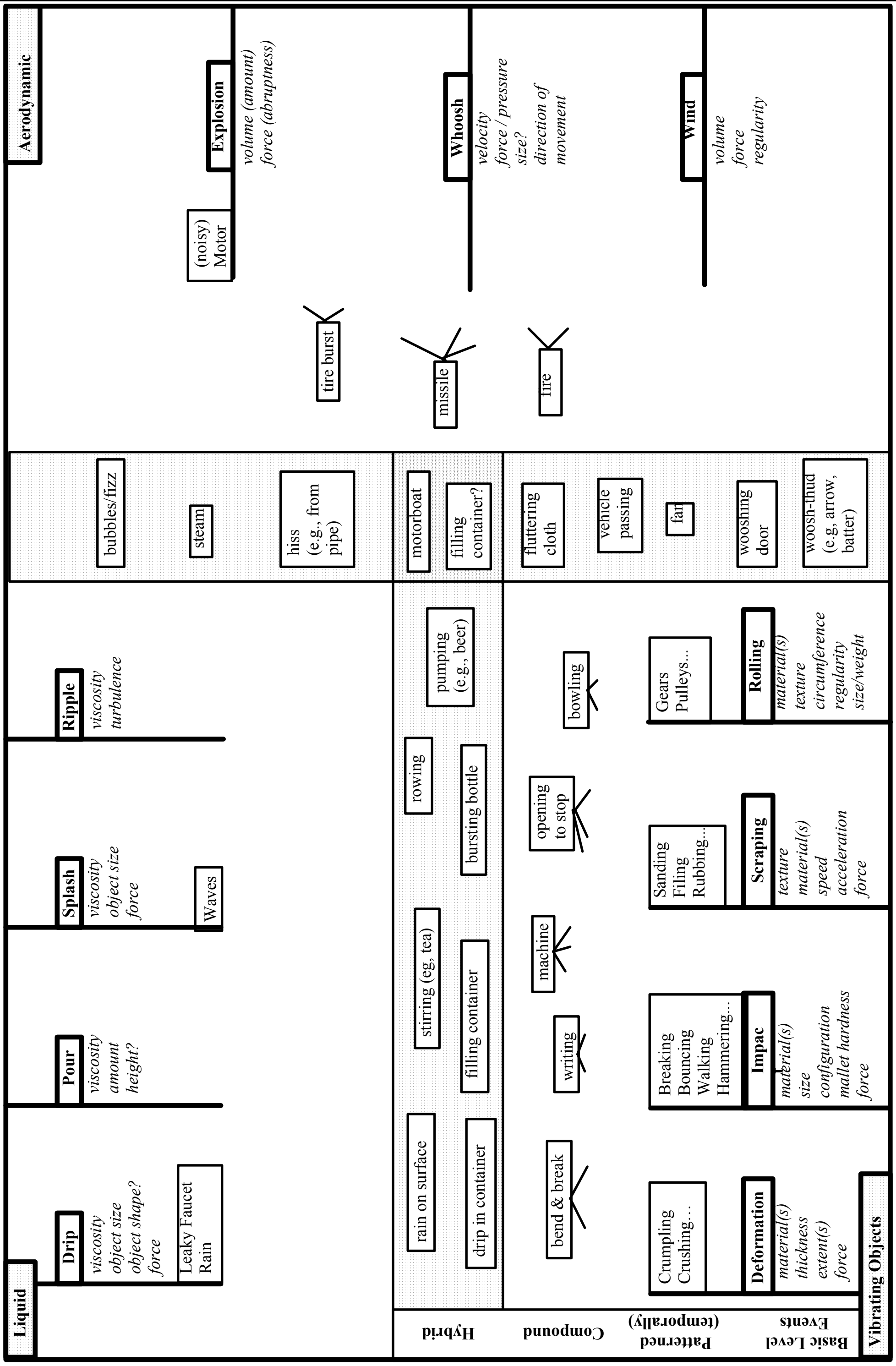

Figure X. *A framework for everyday sounds. Three fundamental sources (vibrating objects, liquids and aerodynamics) are shown in the three overlapping sections of the figure. Within each section, basic sound-producing events are shown in bold, and their relevant attributes next to them in italics. Complexity grows towards the centre of the figure, with temporally patterned, compound, and hybrid sounds shown.*

This simple classification is useful in building up a more comprehesive framework of everyday sounds, as shown in Figure X. This figure is broken into three main overlapping regions, corresponding to vibrating object, liquid, and aerodynamic sounds respectively. Within each region, several levels of sound-producing events are suggested. Finally, the overlapping regions show examples of sound events that involve hybrids of different sources.

Consider, for example, the region describing sounds made by vibrating objects. Four fundamentally differnt sources of vibration are indicated as *basic level events*: deformation, impacts, scraping and rolling. Under each of these basic level events are listed the attributes of these events that are relevant for the sounds they produce. For instance, impact sounds seem to convey information about five aspects of the event: the vibrating object's material, size and configuration, the surface hardness of the impacting materials, and the force of the impact. Scraping, on the other hand, conveys information about the texture of the surfaces, one or more of the materials involved, the speed and acceleration of scraping, and the force or weight with which one object is scraped over another.

Patterned, Compound, and Hybrid Complex Sounds

Above these basic level events are shown three sorts of complex events. The first is defined in terms of *temporal patterning* of basic events. For instance, breaking, spilling, walking and hammering are all complex events involving patterns of simpler impacts. Similarly, crumpling or crushing are examples of patterned deformation sounds. We would expect these sorts of events to convey the attributes made by their basic level constituents; in addition, other sorts of information are made available by their temporal compexity. For example, the regularity of a bouncing sound provides information about the symmetry of the bouncing object; variations in the scraping sounds produced by filing might indicate the general configuration of the object being filed.

The next level of complexity is produced by *compound* events which involve more than one sort of basic level event. For instance, the sounds made by writing involve a complex series of impacts and scrapes over time, while those made by bowling involve rolling followed by impact sounds. Again, these sounds are likely to convey information inherited from their basic level components as well as new information made available by their complexity. It is also worth noting that while some of these events involve more than one sort of source simultaneously (e.g, writing), others involve a series of basic events (e.g, opening a file drawer until in impacts against its stops).

Finally, *hybrid* events involve yet another level of complexity in which more than one basic sort of sound is involved. For instance, when water drips on a reverberent surface, the resulting sounds are caused both by the surface's vibrations and the quickly-changing reverberent cavities, and thus involve attributes both of liquid and vibrating object sounds. Some hybrid events involve attributes of all three basic sources: for instance, the sounds made by a speeding motorboat involves the splashing of the water, the vibrating engine sounds, as well as the rush of air past the body of the boat. As with other complex events, hybrid events provide information about their basic source categories and the basic events involved in their production, as well as more ideosyncratic information specific to their sources.

Problems and Potentials of the Framework

Clearly the framework shown in Figure X is far from complete. For one thing, we know much more about how to characterize the sounds produced by vibrating objects than we do about either liquid or aerodynamic sounds. In addition, we don't yet know how to characterize the attributes of many complex events (though see the discussion of breaking, bouncing and spilling below). Finally, the three basic categories of sounds we propose may not be enough. What of electronic sounds, such as those made by sparks or humming? Should fire be a basic sound-producing category (as suggested by our earth-air-water trichotomy)? What of vocal sounds? Though the inheritence strategy we propose here seems powerful and correct, there is much more to be discovered about how people hear events.

There are several more fundamental problems with this system. Two have to do with the sources of information used in its creation. Insofar as it relies on verbal evidence from subjects, it is liable to confuse the effects of language use for the attributes of perceptual experience. Insofar as it is based on an analysis of physics, it is liable to confuse source attributes that affect sounds with those that are actually heard. For example, the texture of an impacting object certainly affects the physics of dripping, but it is by no means obvious that this property is perceptible. Clearly, other sorts of research will be necessary to test whether these sources of information are informative.

In addition, the idea that the attributes of different sorts of events may be cataloged unequivocally is somewhat questionable. The information people obtained from the various sounds used in protocol studies often seem to be somewhat peculiar to the sounds themselves. Not all impact sounds, for instance, provide equal information about material, while some sounds convey more information about materials than interactions (for instance, many subjects heard metal but not deformation when listening to a crumpling can).

Nonetheless, this framework does seem to describe satisfactorily a great deal of the information inherent in everyday sounds. It captures our intuitions about basic sorts of sound-producing events and the information they make available for listening. It does so in a way that recognizes the mutual constraints of materials and interactions in producing sounds. In addition, it provides a mechanism by which to understand the myriad of complex events we typically hear. Finally, and perhaps most important, it seems useful in suggesting the kinds of source attributes that may be manipulated when creating or shaping everyday sounds that are to be used to convey information.

How Do We Hear It?

Asking what people hear is useful in understanding attributes of the experience of everyday listening. As we have seen, these attributes can be expressed in terms of the physical parameters of sound-producing events -- the kind of interaction that starts an object vibrating, for example, or its force; the viscosity of a liquid, or the size of something dropped into it, and so on. Such an account is useful in knowing the sorts of dimensions we can manipulate in applications which involve everyday sounds.

Asking how we hear it, on the other hand, is meant to get at the attributes of sounds than convey information about the events that caused them. It is one thing to suggest, for instance, that people hear the material of a vibrating object, and another to understand how a sound may be changed so that the perceived material changes. If the account prompted by asking what we hear expands traditional descriptions of elemental sensations, then the account of how we hear it is meant to expand traditional descriptions of the primitive physical dimensions of sounds.

Even less is known about how to characterize the perceptual information that allows us to hear events in the world than is known about how to characterize what we hear of them. This is not surprising, as it is the more difficult of the two tasks. The sort of framework introduced in the last section is not definitive, but suggestive. It organizes a set of semiformal hypotheses based on intuition, physics, and experimentation. Understanding how we hear these attributes requires that we test these hypotheses, that we formalize the information that leads to these supposed experiential dimensions. Asking what we hear is useful in orienting towards the immense range of everyday sounds, but asking how we hear is necessary if we are actually to create them.

Analysis and Synthesis of Sounds and Events

Computer musicians have addressed a problem similar to the one we are asking here. One of the goals for some (but not all) computer musicians is to find efficient means to capture the sounds of acoustic instruments. As we have described above, such sounds can be completely described in digital form by the output of a time-varying Fourier analysis, but this description is likely to be huge. A method known as *analysis and synthesis* (Risset, Deutsch book, xxx) has been developed which aims at understanding how to reduce the data from such analyses so that only that necessary for recreating a perceptually identical sound is retained.

Analysis and synthesis, as the name implies, involves analyzing a real instrument sound and then synthesizing a duplicate on the basis of the analysis. The analysis data can be systematically reduced, and synthesis driven by the results. For instance, straight line segments can be used to approximate complex time-varying attributes of the sound, and the resulting synthesized sound compared to the original. In this way, an understanding of which aspects of the sounds are crucial for perception may be obtained.

Understanding the effective perceptual information can be studied in similar ways. The sounds made by actual events can be analyzed, the data reduced, and then synthesized sounds can be compared to the originals. The purpose of this comparison, however, is not to produce an identical sound (one which produces the same sensations) but to produce a sound that retains information for relevant source properties. For instance, if we were interested in understanding how we hear the texture of a scraped surface, we might record and analyze a number of surfaces being scraped by various objects. In resynthesis, we would be concerned only with maintaining the information relevant for texture, not the size or material of the surface or the scraping object. In this way, we reduce the data from the analysis until only that which conveys the relevant information remains. The result is a description of the information relevant for scraping, independent from that relevant for other source attributes.

We can go a step further, and explicitly consider the source in our account. Now we don't only analyze the acoustics of the source, but the physics of the event. And similarly, we can reduce our description of the source until only those attributes relevant to a particular source attribute are described. Resynthesis, then, can be driven not only by a reduced description of natural sounds, but reduced descriptions of sound-producing events. Basing synthesis on analyses of events as well as sounds is valuable in helping to suggest what acoustic attributes will indeed be relevant and which will not. As will become evident in the examples below, many of the acoustic attributes which provide information for events are subtle and unlikely to be made evident by inspection of acoustic analyses alone. Using analysis and synthesis of events as well as sounds, then, is a powerful method for understanding the effective information for everyday listening.

Breaking and Bouncing Bottles

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

Consider the mechanics of a bottle bouncing on a surface (see Figure Xa). 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 physics). Energy is dissipated with each bounce so that, in general, the time between bounces and the force of each impact becomes less (some irregularities in the pattern are likely to occur due to the bottle's assymetry). 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 Xb). 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 (e.g., Sound Examples N - N).

Sound Examples N - N: Breaking and bouncing bottles.

This informal physical analysis is born out by acoustic analyses of natural tokens of breaking and bouncing sounds (see Figure X). Spectrograms of recorded bouncing sounds clearly show a series of impacts, each with identical frequency components, which repeat at a decaying 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. Individual 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 artifical tokens of breaking and bouncing sounds by combining natural tokens of single impacts in various temporal patterns (see Figure X). The sounds made by four individual pieces of a broken bottle were recorded separately. In order 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.

Subjects were asked to rate natural and artificial bouncing and breaking sounds in order to verify these analyses. When presented with natural tokens and asked to rate them as "bouncing," "breaking," or "don't know," subjects were 99.3% correct for bouncing, and 98.5% for breaking. Clearly subjects were able to obtain and use information for the events. When asked to rate artificial tokens, subjects were 93.0% correct for bouncing patterns, and 86.7 % correct for breaking. Despite some performance degradation, Warren and Verbrugge's (1984) characterization of the information for breaking and bouncing appears to have been substantially correct.

Two things should be noted about these results. First, the breaking sounds they constructed were quite simple. Natural breaking sounds are likely to have an initial impact and rupturing sound different from those following, but Warren and Verbrugge's constructed tokens did not. In addition, natural breaking sounds are probably characterized by many more than four overlapping bouncing sounds. Second, note that the rating task they used is a fairly coarse test. Asking subjects to rate breaking versus bouncing places constrains the events that subjects might think they heard. Subjects were offered a third category, "don't know," to try to reduce this constraint. Nonetheless, we might suspect that their judgements indicate that a particular sound is more representative of breaking than bouncing, for instance, without necessarily sounding like breaking. The coarse grain of Warren and Verbrugge's (1984) empirical methodology, then, seems likely to have balanced the simplicity of their sounds.

Nonetheless, this work is a good representative of analysis and synthesis studies. The combination of an intuitive physical analysis with acoustic analyses seems quite useful in discovering informative properties of everyday sounds. These analyses can be tested by constructing versions of the sounds. In this case, the result is a simple description of the features distinguishing two sound-producing events, a description that can be used in synthesizing representative sounds.

Impact Sounds

In our discussion of the physics of vibrating objects earlier in this chapter, we suggested several properties of objects that might be conveyed by impact sounds, including those of the vibrating object's material and configuration and those of the type and force of impact. A number of studies have been concerned with understanding the information for these properties conveyed by sounds and the accuracy with which people hear them.

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. Four different sized pans were used: one each of 1-, 2-, 3-, and 6-quart sauce pans. Six mallets with heads of different hardnesses were used: metal, wood, rubber, cloth-covered wood, felt, and felt-covered rubber. The sounds were analyzed using a model of the peripheral auditory system. The model first passed the signal through a bank of critical-band filters, squared the magnitude of the output signals, converted the results to decibels, and transformed them by an A-weighting function to approximate loudness (rather than amplitude). The resulting description of the sounds is similar to that provided by a Fourier analysis, but is held to be 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 TWA* (time weighted average) a measure of the ratio of high to low frequency energy in the sounds and its change respectively. Finally, multiple regression was used to assess the usefulness of these parameters in predicting hardness judgements made by nine skilled listeners. The parameters seemed to perform as accurate predictors, with an overall multiple R-squared of .725. However, individual parameters varied widely in their predictive power. Most useful were measures of the spectral centroid and the spectral centroid TWA. To a first approximation, then, mallet hardness is conveyed by the relative presence of high and low frequency energy.

Material and Length

Gaver (1988) studied people's abilities to judge the length and material of struck wooden and metal bars based on the sounds they made when struck. He recorded the sounds made by striking ten wooden and metal bars of five different lengths (10, 20, 30, 40, 50, and 60 cm.) three times each for a total of 30 sounds. A model of the physics of the events was developed which 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.

According to this model, the material of the bars made several effects on the sounds (as indicated by the Fourier analyses shown in Figure X). 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, while the sounds made by vibrating metal decay slowly, with high-frequency components lasting longer than low ones. In addition, metal sounds had partials with well-defined frequency peaks, while wooden sound partials were smeared over frequency space. The sounds made by wooden bars tended to have fewer high-frequency components than those made by metal bars, and each material seemed to support a band of frequencies better than those higher or lower; for wood this reverberant range was lower than for wood. Finally, the sounds made by a given length of metal tend to be higher than those made by a given length of wood because of metal's greater density. The many and 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, frequencies change monotonically with length, but the frequency of the partial with the highest amplitude changes nonlinearly with length (see Gaver, 1988). Thus, according to this model, information for material of the bars is more salient than that for length.

Gaver (1988) asked subjects to judge the material and rate the length of struck bars based both on recorded sounds and sounds synthesized on the basis of his model . Subjects were excellent at judging material -- 96% and 99% correct for natural wooden and metal sounds, and 91% and 97% for synthesized ones. They were much less accurate at judging length: Their judgments showed large interactions with material. However, with a brief training session in which they received feedback about their judgments their accuracy improved dramatically. In addition, almost all subjects' ratings correlated with frequency, so that low sounds were judged as indicating long bars, and high sounds short bars. In sum, Gaver's (1988) studies supported his analysis of the information for material and length, but suggested that judgments of length were disrupted by interactions between the effects on sounds of length and material.

Internal Friction and Material

Wildes and Richards (xxx) also studied the material identification based on impact sounds. Their approach differed from Gaver's (1988) in two ways. First, they rely entirely on analytical physics in their approach, and present no empirical data nor any acoustic analyses. Second, they focussed on identifying a property of the sound that is identified with material, where Gaver (1988) focussed on finding all effects of material on sounds.

Wildes and Richards (xxx) note that for many materials, the amount of deformation produced by an impact, and the material's return to equilibrium after the removal of the impacting force, lags compared to the impact. This characteristic is embodied in their model of a *standard anelastic linear solid*. The dynamic behaviour described by this model depends on an intrinsic parameter of material called *internal friction*. Internal friction determines both the sharpness of the peaks around a vibrating object's partials and the rate of its decay. Thus Wildes and Richards (xxx) conclude hearing material depends on assessing internal friction on the basis of peak bandwidth and decay rate. Note that this conclusion corresponds with Gaver's (1988) observation that wood and metal are differentiated both by their decay rates and by the the fact that partials of metal sounds are more sharply defined than those made by wood.