

AUDITORY INTERFACES

William W. Gaver
Royal College of Art
Kensington Gore
London SW7 2EU UK

INTRODUCTION

Our lives are filled with sound, though we may not be aware of it very often. As I write this, I hear a fan whirring to my right, pushing air through my living room. My stereo is playing quiet music to motivate me and drown out distractions. I hear a breeze rustle through a tree in the garden, shaking a random arpeggio from my windchimes. My cat meows at me, asking me to pet her, an airliner passes overhead, occasional traffic noises filter through from the street, and I hear my upstairs neighbour walking up the steps. None of this bothers me; on the contrary, it immerses me in my surroundings, providing the texture of my environment.

My computer, on the other hand, is conspicuously silent. I hear my fingers tapping on the keyboard, and it might beep at me to warn me of some problem, but that's about all. I don't hear anything as I select icons, move them around, copy them, or delete them. Occasionally, I might hear the hard disk spin up as it accesses data, but other than this my computer doesn't tell me anything about its internal state, its processes, or its modes, without using a visual display. I don't hear other people on the network, nor other events about which I might

want information. My computer is seen but not heard, and its reticence maintains a distance between me and the world it represents.

This is unfortunate, because there are a myriad of opportunities for using sound in computers. Many of the most obvious involve using speech as input or output. In this chapter, however, I focus on ways to use nonspeech audio in the interface—uses for sound that may be less obvious but as powerful as those involving speech (speech interfaces are the topic of another chapter in this volume).

Nonspeech audio can be used in computers as we use it in the everyday world, and new ways to use it are also emerging. I start with three examples of experimental auditory interfaces in order to ground the discussion and give an idea of the space of audio interface design. Then I discuss why sound is a valuable resource for interaction design, and speculate about why it hasn't been used more extensively. In the following section, I describe a variety of ways to think about and manipulate sound. Finally, I devote the majority of this chapter to a more thorough review of systems that have used nonspeech audio to convey information, ending with a discussion of new opportunities for auditory interface design.

Three Orienting Examples
Sound is a relatively unexploited medium in current computers, but

researchers have explored a variety of ways that sound might be used. Their examples define an implicit space for auditory interface design—one which is constantly expanding as new research is done. This space of design is defined by decisions made about three primary issues: First, what information an auditory interface is to provide; second, the kinds of sounds used; and third, the kind of mapping used to establish sound as a representational system. In this section, I briefly describe three examples of auditory interfaces as introductory landmarks for exploring this space.

Data Auralisation

One of the earliest applications of sound in computer interfaces addressed the goal of representing multivariate data. Graphical depiction is strained when faced with more than three data dimensions, but sound offers many dimensions to which data might be mapped. For instance, Bly (1982a, 1982b) demonstrated that sound could be used to discriminate between three different species of iris flowers. Each flower is characterised by four variables: sepal length and width, and petal length and width. These variables are generally sufficient to classify a given flower as belonging to one of three species, but it is difficult to read numerical data in making such a classification. Bly showed that

these classifications are easy to make if the data are mapped to variables of sound. By mapping sepal length to pitch, sepal width to volume, petal length to duration, and petal width to waveform, she represented each flower with its own simple sound. She presented experimental participants with examples of tones representing flowers from each of the species, and then asked them to classify new tones as belonging to one of the three groups. People were able to use the sounds to classify flowers as accurately as they could using most graphical methods (most of the listeners could accurately classify all but one or two of the flowers).

Bly's work is an example of data auralisation, in which parameters of sound are used to represent multidimensional data. It exploits sound's ability to form patterns that we can hear at a higher level than they are explicitly represented by the computer.

Alarms and Musical Messages

Another common use for sound is to alert people about some event. Alarm clocks exhort us to get up in the morning, ambulance sirens warn us to pull over, and foghorns alert sailors to dangerous rocks. The interrupt beep used by computers is an example of such sounds, but the information it provides is so generic that it can easily be customised (from bells to cartoon samples) without confusing its users. In

n Helander, M.G., Landauer, T.K. and Prabhu, P. (eds.), Handbook of Human-Computer Interaction, 2nd edition. Amsterdam, The Netherlands: Elsevier Science.

contrast, many other situations require that alerts sounds are not only noticed, but that they are identified and discriminated from other alarms. For instance, as many as 60 alarms may have sounded during the Three Mile Island accident (Sanders & McCormick, 1987).

Roy Patterson and his colleagues (e.g., Patterson et al., 1986; Patterson, 1982) have designed a system for generating multiple alarms that work together. First, the ambient noise level in the target environment is analysed, and used to design basic sounds that are neither too quiet to be heard, nor so loud as to be disruptive. These sounds are combined to form short tunes, or motives, that serve as the alarms. The motives can be seen as very simple pieces of music; incomparable in complexity or aesthetics to conventional music, but exploiting many of music's basic attributes. Rhythm and pitch are used to give each alarm a distinct identity, to convey the appropriate sense of urgency, and to mimic the alarms meaning (e.g., a motive with an increasing tempo is used to indicate that the speed of an aircraft is too high). The alarms themselves are repeated every few seconds while the situation applies, providing gaps during which operators may discuss the situation and take remedial action; but if the situation persists the alarm is played again, more urgently, to remind them to address the problem. This basic strategy has been tested with target users, and employed in airlines, intensive care wards, and railways.

Patterson's alarms are an example of using simple musical messages in interaction design, one that exploits sound's potential to create distinctive figures that we can associate with events.

The ARKola Simulation

A third use for sound in the interface is to convey information from computer systems about objects, events, and other people. For instance, Gaver et al. (1991) used a collection of everyday sounds to support people collaborating on a process control task. Pairs of people ran a simulated softdrink bottling plant, consisting of eight interconnected processes, that was too large to fit on the computer screen. Sounds were designed to indicate when machines were running, how fast they ran, and when supplies were being wasted—a total of about a dozen sounds playing at once. Users operated the plant to produce as many bottles of softdrink as possible, refilling supplies as they were used and dealing with programmed breakdowns of the machines. Comparing people's performance with and without sound, it became clear that sounds were useful in helping people monitor individual processes, allowed the higher-level activity of the plant as a whole to be perceived, encouraged collaboration without necessitating a shared visual focus, and motivated users by increasing the tangibility of the simulation.

The ARKola simulation is an example of using sound to provide feedback about user actions, to indicate ongoing processes, to support collaboration and to support peripheral awareness of events. In particular, the use of auditory icons—environmental sounds designed to be appropriate for the virtual environment of the interface—allows people to listen to events in the computer as they do to those in the everyday world.

These three examples indicate some of the possible strategies for using sound in interaction design, and some of the applications for auditory interfaces. Each depends on using a set of sound attributes and dimensions to stand for categorical and continuous information, through a mapping that can vary in the degree to which it is arbitrary or constrained.

Note that none of these examples come from the wide range of games and multimedia products that use sound. This is because, in general, the focus of sound design in such applications is on its emotional and aesthetic impact, rather than its ability to systematically convey useful information. In addition, games and multimedia designers have rarely discussed their designs and techniques in a research context. This is unfortunate. As I will suggest in concluding this chapter, there is much we can learn from the sound effects used in games and multimedia. In the body of this discussion, however, I will focus on relatively academic research on auditory interfaces.

Why use Sound? Sound as a Medium

The success of auditory interfaces depends on an understanding of the structure of sound that can be used in design as well as real problems that might benefit from sound. More deeply, they depend on an appreciation of the qualities of sound that make it a valuable medium for some tasks. Implicitly, auditory interfaces all offer answers to the question: Why use sound?

Vision and hearing are our two primary distance senses, allowing us to gather information about the world without physical contact. For most people, vision seems far more

important than sound, to the point that what something looks like is almost inextricably tangled with what it is. Sounds seem secondary. Noises might tell us something about an object, but they seldom really define it. Seeing is believing; hearing only tells us where to look. Vision seems richer, more detailed, and more exact than sound, and thus better used in designing, creating, and communicating new interface worlds.

What this perspective ignores, however, is that sound is a different sort of medium from vision, one that provides information that vision cannot. What we see about an object is not what we hear about it. What we see is patterned variations of light frequencies, usually reflected from the surfaces of objects in the environment around us. These patterns and their variations can tell us about the size of objects, their corners and curves, their textures and materials. But what we hear is patterns of moving air, often emitted from objects as they vibrate due to some event. Sound usually conveys information about the substances and dynamics of objects, rather than their surfaces. Sound tells us about the size of objects, their internal makeup, their parts and their hollows, their textures and consistency. Sound tells us about events, about the contacts and scrapes, bounces and breaks as things move and interact around us. And sound tells us about our environment, interacting with the spaces and surfaces around us before reaching our ears amplified, echoed, and filtered. Sound has different affordances as a medium than vision does; it offers different sorts of information in the everyday world, and therefore different

opportunities for providing information from designed systems.

The affordances offered by the media differ also because the sensory systems with which we gather information from light and sound have different characteristics. Our eyes can register detailed differences in light patterns only over a few visual degrees, and even peripheral vision only extends over about a third of the potentially spherical field of view that surrounds us. Thus we move our eyes, our heads, and our bodies to see what is around us. Visual objects tend to persist in a given spatial location, and we can turn to scrutinise them, or turn away to obliterate them from sight. Our ears, in contrast, register vibrations from all around us (and even from within our own bodies). We need not turn to hear something; in fact we cannot turn away or close our ears.

The different spatial characteristics of our visual and auditory systems interact with temporal characteristics of the media. Sound-producing events tend to have finite duration, and to change in complex ways over time, while visual objects tend to be more stable. So, just as spatial arrangements mean we cannot see everything at once, so do temporal dynamics mean we cannot hear everything at the same time. Visual objects exist in space but over time, while sound exists in time, but over space (Gaver, 1989).

Finally, the ways we structure the world visually and sonically in building cultural artifacts are also different. Music is a language—using the term broadly—developed from sound that has no clear counterpart in vision. On the one hand, music relies on proportional relations of pitch and

rhythm that make it analogous to visual graphs and charts. On the other hand, music can convey mood, emotion, and narrative tension and resolution while remaining abstract from particular content, in ways similar to, but perhaps more developed than, the abstract visual arts. Music certainly seems more universal than either charts or abstract art, and tonal musical structure in particular is recognised very widely, even by cultures in which it is not indigenous. The ability to harness musical power is an important opportunity for interface design.

In sum, sound is a powerful medium for conveying information. Sound complements vision, providing information about distant objects and events, their internal configurations, timing, and dynamics. Sound can reveal patterns in data, give feedback about user actions or allow monitoring of system processes. Sound can provide peripheral awareness about other people and their activities, support an immersive presence in artificial or remote environments, and can be used to create mood, drama, and narrative flow. Because we need not orient our bodies in any particular direction to hear these things, we can look at one thing while hearing another. Moreover, when combined with speech, sound can be used to make computers accessible without any need for vision or visual displays at all.

Why Sound Isn't Used

Given the potential of sound as a resource for interface design, it is worth exploring why it has not been used more extensively in interface design. The first reason is that it has not been clear what sound might

offer; in fact, the term “bells and whistles” is sometimes used colloquially to refer to unnecessary extravagance in design. But sound can go beyond being a mere fillip to being an integral and essential part of the interface. This may be clearest where text and graphics have no role, as in interfaces for the visually impaired, or phone-based systems. In graphical systems, too, sounds may present information more clearly or more effectively than graphics. This will be shown over and over again by the research I discuss in this chapter.

Another pervasive objection to sound, of course, is that it can be annoying. Sounds permeate environments, and many people are disturbed by the idea of increasing noise levels. There are several answers to this objection. First is to note that, by definition, “noise” is unwanted sound. To a great degree, if sounds are useful and meaningful, they will not be annoying. Second, while there is some truth in saying that sound is inherently distracting—after all, we have no “earlids”—we are nonetheless surrounded by sound at all times, and in many cases can simply relegate it to the background of our attention. Thus auditory interfaces that seem irritating when we concentrate on them can fade to the background as we use them, if they are well designed, and provide their information relatively unobtrusively. Finally, sounds' annoyance can be controlled. Research by Patterson et al., (1986), Edworthy et al. (1991), and Swift et al. (1989) has considered how attributes of simple sounds contribute to their perceived urgency, which may be seen as a correlate of their potential to distract or annoy us. They have found that

easily controllable factors such as attack time, spectral type, and rhythmic regularity all contribute to urgency. Matching urgency to the information to be conveyed can help auditory interfaces become an unobtrusive, helpful part of the auditory ambience. There is no reason that sounds need to be annoying to be effective.

A final reason that sound has been underdeveloped in current interfaces is that hardware and software resources for using sound have been limited. It has only been relatively recently that commercially available computers have made it easy to control sound production beyond the ability to trigger a simple beep. The history of auditory interface design is the story of applications that push the boundaries of currently available systems, to make them useful for interface work. And, indeed, there has been phenomenal development of support for sound in current computers, and with it increasing sophistication in auditory interfaces. However, while new sound-making possibilities do encourage the wider use of auditory interfaces, it is important to recognise that the constraints they place may shape research in undesirable ways. The largest impetus for the development of sound-processing software and hardware comes from the music and entertainment industries. Auditory interfaces often require control over different attributes of sound, so there are still serious limitations in the resources available for design. In the next section, I describe a number of frameworks for understanding sound, to give an idea of the resources that could potentially be useful.

PARAMETERS OF SOUND AND HEARING

Designing with sound involves, at the most fundamental level, manipulating sound along various dimensions to produce different effects. This can involve mapping dimensions of sound to dimensions of data to produce the sonic equivalent of graphs (data auralisation), the creation of a discriminable set of musical motives to be used as labels for events (musical messages), or mapping perceptible attributes of source events to those we want to convey (auditory icons). In any case, it is necessary to identify relevant dimensions of sound that can be used in design. Building this common vocabulary is the purpose of this section.

Sound is a rich and many-layered medium, and different frameworks may be used to describe its structure. Here I discuss six. I start with acoustics and psychoacoustics, the basic sciences of sound and hearing that serve as a foundation for any strategy of using sound. Then I discuss everyday listening, which seeks to describe sound and hearing in terms of the perception of events in the world. I focus next on our ability to localise sounds, a topic which overlaps psychoacoustics and everyday listening, and explain systems which allow sounds to be located perceptually in a three-dimensional, virtual auditory space. After that, I turn to musical structures, and propose that the higher levels of organisation they offer might be applied to auditory interfaces more than they have been so far. Related to this are genre sounds, which may be musical or everyday sounds, but are linked to their meanings by long

cultural history. Finally I briefly discuss the effects that sound processing technologies have on how we think about and use sound, suggesting particularly that they offer a large number of separate, overlapping ways to think about timbre, the quality of sounds.

These are broad topics, and I can only summarise some of the relevant issues and research here. For more complete coverage interested readers should consult the references cited in each section as well as more general texts such as those by Bregman, (1990), Handel (1989), or Pierce (1983).

Acoustics and Psychoacoustics

Many of the dimensions used in designing auditory interfaces—e.g., physical dimensions such as frequency, amplitude, or spectrum, or perceptual ones such as pitch, loudness, or timbre—are described by acoustics and psychoacoustics. Acoustics describes the structure of sound itself, while psychoacoustics concerns the mapping between dimensions of sound and dimensions of our perception. This mapping is not linear, and thus it is important to distinguish the two. A design which naively assumes a straightforward correspondence between physical manipulations of sound and perceived differences may be ineffective or misleading.

Acoustics: frequency, amplitude, waveforms and spectra

Sounds are pressure variations that propagate through an elastic medium (usually the air, but also any other substance, such as walls, water, or bone). A simple description of sound can be created by graphing its waveform, which shows pressure variations over time

(Figure 1). The wave shown here is periodic, repeating itself with a frequency described in units of cycles per second, or Hertz (Hz). The amplitude of the wave refers to the degree of departure from the mean pressure level, and relates to the intensity, or force per unit area, of a sound. Since people are sensitive to a vast range of intensity—the most intense sound tolerable is about 10^{12} times as intense as the weakest that can be detected—amplitude is commonly expressed using the logarithmic decibel (dB) scale.

The waveform shown in Figure 1 is a sine wave. Though rare in nature, sine waves are often considered the simplest sort of sound because virtually any complex wave can be described as the sum of a number of sine waves with different frequencies, amplitudes,

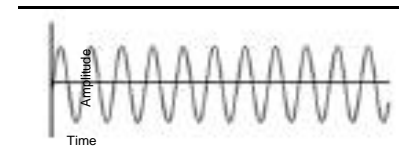


Figure 1. A simple sine waveform.

and phases (phase refers to the how the waveform is shifted in time). Figure 2A shows a complex wave, and Figure 2B the simpler sine waves from which it is made. Fourier analysis is an important analysis technique for analysing waves in terms of their component (sine wave) frequencies, called partials. When a wave is Fourier analysed, the results may be shown in a spectral plot that shows a wave's partials as energy plotted by amplitude and frequency (Figure 2C). Conversely, additive synthesis is

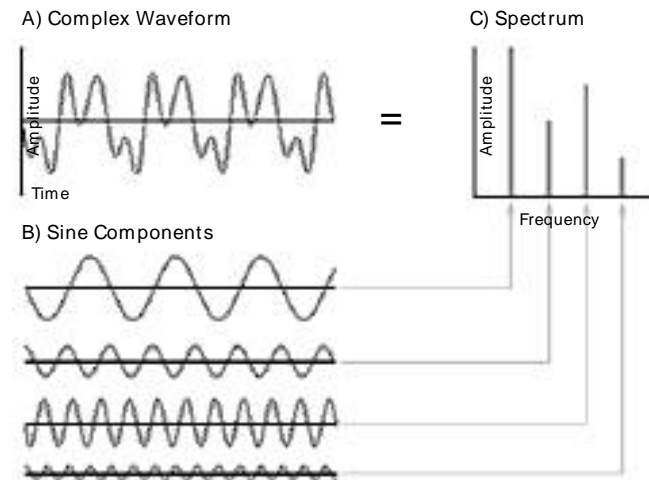


Figure 2. A complex waveform created by adding four sine waves. In A, the waveform is shown; B shows the four component sine waves, and C shows the equivalent spectrum.

a technique for synthesising sounds by adding together sine waves of different frequencies and amplitudes.

Complex waves may be categorised according to the frequency patterns of their partials. Musical sounds tend to be harmonic: the frequencies of their partials (also called harmonics for such sounds) are integer multiples of the lowest, fundamental frequency (e.g., a wave with partials at 100, 200, 300... Hz. is harmonic, with a fundamental frequency of 100Hz.). Harmonic waveforms are repetitive at the frequency of their fundamental and are heard with a distinct pitch.

Many natural sounds are inharmonic: the frequencies of their partials are not integral multiples of their fundamental frequency. Such sounds tend not to be repetitive, and if they are heard with a clear pitch—often they sound like a combination of different pitches—it depends on their average frequency, weighted by amplitude, rather than on their fundamental. Finally, noises are made up of energy at many different frequencies, with continuous spectra, rather than the peaked spectrum shown in Figure 2C. White noise has a spectrum that is flat, while other noises may be bandlimited; that is, they may contain energy within a range of frequencies characterised by their bandwidth. Such sounds, like inharmonic sounds, may be heard with a rough pitch corresponding to their average, amplitude-weighted frequency.

The categories of harmonic, inharmonic, and noise sounds provide useful landmarks in considering the structure of sounds, and in considering the kinds of sound that might be used for design.

Figure 3: A spectrogram showing amplitude (darkness) by frequency (vertical axis) by time (horizontal).

In practise, however, the boundaries are not entirely distinct. For instance, most acoustic musical instruments produce sounds that are only approximately harmonic, and their slight inharmonicities produce a richness that is important for their perception (Pierce, 1983). Similarly, the sounds made by some instruments, such as violins, tend to be so complex that hundreds of sine waves would be necessary to recreate their waveforms. A more efficient approach to synthesising such sounds is to use filters to shape a continuous noise spectrum, a process known as subtractive synthesis.

Sounds tend to evolve over time, as their partials rise and fall in amplitude, and vary in frequency. Thus a series of spectral plots, showing a sound's frequency components at a given moment of its evolution, tend to be joined to form a spectrogram showing the amplitude by frequency over time. Often spectrograms are shown with amplitude plotted as the darkness of a point plotted against time and frequency (Figure 3); alternatively, three dimensional graphs can be used to show all three dimensions.

Waveforms (Figures 1, 2A and B) and spectrograms (Figures 2C and 3) show sounds using different coordinate systems, and tend to be useful for different tasks. Waveforms can show the temporal evolution of sounds' amplitudes more clearly than spectrograms, and thus are often used in sound editing programs (the fact that they are much less computationally expensive also plays a role). Spectrograms can show the detailed frequency makeup of a sound much better than waveforms, however, and tend to be the more powerful of the two. This is due, in part, to the fact that phase relations among sounds' partials are not usually represented in spectral plots, but are very salient in waveform plots. Since phase relations tend not to be heard (Cabot et al., 1976), waveforms that look very different may sound identical. Thus systems that depend on manipulations of the waveform can be misleading.

Psychoacoustics: Pitch, loudness, timbre, masking, and streaming
The acoustic dimensions discussed have correlates in our perception of sound. For instance, the frequency of a sound corresponds roughly to its pitch, that is, whether it sounds high or low. Its amplitude correlates to loudness. Its time-varying spectrum correlates to its timbre, or tone-colour. But the mappings between acoustic and psychoacoustic dimensions tend to be nonlinear, so that equal changes along an acoustical dimension do not necessarily produce equal changes of perception. These mappings are also not orthogonal, but often interact. Changing the amplitude of a tone, for instance, can change its perceived pitch as well as its loudness. The combination of

nonlinearity and lack of orthogonality makes it very difficult to design sounds so that their perceptions may be exactly predicted.

Pitch does not relate linearly to frequency, but logarithmically: In general, doubling frequency raises the pitch by an octave, so that doubling a 220Hz tone to 440Hz raises the pitch by one octave, but to raise it by two octaves one must increase it four-fold to 880Hz. In addition, pitch is affected by loudness: As sine waves get louder, their pitch goes down if they are under 1000Hz, and up if they are over. Moreover, pitch is not a single perceptual dimension (Shepard, 1964). Pitch height, how high or low a note sounds, interacts with pitch chroma, which note is being played. For instance, two notes an octave apart may sound more similar than two that are only a semitone (one twelfth of an octave) apart, because they share the same chroma even if they are more dissimilar in pitch height.

Loudness does not relate linearly to amplitude, either. Roughly speaking, loudness (L) relates to amplitude (I) according to a power law: $L = kI^{0.3}$. This means that a 10dB increase of amplitude only doubles loudness. But loudness depends on frequency, as well, with tones between 1000 and 5000Hz perceived as louder than those that are higher or lower (more exact functions of loudness with frequency can be found in the form of equal-loudness curves; Fletcher & Munson, 1933). Loudness is also affected by bandwidth, the spread between the highest and lowest frequency components of a sound, with high bandwidth sounds seeming louder than equal-energy low-bandwidth

ones. Loudness also depends on duration, with sounds shorter than about a second increasing in loudness with duration (Scharf & Houtsma, 1986), while sounds longer than a second remain constant in loudness.

Timbre is a single term used to describe the most meaningful variations in sound, including those that make raindrops sound different from bells, roosters from train whistles, and flutes from ocean waves—assuming they have the same pitch and loudness. In fact, it is defined as "...that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" (American National Standards Institute, 1973). This negative definition means that timbre is "the psychoacoustician's multidimensional wastebasket category" (McAdams & Bregman, 1979).

Despite timbre's fundamental role in sound perception, remarkably little is known about its structure or dimensions. As Helmholtz (1885/1954) showed, timbre is influenced by the spectrum of a sound, so that changing a sound's partials will generally change its timbre. But it is also critically dependent on spectral dynamics, to the point that the perception of brass tones, for instance, depends more on low partials building up faster than high ones than it does on the exact spectral makeup of the sounds (Risset & Wessel, 1982).

Though research has tried to uncover meaningful perceptual dimensions of timbre and their acoustic correlates (see, e.g., Grey, 1977; Wessel, 1979), no encompassing system has yet been established. Recently, many

computer musicians have been turning away from the goal of determining abstract timbral dimensions. Instead, there is an increasing interest in using physical models of acoustic instruments to understand and control sound (e.g., Borin et al., 1993; McIntyre et al., 1983). These are related to attempts to recast timbre in terms of auditory event perception (as described in the section on Everyday Listening later in this chapter; see also the discussion of Device Models later in this section).

Apart from understanding the perceptual correlates of acoustic dimensions such as frequency and loudness, psychoacoustics also studies other effects of the auditory system on how we hear sound. Here I only touch on two issues that are salient for auditory interface design. Both concern the perception of multiple sounds.

The first, masking, concerns the ability for one sound to make another inaudible (as when you can't hear what somebody is saying in a noisy room). Sounds are typically masked by louder sounds, they tend to be masked by lower sounds rather than higher ones, they can be masked by sounds that come before or after them, and they tend to be less susceptible to masking the more complex they are (Pierce, 1983; Lindsay & Norman, 1977).

The second phenomenon, streaming, concerns the tendency for sounds to be grouped perceptually into perceived sources (Bregman, 1990). Streaming occurs both in grouping sequential sounds and in "fusing" partials in the perception of timbre. Generally, streaming of sequential sounds tends to occur when they are close in frequency, when the tempo is high, and when they share a common

timbre. Fusing of partials tends to be encouraged when they share a "common fate," undergoing similar changes in frequency or amplitude; thus vibrato is often applied to fuse inharmonic sounds (Pierce, 1983). Clearly, controlling both masking and streaming has great practical importance for auditory interface design.

Acoustics and psychoacoustics are the most fundamental sciences of sound and hearing, and a basic understanding of their terms and issues is crucial for all auditory interface design. My description here is far longer than the following descriptions of other frameworks for understanding sound, but it is still woefully short to do the subject justice. Nonetheless, I hope to have introduced the topic well enough to allow readers new to the field to understand the following discussion.

Everyday Listening

A second approach to sound and hearing stresses everyday listening. Where psychoacoustics is concerned with hearing sounds per se, everyday listening is concerned with the experience of hearing sounds in terms of their sources (e.g., Gaver 1993a, 1993b, 1988). For instance, if one drops a crystal vase while carrying it in a darkened room, one is less likely to attend to the pattern of pitched impulses it produces than one is to try to ascertain what it landed on and whether it broke. The experience of listening to the attributes of the sound—its pitch, loudness, duration, or timbre—is an example of musical listening. The experience of listening to determine the source itself—whether it involves a hard or soft surface, a bouncing or breaking object, a threatening or harmless event—is an example of everyday listening.

The approach to audition captured by everyday listening is suggested by Gibson's (1979) ecological perspective on perception. The ecological approach emphasises that perception should be understood in terms of the fit between the organism and a structured environment. For audition, one implication is a new framework for describing sound and hearing, one which complements more traditional approaches (see Gaver, 1993). The vast variety of sounds we hear in the world may be characterised in terms of their sources, their attributes in terms of source attributes. Instead of talking about pulses and buzzes, pitches and timbres, we can talk about impacts and scrapes, size and material. Instead of relating sensations to simple acoustic attributes, as psychoacoustics does, we can relate source perception to more complex acoustic patterns. Most importantly (at least for the purposes of this chapter), we can build auditory interfaces using this framework. Instead of mapping information to sounds, we can map information to events.

Two questions are important in orienting to the new framework suggested by everyday listening. The first is what do we hear? If traditional terms for describing sound and hearing are to be replaced by those referring to source events, what objects, attributes, and dimensions are appropriate? A possible framework is shown in Figure 4., based on a mixture of protocol studies and physical analyses (Gaver, 1993b, 1988). All sound producing events involve an interaction of objects. Basic level events—the simplest combinations of objects and interactions—are grouped according to three basic

material categories of solids, liquids, and gasses, and shown on the outside of the figure with the attributes that might be conveyed by sound. More complex events, which can be classed as temporal patterns,

compound events, and hybrid events (involving materials from more than one category) are shown towards the centre of the figure. Complex events inherit the attributes of their basic components, so that any

sound involving a struck solid, for instance, conveys information about size, material and so on. In addition, complex events convey higher-level information, such as the speed of a machine or space left in a container

being filled with liquid. This is a tentative and simple framework, but it is useful in capturing ideas about everyday listening, guiding future research, and serving as a palette for the design of auditory icons.

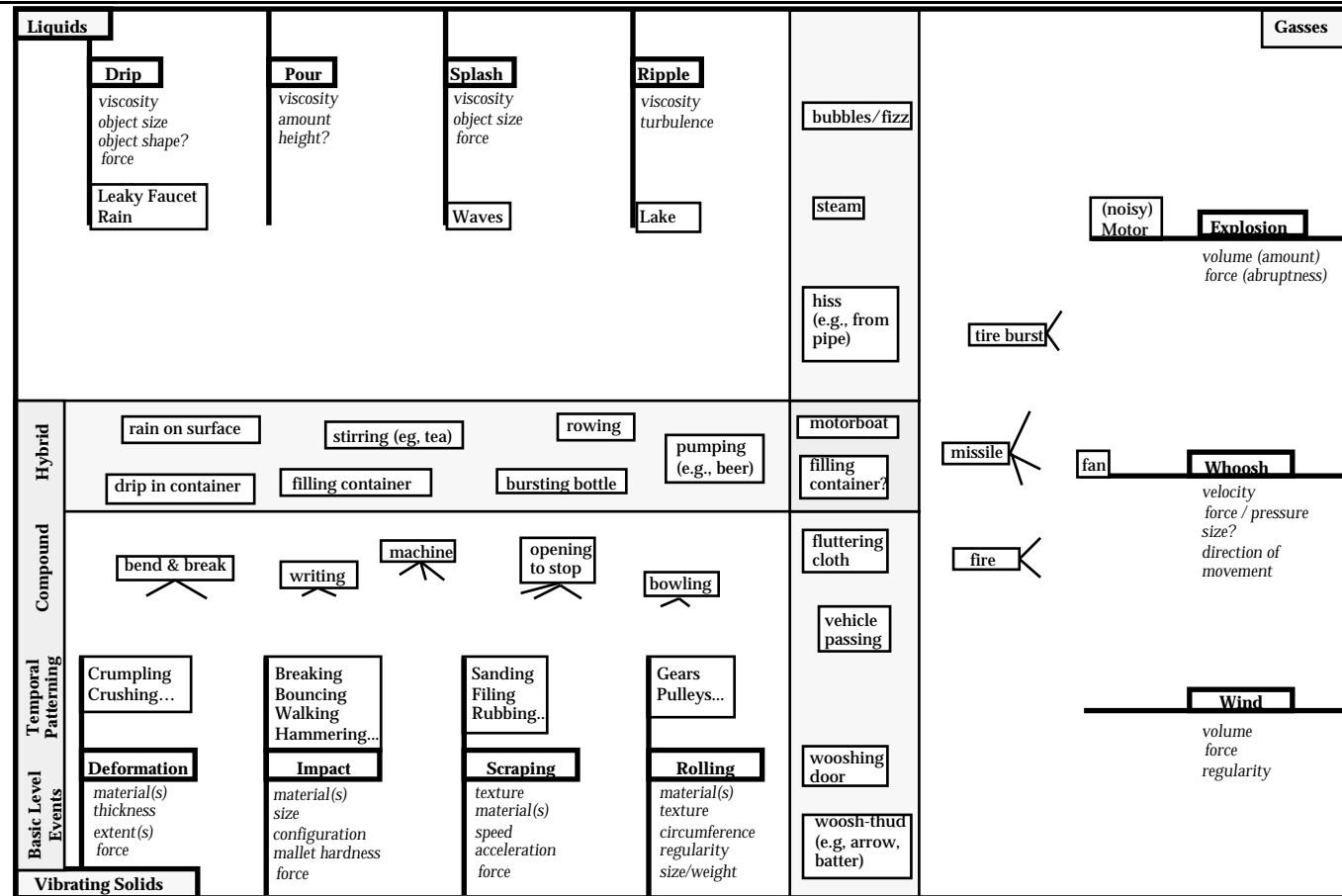


Figure 4. A framework for understanding everyday listening. Basic level events and attributes from three categories—solids, liquids, and gasses—are shown on the edges, with more complex events towards the centre (from Gaver, 1993b).

Of course, listeners are not always accurate in their identification of sound-producing events, or in evaluating the dimensions (e.g. size) of sources they do identify. Ballas (1994a, 1994b) has done extensive research on factors underlying the accuracy and speed with which listeners can identify everyday sounds. He found that simple acoustic variables were not related to accuracy and speed, but that the continuity of spectral information in continuous and discrete sounds was. Identification was also related to sound typicality. When primed with the name of an event, experimental participants were quicker to confirm that a typical sound could have been produced by the event than an atypical one, even if they would agree that the untypical one could have been too. Causal uncertainty, derived by asking participants to estimate the number of sources that could have produced a given sound (the more sources, the higher the uncertainty), is also related to speed and accuracy. Finally, Ballas and Mullin (1991) found context effects for identification of sounds. They

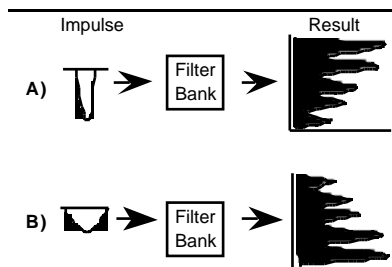


Figure 5. Objects can be modelled as filter banks, and impact hardness by impulses with varying degrees of sharpness (from Gaver, 1993a).

embedded target sounds in sequences that could be consistent, inconsistent, or random, and found that while an inconsistent context had large negative effects on both free and forced choice identification, a consistent one made only small positive effects. All these experiments were performed on relatively short sounds (less than .6 seconds), but they are valuable in indicating that people are best at identifying sounds that are typical, unambiguous with regards to source, and embedded in a congruent auditory ambience.

The second question for everyday listening is how do we hear it? What are the acoustic correlates to perceptual attributes of sources? If frequency is the major acoustic correlate of pitch perception, for example, what are the acoustic correlates to the perception of size, for instance, or texture? These questions can be addressed by comparing acoustic analyses of recorded sounds with experimental data about their perception. For instance, Freed (1988; Freed & Martins, 1986) has shown that people can perceive the hardness of a mallet used to strike a solid object, and that this may be conveyed by the ratio of high- to low-frequency energy in the resulting sound. Physical analyses of sound producing events are also useful in this endeavour. Wildes and Richards (1988), for example, showed analytically that the internal friction characterising different materials determines both the damping and definition of frequency peaks, and thus the material of a struck object may be perceptible. Finally, algorithms that attempt to synthesise sounds directly in terms of events and their attributes can also be useful. For instance, a given

object may be modelled as a filter bank, where the frequencies of the filters depend on the object's configuration and their bandwidths on the material (Gaver, 1993a; see Figure 5). Impact sounds can be produced by passing an acoustic impulse through these filters, where the sharpness of the impulse depends on the hardness of the striking object, while scraping sounds can be produced by passing a longer, noisy signal, where noise's spectral makeup depends on surface texture. Insofar as the sounds resemble those made by the actual events, the algorithms can be said to capture the acoustic correlates to perceived source attributes. Such algorithms are not only useful for understanding the acoustic bases of everyday listening, but are useful in allowing auditory icons to be synthesised and parameterised directly in terms of their source attributes.

Everyday listening represents a new approach to sound and hearing, one that is relatively unexplored as yet. We know little about how best to summarise perceptible events and their attributes in a general system that goes beyond focused descriptions of individual event classes. We know even less about how to characterise the acoustic information for events, the physical variables that underlie our perception of sound-producing events. Nonetheless, the perspective suggested by everyday listening has already proved valuable for design, helping to guide the creation of parameterised auditory icons that convey information as it is conveyed in the everyday world.

Localisation

Our ability to hear the location of sound sources, both in terms of their

distance and their direction, provides a meeting ground for research on psychoacoustics and everyday listening. On the one hand, localisation is a perceptual problem that has been well studied by traditional psychoacousticians. On the other hand, our ability to hear the location of a sound source, and even something about the environment in which it is played, is clearly related to everyday listening. Moreover, the ability to provide location information artificially for sounds has great potential for auditory interfaces, and especially for virtual reality systems. In this section, I briefly summarise the acoustics and psychoacoustics of distance and direction perception, and explain the principle behind systems which allow sounds to be located artificially. For more information, see Begault (1994), on which much of this discussion is based.

Distance

There are several cues for the distance of a source. To some degree, the amplitude of a sound will determine its perceived distance. Physically, the amplitude of a sound made by a point source (which radiates sound equally in all directions), decreases with the square of the distance, so that doubling distance corresponds to a change of 6dB, while line sources (for instance, a river or a road) decrease in amplitude at a rate of only 3dB for every doubling of distance. Perceptually, however, loudness, not amplitude, may be the important variable, such that slightly higher amplitude differences are needed to give the impression of distance doubling.

Amplitude manipulations alone do not always give a compelling

sense of distance, however. First, the effects of amplitude are stronger for unfamiliar sounds than they are for familiar ones (otherwise we might think a car passing ten feet away was closer than a watch held next to our ear). Second, most research on loudness has been done in anechoic environments, where echoes and reverberations are kept to a minimum. In a normally reverberant environment, the effects of sounds reflecting from nearby surfaces can drastically change the way that amplitude fades with distance.

Reverberation itself provides information for distance (as well as for other aspects of the environment). In a normal environment, sounds not only reach our ears directly from the source, but also after reflecting from various surfaces. These later reverberations, by definition, travel a longer path than the direct sound, and thus have a lower amplitude. But as a listener moves further from the direct source, the total amplitude of reflected sounds increases compared to that of the direct source. Thus increasing the R/D ratio, the ratio of reverberant to direct sound, can give a convincing impression of increasing depth up to the auditory horizon, at which point further increases in reverberation will not have an effect. Reverberation is a very effective way to produce depth impressions, one that is becoming increasingly easy to achieve with new technologies.

A final cue for distance is in the spectral content of a sound. There are two aspects to this. First, as sound travels through air it tends to lose high frequency energy to a degree that depends on factors like humidity and temperature. Second, the spherical waveform produced by

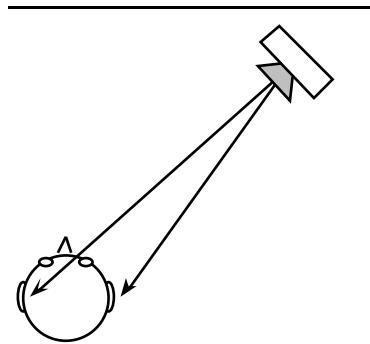


Figure 6. Sound travels along different paths to reach our ears.

a point source emphasises low frequencies to a degree depending on the waveform's radius, up to about 6 feet when the wavefront becomes effectively planar. There is little experimental evidence that either of these cues produce strong effects on distance perception. Nonetheless, experience suggests that low-pass filtering of a sound to diminish high frequencies can be used (especially with changes of loudness and reverberation) to make sounds seem more distant.

Direction

Direction is expressed with respect to the listener's frame of reference, with the azimuth and elevation of a sound's direction measured in angles horizontally and vertically from a point directly in front of the listener. The cues for distance described above need only one ear to work; they are monaural cues. In contrast, cues for direction are largely binaural, depending on differences in the sounds that strike our two ears. Sound waves produced by a given source travel different paths to each ear (Figure

6), producing two cues for direction on the azimuth. The first is the interaural time delay (ITD): The sound reaching the further ear will be slightly delayed with respect to that reaching the nearer (about .00065 seconds maximum). The second is the interaural intensity difference (IID): The sound reaching the further ear is shadowed by the head, and thus of lower amplitude than that reaching the nearer ear (this is basically the principle used in stereo systems). There has been a great deal of research on the details of both these sources of distance information (see Begault, 1994), but each can be an effective cue for direction.

The difference in sounds reaching the two ears is not sufficient to specify direction, however. In Figure 7, it is easy to see that the sound source could be

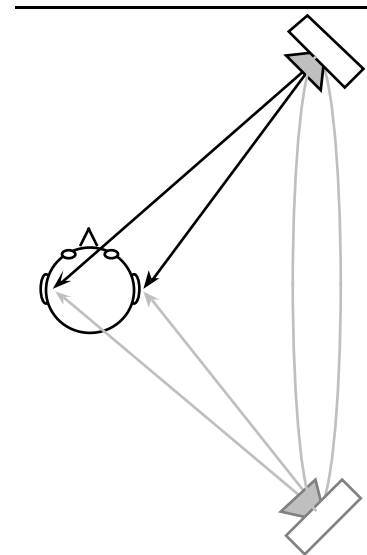


Figure 7. The cone of confusion.

behind the listener's head, with the sound waves travelling the same distance to each ear. In fact, the sound could be anywhere along the cone of confusion shown in Figure 7; each location on the surface of the cone would involve the same ITD and IID. Because of this, using time and intensity differences alone can give rise to both front-back and up-down errors.

One source of information for disambiguating direction along the cone of confusion is movement. Imagine that the listener in Figure 7 turned her head towards the source of sound, in front of her and to her right. As she oriented towards the source, both the ITD and IID would be reduced. If the source were actually behind her, turning towards the right would increase both differences. Not surprisingly, it has been shown experimentally that allowing listeners to move improves their ability to judge direction.

We need not move to hear the direction of a sound, however. It turns out that this is possible because our outer ears, or pinnae, filter incoming sounds before they reach our eardrums. As the sound waves pass over the curved, convoluted surfaces of our ears, energy at various frequencies is delayed by different amounts, and the pattern of these changes depends on which direction the sound is coming from. These spectral changes—which also depend on reflections and shadowing from the torso, shoulders, and ear canal—are referred to as head-related transfer functions (HRTFs). They can be measured by inserting a microphone deep into the ear canal, playing a known sound (e.g., a click) from many different directions, and recording the sound that actually

reaches the microphone. The results can be used to capture the spectral changes produced by the pinnae and body, as well as ITD and IID information.

Head-related transfer functions can be used to build systems that artificially locate sounds in virtual space. The Convolvotron (e.g., Wenzel et al., 1988) is a seminal example of such a system. Designed by Scott Foster and Beth Wenzel for NASA in the late eighties, it filters sounds (produced elsewhere) using HRTF data collected by Wightman and Kistler (e.g., Wightman and Kistler, 1989). When the resulting sounds are played over headphones, they produce an impressive illusion of a sound source at a location exterior to the listener (in contrast, stereo recordings played over headphones always sound like they are coming from inside the head). Moreover, the Convolvotron uses a head-tracking device so that listeners may move their heads, while the virtual sound source will seem stable. New versions of the system even allow the specification of localised early reflections from surfaces in the environment, giving a still more realistic impression of being immersed in a virtual sound space.

The Convolvotron was an impressive technical achievement for its time, and thus very expensive. But newer systems are appearing which make artificial localisation more approachable. For instance, the Focal Point system developed by Bo Gehring uses specialised audio cards on either Macintosh or PCs, while Burgess (1992) managed to implement a HRTF based localisation system in software for use on Sun workstations. Though not as sophisticated as the Convolvotron family of systems, both produce

reasonably realistic spatial impressions, and we may expect that more, better, and less expensive systems for localisation will appear over the next few years.

Music

Psychoacoustics, everyday listening, and localisation are all frameworks in which sounds are analysed. Music, in contrast, involves ways that sounds can be combined. Music has given rise to what are by far the most sophisticated systems for thinking about and manipulating groups of sound. It is structured at many different levels, from details of individual notes that can be understood in terms of acoustics and psychoacoustics, to overarching thematic and harmonic movements that give relatively simple shape to compositions that may last hours. These overlapping structures could be exploited by auditory interfaces to convey multileveled information. Moreover, music has profound expressive possibilities which are, at least to some degree, accessible to analysis and manipulation.

Auditory interfaces have so far drawn relatively little on the possibilities suggested by music (though see the section later on Musical Messages). One reason may be that the control needed for research on auditory interfaces implies a level of explicit articulation which the complexity of music resists. This situation contrasts with designers of multimedia or games environments, who happily exploit music's potential to create mood without needing to articulate exactly how they are doing so. In addition, it may be that many researchers on auditory interfaces are not themselves musicians (including myself, for instance), and that musicians have not been involved in

the design process early enough to deeply influence the approaches taken (a striking counter example is Cohen's 1994b OutToLunch system, described later). Nonetheless, it seems useful to speculate here about some of the possibilities for exploiting higher-level musical structures in auditory interfaces.

Rhythm offers a great deal of inherent structure that could be mapped to the structure of data or events that designers want to convey. Different rhythms are often used to distinguish musical messages, as Patterson did in creating the auditory alarms described earlier (Patterson et al., 1986; Patterson, 1982). But rhythm offers other possibilities beyond allowing us to tell one sequence of sounds from another. For instance, tempo, the overall repetition rate of a rhythm, might be used to convey a sense of activity, excitement, or even alarm. The number of divisions applied to a basic repetition rate (i.e., the use of quarter, eighth, sixteenth... notes) provides one way to build a sense of rhythmic complexity, as does the degree of syncopation used. These might be used to convey any quantity, and particularly those that are semantically related to complexity (e.g., the relative timing of related processes, as in the ARKola simulation described above). Finally, rhythm may be varied from measure to measure, building levels of repetition and change over increasingly long periods of time. This might be used to convey a sense of the overall flow of complex processes with many subordinate parts.

Melodic or harmonic structures also offer a rich resource for auditory interface design. The overall shape of simple tunes may suggest

semantic interpretations, a fact exploited in creating musical messages. But musical relationships offer meanings that have yet to be exploited. Major and minor keys often seem to sound happy and sad respectively, at least within western cultures. This sort of effect works at many different levels simultaneously. Individual notes set up varying patterns of tension and resolve within their harmonic contexts. The same is true for chords. At an even higher level, modulating from key to key sets up the same shifting patterns of extending and returning. Like rhythm, then, harmony might be used at a number of different organisational scales simultaneously.

A speculative example may help explain the uses of high-level musical structure I am suggesting here. Each time I read email from home, a complex and repetitive set of processes are set off. Imagine each one was linked to a melodic movement. A drone at the fundamental note might indicate that my reader is watching for mail on the server. When new mail comes, a melodic sequence building in tension and rhythmic complexity might play as each message is read into memory, then a related one, resolving the last both harmonically and rhythmically, might indicate that the message has successfully been written to disk. The chords from which these sequences were chosen could be varied for each new message, building and resolving tension at a larger level as all the mail was read. Finally, the overall key could be changed as I switched to my actual mail reader, first increasing tension as it accesses my spool file and reads the new messages, then resolving it again,

by shifting keys once again, as I actually start to read my mail.

This scenario might not work in practice (it is a bit melodramatic), but it captures the possibility of multilevel structure that music might offer interface design. Currently, auditory interfaces tend to work only at the lowest level, associating sounds with events defined at a single level. Using the hierarchical structure of music, we might be able to give equally meaningful information about multiple levels of events, so that sound could inform us, not only about a single email message being received, but about the entire process of reading email. In any case, it should be clear that musical structure is described along dimensions going well beyond pitch, loudness, timbre, and the other simple attributes of individual sounds described by acoustics and psychoacoustics.

Genre Sounds

Genre sounds are musical or everyday sound that are strongly associated with a given event, not because the event causes them, but because they have been linked with that event regularly over a long time. Telephone ringing sounds are a good example of such sound stereotypes. Although they are sometimes used mistakenly as examples of everyday listening, there is no necessary physical mapping between the sound and the fact that somebody is on the line waiting to be answered. Instead, particular ringing sounds are the result of design, particularly product design, and cultural standardisation. They have been remarkably stable within cultures, and so serve as almost irresistible symbols for telephone calls. But they differ substantially between countries, so

that US phone bells may not be recognised by Europeans. In addition, telephone rings are becoming more diverse with new phone designs, so that the classic ringing sound may become a kind of dead metaphor, reified by a culture that doesn't use it anymore.

Cohen (1993) pointed out the huge number of genre sounds with which people are familiar, and suggested their use in interface design. He used sounds from a cult science fiction television show for an interface, and found that most people had no problem recognising the sounds. Genre sounds can be drawn from a wide variety of sources—television and movies, cartoons, popular music, appliance, and so forth. In most cases, genre sounds themselves are particularly successful, or at least well-known, examples of auditory interface design. Telephone bells, the theme from *Jaws*, the “boing” sound indicating an impact in cartoon-worlds, are all examples of sounds that have been designed to convey certain meanings. In their historical success, they have the advantage of being recognisable, and thus potentially applicable to new designs.

Note, however, that recognising the sounds is not the same as recognising their meanings within a given interface. Genre sounds serve as a vocabulary for auditory interface design, which still must be mapped to the meanings to be conveyed. There are several potential drawbacks to using genre sounds. First, most genre sounds are culturally specific. Just as telephone bells vary between cultures, so do science-fiction shows, cartoons, and movies. Second, genre sounds are usually not easy to parameterise. They tend to serve as labels for

events, with no ready-made system for indicating variations of the basic message. Finally, genre sounds are mapped closely to their cultural meaning, and this may make them resistant to new mappings. Imagine, for instance, that the motive usually used to indicate an approaching shark is newly mapped to a reminder that you are to meet your boss. Even if this new mapping is remembered when the sound is heard, it is fairly certain that the old one will be too. This may interfere with hearing the intended message, and may well add an undesirable (at least to your boss) emotional tone to the interface.

Despite these problems, however, genre sounds are potentially a powerful resource for auditory interface designers. Well-known genre sounds are powerful indications of their messages, and can be a strong foundation for building new meanings. They are relatively easy to find, and to appropriate for new uses. Finally, they can be fun and expressive, enlivening auditory interfaces.

Idiosyncratic Device Models

The frameworks I have discussed so far are all relatively explicit. The devices we use to record, manipulate, and synthesise sounds, however, implicitly convey their own frameworks for thinking about sound and hearing. To some degree, the controls of such devices reflect dimensions described by acoustics and psychoacoustics, and, at a higher level, dimensions used in creating music. At the same time, however, many of their controls are much more idiosyncratic to particular devices. These controls embody their own models of sound, and in particular of timbre, the little understood but vital “quality” of

sounds. The device models they embody are important because they are the ones that auditory interface designers actually use to create informative sounds.

The degree to which device controls are standardised not surprisingly reflects accepted theories about psychoacoustics and music. Almost all sound-producing systems allow control over pitch, loudness, and duration. In addition, many devices allow pitch bending, a continuous change in frequency, which can be used to create glides between notes or microtonal variations of traditional scales. They also support vibrato and tremolo, repetitive microvariations of frequency and loudness similar to those produced by violin players. Finally, some parameters of timbre control are also reasonably standard. Many devices allow the “onset velocity” of sounds to be varied, which usually varies attack times and brightness (bandwidth), by analogy with the way that the sharpness and brightness of piano notes depends on the force with which keys are struck. It is also common to provide some control over sounds' amplitude envelopes, which determine how amplitude grows and decays within the course of the note duration, though the degree of control can vary considerably.

The standardisation of basic parameters offered by most sound-producing systems has both caused and been encouraged by the Musical Instruments Digital Interface (MIDI; Loy, 1985; IMA, 1983), a standard protocol created for communicating among computers and commercial signal processing devices (samplers, synthesisers, effects units, etc.). Typically, notes are triggered at designated pitches and “key

velocities," varied as they play via devices like pitch benders and foot pedals, and stopped on command. MIDI has been extremely successful in allowing the integration of sound systems, but it is limited in a number of ways. Perhaps most important is its adherence to a keyboard-centred view of control, which leads to a focus on triggering and stopping discrete notes. More continuous control parameters are not standardised, and the ways that sounds vary in response to continuous controllers must be defined using the devices themselves. For instance, it is easy to use MIDI to create simple motives, such as those used by Patterson (Patterson et al., 1986; Patterson, 1982) and Blattner (e.g., Blattner et al., 1989; see below). It is more difficult to use MIDI to make continuous variations to the timbres used in a motive, beyond selecting from a variety of device-dependent presets.

In dealing with timbre, devices vary widely, and MIDI does not offer a standard set of parameters for controlling them. Instead, timbres and control parameters are usually specified separately on each sound-producing device to be used. Devices differ in the ways they actually create sound, with the most basic distinction between samplers and synthesisers. These two classes of device tend to have different sounds, though some systems share characteristics of both. Samplers allow any sound to be recorded, and often manipulated to change its pitch, loudness, duration, attack rate, decay, and so forth. Synthesisers, in contrast, create sounds algorithmically. This means that, in principle, they allow almost complete specification of any sound. In practice, however, the sounds

synthesisers can create are limited by the algorithms they use. For instance, in realistic instrumental sounds, the amplitude of every component frequency varies independently and continuously. Early synthesisers, however, typically allowed only the specification of a simple amplitude envelope to be applied to a repetitive waveform; even in many current systems, independent control over each partial is not generally available.

The algorithms used to synthesise or manipulate sounds mean that some sounds are easier to create than others, and that some manipulations are easier to make than others. In effect, each technique carves up the "wastebasket category" of timbre differently, representing different ways to understand its underlying dimensions and attributes. Additive synthesis, for instance, involves adding together time-varying sine waves to create more complex sounds, in a sort of inverse to the Fourier analysis described earlier. For this sort of technique, the focus is on controlling parameters of these component sine waves. In contrast, frequency modulation (FM) synthesis, another powerful technique, involves controlling the frequency of a carrier waveform by the amplitude of another waveform. When this is done at audio rates, modulation gives rise to many "side-band" frequencies, thus providing an efficient way to generate rich sounds. For this sort of technique, the emphasis is on controlling groups of component frequencies, their density and frequency relations, rather than individual components themselves. Finally, synthesis based on physical modelling is becoming more prevalent. This allows control

parameters to be expressed in terms of virtual sound sources, such as the hardness of a drum head, or where a string is plucked.

Different techniques embody different frameworks for thinking about timbre, and these are not always comparable. Moreover, because creating complex sounds is computationally expensive, many synthesisers exploit nonlinear synthesis techniques. For instance, continuous changes of the carrier or modulator frequencies in FM synthesis produce categorical changes in the sounds that are produced, because only small integer ratios of carrier and modulating frequencies create harmonic sounds. This may mean that the control parameters offered by some devices are more closely related to their implementation than to perception, which is clearly a problem for their use. More often, though, these techniques simply offer different ways to think about sounds. It is difficult to control the amplitude and evolution of individual harmonics using FM synthesis, for instance, but likewise, it is difficult (and computationally expensive) to set up large numbers of hierarchically related partials using additive synthesis. Each technique affords its own style of sound creation and manipulation., its own way of thinking about sound.

The idiosyncratic control parameters offered by commercial equipment reflect the difficulty of establishing a common vocabulary for sound manipulations beyond the obvious ones of pitch, loudness, etc. Rushing to form a standard description of timbre would clearly be a serious mistake. Given the rich variations of sound lumped under the term, any such system would almost certainly be contrived and

incomplete. But the sheer range of techniques available in commercial equipment may paradoxically have the effect of constraining the possibilities that researchers explore in creating auditory interfaces. It is difficult to report new interfaces, or to transfer them to new equipment, if they depend deeply on specialised techniques. There may be a tendency to use preset voices instead of using the ability to create and control timbre along the many dimensions that are available. Clearly, this is unfortunate, given that timbre's underlying dimensions are very powerful resources for auditory interface design. The best tactic to take, it seems, is to understand the frameworks being used as well as possible, to exploit them fully in design, and to translate the effects employed into standard psychoacoustic terms in reporting the results. This will allow us to take full advantage of the resources offered by current equipment, without losing the ability to communicate about our designs.

AUDITORY INTERFACES

In this section, I review examples of working interfaces that use sound to convey information from computers. Three dimensions are useful in characterising these systems. The first has to do with the choice of sounds to be used—from simple multidimensional tones, to musical streams, to everyday sounds. This further implies a choice of frameworks for manipulating sound, as described in the last section. The second dimension has to do with the way sounds are mapped to information, from completely arbitrary mappings on the one hand to metaphorical and literal ones on the other. Finally, the third dimension concerns the kinds of

functionality that sounds have provided, from allowing exploration of multidimensional data to coordinating distributed work groups. Together, these three dimensions define a design space that encompasses existing auditory interfaces.

In the following section, I discuss the space of auditory interfaces in more detail. Here it is important to note, however, that sounds, mappings, and functions are not independent dimensions. Instead, systems have tended to cluster in certain areas of the space. In this section, I describe the three most well-developed groups: data auralisation, musical messages, and auditory icons. In the next, I point to ways that the design space can be filled out and extended.

Data Auralisation

One of the earliest and most recurrent form of auditory interface uses sound to support perception of complex, often multidimensional, data. The basic strategy is simple: data parameters are mapped to parameters of sound, forming individual tones or continuous streams of sound that change with variations in the data. By analogy to graphs or scatter plots, the perceptible variations in the sounds thus formed may be used to form an impression of patterns, groups, or variations in the represented data. These techniques, then, serve as an auditory analogue to data visualisation.

Patterns in Multidimensional Data

The essence of a good auralisation is to allow multidimensional data points to be perceived as integral

wholes, so that high-level patterns become perceptually available. To accomplish this, a good knowledge of psychoacoustics is necessary to avoid pitfalls such as mapping data linearly to sound dimensions that are heard logarithmically. In addition, systems should offer a wide variety of sound dimensions and manipulations, and allow users to explore their data sonically once these mappings have been established. Ideally, the auralisation should present data so that people can perceive structures at a higher level of organisation than they are represented by the computer.

Consider the iris data described at the beginning of this section, for example. The length and width of the petals and sepals (the four parameters used for classification) are integral aspects of every iris, so that it is possible to discriminate species just by looking at them. Measuring these variables abstracts them from their original context, requiring that they be approached analytically if patterns are to be found. The goal of auralisation is to reunite the variables in a single presentation. Thus when Bly (1982a, 1982b) mapped the iris data to sounds, the integration of the four variables could be heard in much the same way that a flower is seen.

Traditional graphing techniques also have the goal of bringing multidimensional data together into a single presentation. The location of each point in a scatter plot, for instance, conveys information about two dimensions simultaneously. Beyond using the three spatial dimensions, however, graphical techniques become problematic.

Although other attributes, such as colour, shape, or size, can also be used, it is surprisingly difficult to come up with many good candidates. Mapping dimensions to more literal pictorial representations can be effective (e.g., Holmes, 1993), but pictures can be misleading when data dimensions only map approximately to pictorial ones (Tufte, 1990, 1983). In any case, the results of such mappings can be cluttered, and patterns difficult to see. Sound has the advantage that it is naturally multidimensional, perceived integrally, yet abstract enough not to mislead listeners.

Auralisations can complement visualisations in presenting complex data. Bly showed this in another example, in which she mapped sounds to six-dimensional data sets that overlapped in any five dimensions, and asked people to classify sounds based on examples from each set. The technique was again effective, with subjects scoring as well using the auditory display alone as they did using visual scatterplots. Moreover, they performed even better with a mixed auditory and visual display. This suggests that the sounds and graphics can work together in allowing the perception of high level patterns.

Auralisation is also clearly relevant for the visually-impaired. Lunney and Morrison (1990; 1981; Lunney et al., 1983) use sound to present the results of infrared spectrometry, an important tool for identifying inorganic compounds. They preprocess a continuous infrared spectrum, replacing

absorption peaks with single vertical lines. The frequency of these peaks are then mapped to notes from several octaves of a chromatic scale. The sets of notes thus produced is presented in three ways. In the first, notes representing peaks are played in a descending scale, with the absorption intensity mapped to the duration of the corresponding note. Second, notes are played in order of decreasing peak intensities, with all notes of equal intensity. Finally, the notes for the six strongest peaks are played in a chord (this is the most concise but least memorable representation). Informal testing has shown that these representations allow reliable matches to be made between test and sample spectrograms. The ability to listen to explore the data in three different forms seems particularly useful, with each representation allowing the data to integrate in new ways.

Time-Varying Data

Because sound is an inherently temporal medium, it is well suited for representing streams of data that vary with time. Bly (1982a, 1982b) demonstrated this possibility by representing simulated battles between two opposing armies. Each army was mapped to a distinctive timbre. The number of troops at the front was indicated by pitch, while intensity was used to encode the number approaching the front. The resulting "battle songs" enabled observers to distinguish the evolution of battles with the same outcome, though listeners could not always track the individual streams of sound.

Similarly, Mezrich, Frysinger and Slivjanovski (1984) used a visual and auditory display to track changes in four dimensional economic data over time, and tested the ability for people to perceive correlation among the dimensions. The visual display was created by mapping each data dimension to a pair of vertical lines that grew and shrank symmetrically around the middle of the display (Figure 8). The auditory display was created by mapping the four parameters to notes over three octaves of a major scale. Thus each data dimension was associated with a pattern of moving lines, and with a stream of sound over time, with the four dimensions together creating a polyphonous "tune" which rose and fell with the changes of data. When people used this system, they could perceive much smaller correlations than they could using either overlaid or stacked graphs. Although they did not assess the utility of the auditory display alone, Frysinger and Slivjanovski's (1984) results again indicate the effectiveness of this sort of technique, and particularly that of integrated dynamic auditory and

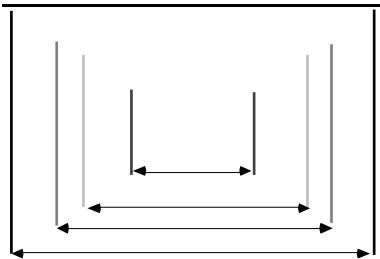


Figure 8. Frysinger and Slivjanovski (1984) mapped time-varying data to mirrored, vertical lines which appeared to move in and out as data changed over time.

visual displays.

Interactive Auralisation

While most auralisation strategies have focused on presenting data to listeners, Grinstein, Smith and their colleagues (Smith et al., 1990; Grinstein and Smith, 1990; Smith et al., 1994). have developed a system that allows interactive control over auralised data. The system involves creating visual and auditory "textures" from large amounts of data, from which higher-order variations and patterns may be detected. Visual textures are created by mapping each multidimensional data point to a simple multi-limbed stick figure called an "icon" with the length, intensity, and angles determined by individual data dimensions. Two data dimensions are typically used to specify the horizontal and vertical placement of each icon, creating a texture field when many are used. Each icon is also associated with a sound (usually short tones or noise bursts), which may be determined by the same or different data dimensions. Sounds are played as the user moves the mouse over the icons, so that an aural texture may grow and change as multiple icons are swept. This aural texture can be used to identify and explore different regions of the data space, and allows discrimination even of data structures that may be invisible. The significant development of this system is that it allows people to explore and probe the data space quickly and interactively, which may be expected to help them better understand its structure.

Towards General Auralisation Tools

Recently there has been a move away from designing specific auralisation techniques towards

designing more general hardware and software systems that allow experimentation with many different sorts of auralisation for a variety of application domains. A good example of this is the Kyma system, designed by Scalletti and her colleagues (Scalletti, 1994, Scalletti and Craig, 1991). Kyma is an object-oriented, stream-based system which allows direct manipulation of data sources, sound streams, and a variety of manipulators. In particular, it offers a suite of prototype tools which exemplify some of the most useful techniques for turning data into sound:

- Shifter interprets a data stream as an acoustic waveform, potentially shifted from the sub or ultra sonic range to be audible, allowing it to be heard "directly". Shifters can also—and are usually—used to produce data streams that control other parameters.
- Mappers implement the basic auralisation strategy of allowing data streams to control sound parameters such as frequency, deviations from a base frequency, amplitude, low pass filters, event density, and stereo panning.
- Combiners arithmetically combine two or more sound streams, via sums, differences, or products. Using amplitude modulation, for instance, in which the two streams are multiplied, frequencies are produced that correspond to the magnitude of their differences, as well as to one of the original streams.
- Comparator is a variation of a combiners that plays two sound signals simultaneously to the right and left stereo channels respectively, allowing the signals to be compared for the timing and degree of any differences.
- Markers allow sounds to be triggered when specific conditions are met—for instance, when a signal exceeds a threshold.
- Histogram represents classes of data by single frequencies, and their magnitude by the corresponding amplitude. The result can sound like a single time-varying timbre, a chord, or a collection of voices.

These tools, in combination with standard arithmetic operators, allow the construction of any of the specific auralisation schemes described above. Thus they are useful in abstracting away from specific case studies to a more general range of techniques. For instance, though the Shifter allows data to be heard directly, which would appear the simplest form of data auralisation, Scalletti (1994) reports that it is useful only for quasi-periodic and slowly changing data, such as tide movements, frequency measurements, and the like. The Mapper, on the other hand, is a generally powerful tool that captures the essence of most auralisation strategies, including, for instance, Bly's (1982a, 1982b), Frysinger and Mezrich's (1984), and Grinstein and Smith's (1990). Finally, the sonic histogram seems functionally equivalent to the chord mapping used by Lunney and Morrisson (1990) to represent infrared spectra; it has also been found useful in

exploring other sorts of data such as the effects of fire prevention on forest ageing (Scalietti, 1994). The system suggests other strategies as well, such as amplitude modulation to compare streams, or markers to signal significant events. In general, tools such as Kyma may be expected to encourage many new and different applications of data auralisation.

Auralisation: Discussion

The basic strategy of mapping data parameters to dimensions of sound appears to be simple, powerful, and easy enough that systems using this strategy have appeared with regularity over the last number of years, most applying variations of the sort of strategies discussed above to specific application areas. The advent of systems like Kyma, and the increasing power of PC-based signal processing in general, is likely to increase this trend, and to make auralisation increasingly available to non-expert users. This is encouraging insofar as the work described here indicates that sound can be a useful way to represent data for exploration. Nonetheless, auralisation remains a less mature endeavour than visualisation. A number of issues may be responsible for this.

First, progress has been slowed by hardware and software limitations on sound synthesis and processing. Much of seminal work described here was accomplished with much difficulty: for instance, Bly (1982a) had to build special hardware and a custom computer interface in order to undertake her work. The practical difficulties of using sounds has tended to seriously constrain the ability to experiment with novel sound parameters and mapping techniques. Recently,

sophisticated digital signal processing has moved from being the exclusive province of academic computer science departments to being more widely available on PC's and off-the-shelf synthesisers, effects units, etc. Even so, these new tools offer only partial support for auralisation; as Smith et al. (1994) have pointed out, MIDI equipment—and most software as well—is designed for making popular music, and thus does not always allow the appropriate controls for auralising data.

In part because of these limitations, widespread publication of sound, especially in academic contexts, is much rarer than publications of graphics. It is common to receive academic journals with a variety of printed graphs, charts, and visualisations, but rare to receive an audio CD or cassette. In addition, graphics can be seen on the page, while sound examples require playback systems extrinsic to the publication. This barrier, too, may fall as interactive CDs and the Internet become more popular outlets for publication, allowing the seamless incorporation of auralisations into research reports.

A more fundamental issue for using sound to explore data involves it being a temporal medium, rather than one that is spatial and relatively static, like vision. The ability to scan a visualisation, to focus on particular regions, to quickly look back and forth between different parts all combine to help visualisations support the perception of data patterns. The equivalent of these abilities seems difficult to define for sound, much less to provide. The ability to interact with an auralisation, as demonstrated by Smith et al.'s work (1990), is a

significant step in towards providing similar capabilities, but they may not come naturally to this medium.

With the serial nature of sound comes other difficulties. By their nature, data visualisations and auralisations tend to involve many arbitrary mappings between the semantics of data dimensions and values and their representations. For visualisations, tools such as keys, scales, and indices have developed to represent those mappings directly. For auralisations, such tools have not been well developed, and there are practical difficulties for doing so. For instance, reference tones might be played to indicate the extremes of a given dimension, but if many dimensions need to be indicated it is impossible to play them all at once, and playing them sequentially raises memory problems. Again, a high degree of interactivity may provide a solution to this problem, but these issues have only begun to be addressed.

Perhaps most importantly, though there are many examples of auralisation, there are few or none in which auditory representations have been shown clearly to lead to new insights about real data (Scalietti, 1994). This may be equivalent to saying that there has been no "killer app" for auralisation, and asking or a single convincing example of its effectiveness (rather than a host of lesser examples) may be unfair. Nonetheless, most work so far has focused on techniques and implementations of auralisation rather than on specific insights about real content areas. It may be only when it is natural to report the content of data first, and the auralisation techniques used to discover it second, that these techniques will have come of age.

Musical Messages

Auralisations focus on using sound to understand data, a domain in which the computer is a tool used to accomplish some task. Most other auditory interfaces are concerned with supporting interaction with the computer itself, by providing information about events, objects, and processes within it.

One strategy for conveying information about events using sound is to create auditory messages from a musical vocabulary. Patterson's alarm sequences, described at the beginning of this chapter, are an example of this strategy (Patterson et al., 1986; Patterson, 1982). Different alarms are created by varying their rhythms, melodies, and timbres to create short, distinctive tunes. These map to the messages they are meant to convey in a variety of ways. Sometimes the mapping is simply arbitrary, and must be memorised by listeners. Other times the tunes are modelled on the intonation and rhythm of the equivalent spoken message, or on a sort of analogy to the message to be conveyed.

The motives used by Patterson and his colleagues (Patterson et al., 1986; Patterson, 1982) are the culmination of work focused primarily on the basic psychoacoustical requirements for effective musical messages. His work is exemplary in showing how sounds can be shaped to fit the auditory environment in which they will be heard, using acoustic analyses in the design of sounds that will be loud enough to be heard, while not so loud as to interfere with communication or concentration (a real problem for commercial aircraft alarms, for which existing alarms have often used mechanisms

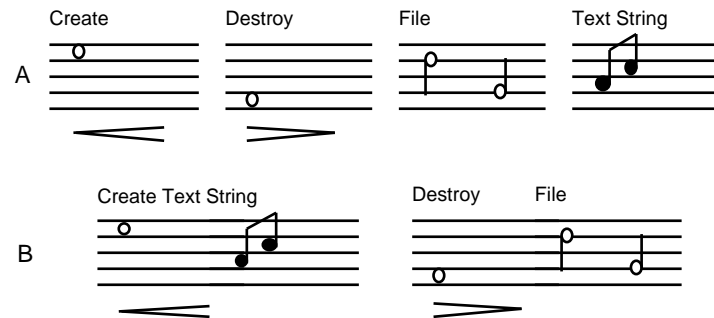


Figure 9. Simple earcons A, can be concatenated to form combined earcons B (after Blattner 1989).

originally designed for World War II bomber cockpits). In addition, he and his colleagues (Patterson et al., 1986; Edworthy et al., 1991; Swift et al., 1989) have pioneered work on the acoustic factors that make sounds more or less urgent, showing that high frequencies, abrupt onsets, inharmonic timbres, and changes in frequency and amplitude all contribute to urgency. Finally, his use of short alarm bursts punctuated by silence is a good example of the restraint desirable in creating auditory cues, especially those that are meant to provide information over protracted periods of time.

Earcons: Creating Families of Musical Messages
Blattner and her colleagues have developed a similar approach, applying it to the domain of computer interfaces (e.g., Blattner et al., 1989; Blattner et al., 1994; Stevens et al., 1994; Brewster et al., 1994). The idea is that such messages could be used to indicate a wide variety of events in the interface, such as error conditions or feedback about user operations.

However, where Patterson focuses on psychoacoustical factors, Blattner has been concerned with developing a systematic methodology for designing families of musical messages in which related messages sound similar. The system she has developed uses simple motives as building blocks, which are used to construct more complex cues called earcons—a bad pun first used by Bill Buxton—through combinations, modifications, and hierarchy.

Motives are essentially very short melodies used as individual, recognisable entities (Blattner et al., 1989). For example, Peter and the Wolf uses motives to stand for the various characters; similarly, the familiar pulsing sound used to indicate the shark's presence in *Jaws* is also a motive. Blattner et al. (1989) identify rhythm and pitch as the fixed parameters of a motive, those which are used to give each its individual identity. Families of motives are created using variable parameters of sound, such as timbre, register, and dynamics. According to Blattner et al. (1989),

these parameters may be varied fairly widely, while leaving the basic identity of the motive unchanged.

Some of the basic recommendations for earcon construction suggested in Blattner et al. (1989) have been modified with experience and experimental work. For instance, early examples of motives used very simple waveforms (e.g., sine, square, and triangle waves) as variations of timbre. However, experimental work by Brewster et al (1994) has shown that such waveforms are often confused, so more recent work has tended to use instrumental timbres as found on MIDI controlled synthesizers. Similarly, Blattner et al. (1989) recommended using pitches from a musical scale within a single octave to avoid octave confusions, but again, Brewster et al. (1994) found that this makes motives relatively indistinguishable from one another and recommend that a wider range of pitches be used.

More complex earcons, and families of earcons, can be constructed from simple motives in several different ways. Combined

earcons are created by playing two or more basic ones in succession. For instance, Figure 9A shows simple earcons denoting two operations (create and destroy) and two entities (file and text string). Figure 9B shows combined earcons made by concatenating these simpler elements, representing "create text string," and "destroy file;" other combinations can obviously be made in a similar way.

Families of earcons can also be made by constructing hierarchies in which earcons at each subordinate level are differentiating using a new musical parameter. For example, a class of error messages (Figure 10) can be defined by a rhythmic pattern of unpitched sounds (i.e., clicks). Different classes of error messages—in this case, operating system vs. execution errors—can be given different melodies, while specific messages (e.g., "file unknown" or "underflow") can be distinguished by using different timbres.

Blattner et al. (1989) suggest playing the earcon corresponding to each level of a hierarchy in series, so

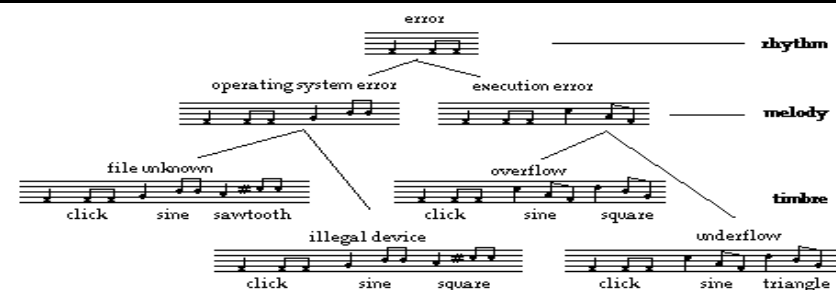


Figure 10. A family of error messages can be built as a hierarchy in which each subordinate level is distinguished by a new musical parameter. In this example, the earcon corresponding to each level is played sequentially. Combining all cues into a single rendition of the motive (the last in each example) is also possible. (After Blattner et al., 1989).

that, for instance, the earcon for “overflow” would consist of first the rhythmic pattern indicating “error,” then the melody indicating “execution error,” and finally the melody using a square wave timbre that indicates “overflow.” This strategy has the advantage of being very clear for beginners, but produces earcons that may be unacceptably long—especially considering that hierarchies of five or more levels are possible according to their scheme. A more succinct version of these earcons would use only the earcon associated with the final nodes of the hierarchy, so that “underflow” could be indicated simply by the square-wave version of the basic melody. These are called transformed earcons by Blattner et al. (1989), because the family of earcons in this case is created using musical transformations of an implied prototypical earcon, rather than by explicitly traversing the branches of a hierarchy.

Using Earcons to Create an Auditory Map

Blattner et al. (1994) report on the design on an auditory map, in which earcons were added to an interactive, digital aerial view of Lawrence Livermore National Laboratory. Sounds played as the cursor was moved over various buildings on the map, indicating access privileges, administrative units, and computers. The earcons were associated to these entities by way of two sets of analogies. First, entities were mapped to actions that could be visualised: knocking for access privileges, a schematic tree diagram for administrative units, and a visual icon of a computer for computers. Then musical analogues were found for each of these visual images: a uniform tom tom drum

sound for knocking, three notes on a saxophone to stand for walking up a tree, and a “whirring noise sounded out by four notes on a flute” to indicate computers. Animated versions of the visual analogues were presented when the map was opened to help users learn the meaning of the sounds.

The earcons were combined, transformed, and used inheritance to indicate more than the mere existence of an entity at a particular location. The level of access privileges was mapped to frequency, so that higher required privileges sounded like higher, faster knocking. Summary data about the number of administrative units and computers in a given area was indicated by varying the amplitudes of their respective sounds. In addition, very short versions of each earcon—consisting only of a short tone of the appropriate timbre—were played as the user moved the cursor over a region, to avoid having to play the entire messages.

Algebra Earcons: Summarising Equations for the Visually-Impaired
Stevens et al. (1994) developed an innovative system that used earcons to represent algebraic equations. The system was targeted at visually-impaired users, and was meant to give the sort of high-level information about an equation’s structure that readers get at a glance—information about an equation’s length, complexity, the presence of fractions, and so on. The intent was to provide the sort of information used by sighted readers to plan their reading of an equation, and to form an idea of the structure into which various terms should fit. Moreover, Stevens’ system allowed users to explore mathematical

structures in more detail, by navigating in and out of different levels of auditory information.

Sounds were designed to reflect the syntax of algebra equations. Different syntactic types (base level operands, binary operators, relational operators, superscripts, fractions, and subexpressions) were mapped to different musical timbres (e.g., pianos, pan pipes). The timing, pitch, and amplitude of each sound was then varied according to rules based on prosodic cues. For instance, each new term in an equation started at middle C, except the last one which was played at A4 to mimic the way people’s voices drop as they finish reading an equation. In addition, some of the sounds were designed to map intuitively to their referents. For instance, superscripts and subscripts were at high and low pitches, respectively, matching their graphical appearance.

The strategy was tested using a forced choice test in which subjects heard a target equation, and then picked its written version from three distracters. Since the distracters were designed to differ from the original equations in only one respect, the data could be used not only to assess the strategy’s overall effectiveness, but particular problems as well. The system was accordingly redesigned, and retested using the previous subjects. Most of the subjects showed marked improvement and, moreover, most scores increased on the previously problematic equations. Though the experiments did not directly address the question of whether the system will give visually impaired users the same sort of information that sighted people obtain by glancing at equations, they do indicate that people can recover

syntactic information about equations using these sounds.

This system is an excellent example of the potential for tailoring earcon design to specific tasks. In general, the rules used to generate these “algebra earcons” do not reflect the general principles for designing earcons suggested by (Blattner et al., 1989), though they can be seen as examples of combined earcons. Instead, the system builds on the original principles, basing its mappings on more specific considerations of the syntax of equations, and on analogies to prosody and typographic layout. This approach allows an appropriate mapping between algebraic and sound structures to be realised, one which matches relatively abstract sounds to the abstractions of mathematics.

Earcons: Discussion

There are relatively few examples of systems that use earcons, and many of the ideas from the original proposal—most notably, that of creating rich hierarchical families of earcons—have yet to be demonstrated in working prototypes. Nonetheless, the strategy is an interesting one for creating auditory cues. The appeal is in creating a system which allows the systematic development of families of auditory cues from simple, readily-created musical sounds. It makes use of people’s ability to make symbolic mappings between essentially arbitrary combinations of representations and referents, as when using a ballpoint pen to stand for a bus in a story about motorcycling home.

In addition, the idea of creating hierarchical families of cues from basic motives is a powerful one, particularly when all the variations

are incorporated in one sound (c.f. parameterised auditory icons, described in the next section). This strategy makes use of people's ability to listen to sounds at different levels of detail, perhaps obtaining information at a high level ("that's an error sound") even if they cannot distinguish lower levels (e.g., confusing overflow and underflow errors). An everyday analogy to this ability is our differing sensitivities to automobile noises: I may take my car to a mechanic because "it's making an awful rattling noise," while to her it may sound clearly like a faulty carburettor. Even the high-level information can be useful, and with experience and motivation we may learn to understand the details as well.

There are several potential problems with earcons, however. Perhaps most important, they tend to be arbitrarily mapped to their referents, which means that they must be learned without benefit of past experience. I will address this issue at greater length in the next section. In addition, the tendency to use musical sounds, as well as arbitrary mappings, suggests that there will be difficulty in designing them to integrate with graphical interfaces. Musical phrases may not mesh well with working environments, as well, especially as people tend to find repetitions of simple tunes annoying (Gaver & Mandler, 1987). Finally, earcons tend to have relatively long durations. While this may be necessary to achieve good recognition (especially in experimental trials with limited training), the designs are often less subtle and more intrusive than might be possible.

These problems do not seem fatal for the basic approach,

however. It is easy to imagine systems which use earcons designed carefully to be short, subtle, and recognisable. In fact, we already have good examples of what such systems might sound like: The melodic chirping sounds made by many of the devices in Star Trek seem recognisable as future cousins to the earcons that have so far been demonstrated.

Auditory Icons

Another way to use sound as an integral part of the interface involves creating auditory icons, everyday sounds mapped to computer events by analogy with everyday sound-producing events (e.g., Gaver 1993b, 1986). Auditory icons are like sound-effects for computers: For instance, selecting a file icon in a graphical user interface might make the sound of a notebook being tapped, with the type of file indicated by the material of the object, and the file size by the size of the struck object. In contrast to earcons, which rely on essentially arbitrary mappings between sounds and interface entities, auditory icons are similar to visual icons in that both rely on an analogy between the everyday world and the model world of the computer.

Like earcons, auditory icons can be varied to produce "families" of related sounds. Unlike earcons, this is done by parameterising auditory icons along dimensions relevant for events that make sound. Thus auditory icons can reflect not only categories of events and objects, but can also reflect their relevant dimensions as well. For instance, the material involved in some sound-producing event might be used to represent the type of interface entity involved in a computer interaction. In this case, all auditory icons

concerning that type of object would involve sounds made by that kind of material. So text files might always sound wooden, but selecting them might make a tapping sound, moving them might make a scraping sound, and deleting them might make a splintering sound. The result is a system that maps the attributes that are perceptually relevant for a given sound-producing event to those relevant for the event to be represented. In this way, a rich system of auditory icons may be created that relies on relatively few underlying metaphors.

When the same analogy underlies both auditory and visual icons, the increased redundancy of the interface may help users learn and remember the system. For instance, one novice user reportedly didn't realise that icons stood for files until hearing the sounds made when they were selected. In addition, using auditory icons may allow more consistent model worlds to be developed, because some computer events may map more readily to sound-producing events than to visual ones. The timing and nature of difficult to visualise events like disk accesses may be one example of this. Finally, making the model world of the computer consistent in its visual and auditory aspects seems to increase users' feelings of direct engagement (Hutchins et al., 1986) with that world—the feeling of working in the world of the task, not the computer.

A number of systems have been created which illustrate the potential for auditory icons to convey useful information about computer events. These systems suggest that sound is well suited for providing information:

- about previous and possible interactions,

- indicating ongoing processes and modes,
- useful for navigation, and
- to support collaboration.

In the following sections, I describe a variety of systems that have used auditory icons.

The SonicFinder: Extending a Graphical User Interface

The SonicFinder (Gaver, 1989) was the first system to incorporate auditory icons. Developed for Apple, it was an extension to the Finder, the application used to organise, manipulate, create and delete files on the Macintosh.

The SonicFinder extended the underlying code for the Finder to play sampled sounds modified according to attributes of the relevant events. This allowed auditory icons to accompany a variety of interface events (see Table 1). Most of the sounds were parameterized, although the ability to modify sounds on the Mac was limited at the time the SonicFinder was developed. So, for instance, sounds which involved objects such as files or folders not only indicated basic events such as selection or copying, but also the object's types and sizes via the material and size of the virtual sound-producing objects. In addition, the SonicFinder incorporated an early example of an auditory process monitor in the form of a pouring sound that accompanied copying and that indicated, via changes of pitch, the percentage of copying that had been completed (auditory process monitors have been explored more extensively by Cohen, 1994a, as described later).

The SonicFinder was never released commercially by Apple, largely because the size of the sampled sounds was deemed too large (at the time, Macintosh system releases could be distributed on single 800K floppies). However, it was distributed quite widely in an informal fashion, with users in North America, Europe, Australia, and Asia. No formal user testing or evaluation studies were ever carried out on its reception, but anecdotal evidence suggests that while some people stopped using it after a short time, others found it valuable, or at least appealing, and when it finally

became obsolete due to new system releases, lobbied for its return. Because the Macintosh is a single-processing machine with a fairly simple interface, the sounds used in the SonicFinder basically provided feedback and information about possible interactions (as well as more general information about file size and type, dragging location, and the like). Nonetheless, the fact that it was a popular extension of a widely-used interface meant that it provided a valuable example of the potential of auditory icons, showing that sounds such as these can be incorporated in graphical user

interfaces in intuitive and informative ways.

SoundShark: Multiple Sounds, Navigation, and Collaboration
The SonicFinder was useful in incorporating auditory icons into a well-known and often-used interface, but the simplicity of the Macintosh's operating system and the finish of its interface limited the possibility of exploring novel functions for auditory icons. For this reason, Gaver and Smith (1990) demonstrated their use in a large-scale, multiprocessing, collaborative system called SharedARK, and dubbed the resulting auditory interface SoundShark.

SharedARK was a collaborative version of ARK, the Alternate Reality Kit. Developed by Smith (1989), ARK was designed as a virtual physics laboratory for distance education. The "world" appeared as a huge flat plane over which the screen could be moved, on which a number of 2.5D objects could be found. These objects could be picked up, carried, and thrown using a mouse-controlled "hand." They could be linked to one another, messages could be passed to them using "buttons," and, because SharedARK used a multiprocessing system, their associated processes could run simultaneously. In addition, SharedARK allowed the same world to be seen by a number of different people on their own computer screens (and was usually used in conjunction with audio and video links that allowed users to see and talk to one another). They could see each other's "hands", manipulate objects together, and thus collaborate within this virtual world. The result of all this was a large and complicated world, with pockets of activity separated by blank spaces:

navigation and orientation were problematic in this environment.

This interface was extended by adding auditory icons to indicate user interactions, ongoing processes and modes, to help with navigation, and to provide information about other users. Sounds were used to provide feedback as they were in the SonicFinder: Many user actions were accompanied by auditory icons which were parameterized to indicate attributes such as the size of relevant objects. Collaborators could hear each other even if they couldn't see each other, which seemed to aid in coordination. In addition, ongoing processes made sounds that indicated their nature and continuing activity even if they were not visible on the screen. Modes of the system, such as the activation of self-perpetuated "motion," were indicated by low-volume, smooth background sounds. Finally, distance between a given user's hand and the source of the sound was indicated by the sounds' amplitude and by low-pass filtering, aiding with navigation. The apparent success of this manipulation led to the development of "auditory landmarks," objects whose sole function was to play a repetitive sound that could aid orientation.

ARKola: Sound Ecologies and Peripheral Awareness

Experience with SoundShark suggested that auditory icons could provide useful information about user-initiated events, processes and modes, and about location within a complex environment. This work led to the development of the ARKola study (Gaver et al., 1991) described at the beginning of this chapter. The ARKola study allowed exploration of issues concerning the design of

Events	Auditory icons
Icons	
<ul style="list-style-type: none"> • Selection type (file, application, folder, disk, trash) size • Opening size • Dragging size where (windows or desk) possible drop in • Drop-in destination size • Copying amount copied 	<ul style="list-style-type: none"> • Hitting sound source material (wood, metal, etc.) source size (frequency) • Whooshing sound size (frequency) • Scraping sound size (frequency) texture (bandwidth) selection sound of container • Noise of object landing size (frequency) • Pouring sound frequency
Windows	
<ul style="list-style-type: none"> • Selection • Dragging • Growing size • Scrolling revealed area 	<ul style="list-style-type: none"> • Clink • Scrapping • Clink on release size (frequency) • Tick sound size (frequency)
Trashcan	
<ul style="list-style-type: none"> • Drop-in • Empty 	<ul style="list-style-type: none"> • Crash • Crunch

Table 1. Events, sounds, and parameters used in the SonicFinder (after Gaver, 1989).

many sounds that could be played continuously and simultaneously, the possibility of using sound to integrate perception of complex processes, and the use of sound in supporting collaboration.

With as many as 12 sound playing simultaneously, designing the sounds so that all could be heard and identified was a challenge. In general, temporally complex sounds were used to maximise discriminability, and the sounds were designed to be semantically related to the events they represented. Two strategies were found to be useful in avoiding masking. First, sounds were spread fairly evenly in frequency, so that some were high-pitched and others lower. Second, continuous sounds were avoided, and instead repetitive streams of sounds were used to maximise the chance for other sounds to be heard in the gaps between repetitions (cf. Patterson's use of intermittent alarms).

As expected, the sounds helped people keep track of the many ongoing processes, allowing them to track the activity, rate, functioning, and problems with individual machines. More surprisingly, the auditory icons merged together in an auditory texture that encouraged people to hear the plant as an integrated complex process. Sounds were also used as an important resource for collaboration. Without sound, participants had to rely on their partner's reports to tell what was happening in the offscreen part of the plant. With sound, each could hear events and processes they could not see. This shared audio context seemed to lead to greater collaboration between partners, with each pointing out problems to the other, discussing activities, and so forth. The ability to provide

foreground information visually and background information using sound was essential in allowing people to concentrate on their own tasks, while coordinating with their partners about theirs.

Finally, sound also seemed to add to the tangibility of the plant and increased participants' engagement with the task. This became most evident when one of a pair of participants who had completed an hour with sound and were working an hour without remarked "we could always make the noises ourselves..." In sum, the ARKola study indicated that auditory icons could be useful in helping people collaborate on a difficult task involving a large-scale complex system, and that the addition of sounds increased their engagement as well.

EAR: Environmental Audio Reminders

Where SoundShark and ARKola explored the use of auditory icons to support collaboration in software systems, another system, called EAR (for Environmental Audio Reminders), demonstrated that auditory icons are also helpful for supporting collaboration in the office environment itself (Gaver 1991). This system distributed a variety of nonspeech audio cues to offices and common areas to keep people informed about a variety of events around their workplace.

A wide variety of sounds were used to remind people about a range of events managed by the Khronika system (Lövstrand, 1991). For instance, when new email arrived, the sound of a stack of papers falling on the floor was played. When somebody connected to a video camera in our media space, the sound of an opening door

was heard a few seconds before the connection was made, and the sound of a closing door just after the connection was broken. Ten minutes before a meeting, the sound of murmuring voices slowly increasing in number and volume was played to subscribers' offices, then the sound of a gavel. And finally, when somebody decided to call it a day, they might send their friends the "pub call," the sound of laughing, chatting voices in the background with a beer being pulled in the foreground, to suggest the evening's arrival.

Although the sounds used in EAR were somewhat frivolous, they were easily learned and recognised. In addition, they were designed to be unobtrusive, following the findings of Patterson and his colleagues (Patterson et al., 1986; Edworthy et al., 1991; Swift et al., 1989). Most were short; those that were longer had a slow attack so that they could subtly enter the auditory ambience. The sounds had relatively little high-frequency energy, and noisy or abrupt sounds were avoided. In sum, the sounds fit into the existing office ambience, rather than intruding upon it. Most telling, though the sounds were designed rather quickly and informally, they were used intensively over a long time, with hundreds of sounds being played each day over several years.

ShareMon: Monitoring Background Activities

Experience with SoundShark, ARKola, and EAR suggests that sound is useful for monitoring peripheral events without disrupting foreground activity (Gaver, 1991). Cohen (1994a) focused on this possibility in producing ShareMon, a system that allows users to monitor

network file sharing activity using a number of auditory icons. In addition, he used an iterative strategy to develop sounds that were useful, learnable, and unobtrusive.

ShareMon monitored file sharing activity on networked Macintosh computers. Auditory icons were designed to indicate a variety of the events associated with file sharing. In an initial version, knocking was used to indicate logging in, a slamming door to indicate logging out, an "ahem" sound to remind of an idle user's continued presence, and varying speeds of walking, jogging, and running to indicate CPU usage. Several alternatives for the CPU sounds were also designed: continuous humming or pink noise sounds, pulsed humming sounds, and ocean waves. The continuous sounds raised in pitch with greater CPU usage, while the pulses and waves increased in repetition frequency. Text-to-speech and visual notifications (text messages that moved across the screen autonomously, unless users "trapped" them by holding down the mouse button) were also used as comparisons to nonspeech feedback.

User feedback was obtained from people who tried the system while ostensibly working on an independent "foreground" task. File sharing events were automatically generated with high frequency during the sessions, and users were asked periodically to report on file-sharing activity. The results of this study, and more informal comments made during development, indicated mixed success for the sounds. One problem was that several of the sounds were annoying to users. This was a particular

problem for sounds indicating CPU usage, which seemed to convey little information but were played quite frequently (though the ocean wave sounds were reported to be pleasant in this context). Another problem was in mapping the sounds to the events. Most users could not initially guess what the sounds meant, though they seemed to remember their meanings after being told. Some users seemed to build elaborate narratives from the sounds (e.g., of somebody knocking on a door, not being answered, and then pacing outside) which were not intended. Others found it difficult to distinguish the system's sounds from those occurring in the space around them, and vice versa.

On the basis of these results, the sounds were changed in several ways. In the final system, CPU usage was not indicated, to avoid the problem of frequently notifying users of a relatively unimportant event (though Cohen opines that such sounds are probably useful, and that new CPU sounds might be designed). Instead, the sound of a file drawer opening and closing was used to indicate files being opened and closed. The misleading connotation of knocking indicating a request to log in, rather than an actual connection, was corrected by adding the sound of a door creaking open; in addition, registered users were distinguished by using the sounds of keys jingling and turning a lock rather than knocking. Finally, the connection reminder was changed from the annoying "ahem" sound to the sound of a creaking chair. A new round of user testing showed these sounds were more successful than the initial ones. People enjoyed the sounds more, and were better able to correctly guess what the sounds meant.

There were still problems with interpretation, however, as well as with the affective connotations of some of the sounds (most notably the use of a slamming door to indicate log off, which sounded angry to some participants).

ShareMon provides a useful example of auditory interface design, for several reasons. Perhaps most importantly, it illustrates the difficulty of designing auditory icons which succeed at a number of levels, from being unobtrusive to meaningful. The domain addressed by ShareMon was particularly challenging, because there was no complementary visual interface to aid interpretation of the sounds, the sounds played autonomously rather than in response to user activity, and they were heard quite frequently. In this context, the utility of testing such systems was well illustrated. Cohen's (1994a) strategy of testing mapping by asking people to guess the sounds' meanings may have been overly severe, and perhaps a more realistic criterion would have been whether people could remember the meanings once they are introduced. Nonetheless, the studies were not only useful in guiding the design of new sounds, but in emphasising the affective and narrative power of everyday sounds. Finally, a major success of the ShareMon system is that it was used beyond the studies by a number of "faithful adherents" in their working lives. This sort of long-term experience, and the insights it brings, is difficult to simulate in experimental settings.

Mercator: Giving Access to the Visually Impaired
Possibly the most ambitious use of auditory icons is in the Mercator system, which provides access to X

Windows applications for visually impaired users (Mynatt, 1994a, 1994b). The system addresses a problem that is clearly significant: while direct manipulation interfaces may provide significant advantages to sighted users, they are more difficult to translate into other modalities than are language-based systems, and thus represent a step backwards for the visually impaired (Boyd et al., 1990; Buxton, 1986).

Completely mapping the spatial layout of a graphical interface to the temporal medium of sound seems difficult in principle, since the ability to play and hear simultaneous sounds is far more limited than the ability to display and scan multiple graphical entities. Moreover, as Mynatt (1994a, 1994b) notes, many elements in graphical interfaces are artifacts of a limited screen size (e.g., scrollbars) and may be unnecessary or distracting in auditory versions. For this reason, the Mercator system doesn't map the visual interface *per se* to sound, but instead uses sound to convey the semantics of interaction components. It does this by translating the hierarchy of widgets used to implement an interface into a hierarchy of "auditory interface components" corresponding, for example, to buttons, menus, and windows. This is not a simple mapping, since some widgets are idiosyncratic to visual interfaces, and each component may consist of many different widgets (e.g., a menu is typically created using widgets such as lists, shells, and buttons). Thus the components presented to users of Mercator are an abstracted version of those used to actually build the corresponding graphical interfaces; the attempt is to make the mapping at the level of

units that are semantically meaningful to users.

Users navigate Mercator by using the keyboard to traverse the component tree structure, moving up, down, and sideways from superordinate to subordinate nodes. This eliminates the problem of dealing with the empty spaces between meaningful items in graphical interfaces, which are essential for spatial layout but lead to useless and frustrating "dead space" in auditory translations. Moreover, as users move from component to component, they can hear where they are, but must select the component explicitly if they want to interact with it. This allows users to "scan" the interface without actually causing any unintended actions (a lesson learned from Stevens et al., 1994).

Though speech output can be requested as needed (and is a necessary and integral part of interactions with entities such as text fields), by default Mercator uses auditory icons to provide feedback. A variety of sounds are used, such as the sound of tapping on a glass pane when a window is reached, various push button sounds for the various types of interface buttons, and a typewriter sound when a text field is reached. Auditory icons are parameterised as well, so that location in lists and menus is indicated by pitch, the complexity of containers by reverberation, etc. This strategy was suggested by Ludwig et al's (1991) work on filters, simple manipulations (such as pitch-shifting, filtering, and reverberation) that can be made on any sound to convey information.

The design of auditory icons for such a system is challenging, because sounds must be found for every interface component, and they

must work without corresponding graphical displays. User feedback was employed in finding the sounds for Mercator: people were presented interface entities and asked to choose the best sound from several alternatives, or conversely were presented a sound and asked to choose the interface entity that corresponded best. These studies were useful in guiding the sounds used in Mercator, but in the end their results depend fundamentally on the sounds offered as alternatives. For instance, early versions used short, heavily edited sampled sounds for these icons, but these were lengthened somewhat to improve recognition (Mynatt, 1994b).

Using Mercator is surprisingly simple, given the lack of visual feedback, the model of traversing a hierarchy of interface components, and the reliance on nonspeech audio cues. More formal user testing, with both sighted and visually impaired users, also suggests that the system does a good job of making X Windows applications accessible. However, it is difficult to avoid the feeling that something is lost in translating the spatial layout of a graphical interface to a hierarchy that is examined node by node. Space is an important resource for graphical displays, affording grouping, scanning, focusing, and overlooking multiple entities; these affordances are lost in Mercator's hierarchical system. Future work should focus on regaining these affordances, perhaps through the use of multiple sounds, shifting focus, and novel input techniques. To do this will probably require deeper investigations about how visually impaired people use and represent space. For now, however, Mercator represents a significant

milestone in translating visual interfaces to nonvisual form.

Auditory Icons: Discussion

Auditory icons and earcons represent different strategies for using sound to convey information from interfaces. Auditory icons use everyday sounds, earcons use musical ones. Auditory icons stress the importance of metaphorical or iconic mappings between sound-producing events and the things they are to represent, exploiting our existing skills of listening to the everyday world. Earcons use arbitrary mappings, relying on people's abilities to learn such associations. Auditory icons use the natural structure of events and their attributes to parameterise sounds, creating families of sounds that map to interface events. Earcons use invented hierarchies of sound attributes to do the same thing. Auditory icons are designed to be easy to understand, but this may require exacting sound design. Earcons must be learned, but they are easy to create and manipulate, especially using MIDI equipment. In many ways, auditory icons and earcons take diametrically opposed approaches to providing much of the same functionality.

Attempts have been made to experimentally compare auditory icons and earcons (e.g., Sikora et al., 1995; Lucas, 1994). Sets of auditory icons and earcons are designed for a selection of interface entities, and participants asked to rate how well the sounds communicate as well as how attractive they are. Auditory icons tend to be judged as mapping better (more clearly, more memorably) to interface events, but earcons tend to be judged as more pleasant.

These experiments may seem useful in examining the differences between auditory icons and earcons, but they are actually quite limited. Such studies depend on the design of the sounds meant to represent auditory icons and earcons. If a well designed set of auditory icons were compared with a poorly designed set of earcons, for instance, it is easy to suppose that auditory icons would be rated as better mapped and more likeable. The generalisability of such results, however, would be suspect at best. In addition, ratings of mapping may not correspond well to long term memorability. Similarly, preference judgements made after brief exposure in laboratory settings may not relate well to preference after long term exposure in a work setting. In general, the results of experiments meant to compare these different approaches must be taken with some scepticism.

Because of their differences, there has traditionally been a kind of implicit rivalry between auditory icons and earcons. The two approaches may turn out to complement one another, however. Auditory icons map more closely to interface entities than earcons when metaphorical or iconic mappings are used, but such mappings may be impossible to develop for some interface events. In these cases, the arbitrary but systematic mappings offered by the earcon approach may be more suitable than inappropriate or misleading everyday sounds. Similarly, parameterising auditory icons along dimensions of sound-producing events is an ideal that is sometimes only approximately implemented (e.g., changing the pitch of a sampled sound to indicate size is only an approximation of the actual acoustic correlates of size).

When this is the case, the acoustic dimensions used to parameterise auditory icons are not much different from those used to build hierarchical earcons. Similarly, recent examples of earcons have moved towards varying sounds by analogy with everyday sound-producing events (e.g., the auditory map and algebra earcons described earlier). This represents a move away from the development of arbitrary languages for sound, and towards the metaphoric and iconic mappings of auditory icons. In practice, then, the differences between these approaches may not be as great as the theories would suggest.

Systems have also developed recently which merge auditory icons with techniques used for auralisation. In Alber's (1994) Varèse system, for example, auditory icons were designed to indicate the operational states of six satellite subsystems. The proximity of each subsystem to its next operational state was mapped to the speed with which its auditory icon was played and repeated. This may be seen as an example of parameterising auditory icons, but it also draws on auralisation techniques that set thresholds between operational states rather than merely reflecting their basic parameters (see, e.g., the Kyma system described earlier). Similarly, Fitch and Kramer (1994) developed a number of auditory icons representing various physical variables (e.g., heart rate, temperature, blood pressure) of a simulated patient in an anaesthesiology task, and modified them using more abstract parameters such as pitch, timbre, and filtering. They found this hybrid technique to be very successful: Participants in their experiment

actually performed better using the sounds than with visual displays.

Despite the number of systems that have used auditory icons, the strategy has only begun to be developed. One issue that deserves more emphasis is the design of unobtrusive sounds than still convey adequate information. The work by Patterson et al., (1986), Edworthy et al. (1991), and Swift et al. (1989) on perceived urgency and annoyance could be a useful guide for this. Nonetheless, the design of truly subtle auditory icons—sounds as subtle as that made by moving a sheet of paper across a desk—will require careful attention to design. Such attention can be found in the design of some of the auditory icons beginning to appear in commercial systems: the whoosh as a window is compressed, for instance, or the click of a software button. But these sounds seem designed in an ad hoc fashion, without taking advantage of parameterisation or the creation of ecologies of sound. We may be beginning to see the emergence of auditory icons as a standard interaction technique, but there are still many possibilities that have yet to be explored or developed in commercial products.

NEW DIRECTIONS IN THE DESIGN SPACE

Though it is a cliché to say that auditory interface design is a new field, it should be clear at this point that this is not really the case. The earliest work I describe here, by Bly (1982a, 1982b), is already 15 years old, an extraordinarily long time in interface-years. The precursors to this research are even more ancient: With anecdotal reports of computer scientists tuning radios to their computers or wiring speakers to shift registers, it is likely that people

have been trying to listen to computers from their very inception.

Nonetheless, it is fair to say that research on auditory interfaces is not mature. A number of disparate strategies for creating auditory interfaces have been explored over the last fifteen years or so, but they can be differentiated along relatively few dimensions. First, the kinds of sounds employed vary at a number of levels. Second, the kinds of mappings created between sounds and the meanings they are to convey also vary, along a dimension from completely arbitrary mappings to very literal, iconic ones. Finally, the kinds of functionality auditory interfaces are to provide vary as well, from allowing people to perceive patterns in data, to providing information about events and processes that are useful for social coordination.

Though research has ranged over a broad space of sounds, mappings, and functions, this space has been relatively sparsely populated by working systems and rigorous research. Moreover, the area is still relatively restricted. In part, this is because the dimensions of sound, mapping, and functionality are not really independent: a choice on one dimension constrains the possible choices along the others. In this section, then, I discuss the space of interfaces that has developed, and suggest several ways that it might be better filled by new research. I end by discussing the need to expand the field, to help research on auditory interfaces merge with new developments in interaction design.

Sound: The Sonic Palette
The most obvious distinguishing characteristics of auditory interfaces is in the auditory vocabulary they

employ. A primary distinction can be made between those systems that use musical sounds and those which use everyday sounds. Within these categories, however, the sonic structures used vary in complexity, and simpler systems have typically been explored more thoroughly than more complex ones. The choice of a framework for understanding sounds, and the level of complexity explored within that framework, together have strong implications for the sorts of mappings that can be achieved between sounds and their meanings, and ultimately for the kinds of functionality that can be addressed.

Musical sound is a very broad category, encompassing everything from the discrete tones used by Bly (1982a, 1982b), the simple motives used by Patterson (Patterson et al., 1986; Patterson, 1982), Blattner et al. (1989) and their colleagues, to more complex, expressive music. Each of these subgroups offers its own parameters for control or encoding, from the basic psychoacoustic parameters used to describe simple tones, to the simple musical attributes used in motives, to the higher level musical structures that might allow even more complex information to be integrated. In addition, there is the potential for controlling—and the impossibility of avoiding—the expressive dimensions that characterise music, and which are used, for instance, in the design of films, games, or multimedia products. The sorts of dimensions we might use to describe musical sounds at this level—mood, atmosphere, texture—are not well understood from a scientific point of view, but could be a powerful addition to auditory interface design.

Similarly, everyday sounds range from simple sound effects

used without variation, to parameterised auditory icons, to ecologies of everyday sounds meant to be played together. Opportunities for control here range from simply choosing a representative sound from the infinite variety of possibilities, to determining the source-specific attributes that might be used to parameterise auditory icons, to controlling aspects of a virtual auditory environment such as localisation and room acoustics. Beyond this, entire auditory environments might be defined, similar to that demonstrated by the ARKola system (Gaver et al., 1991), in which numerous everyday sounds combine to form an integrated whole. This sort of potential, which mirrors the possibility of using more sophisticated musical structures, is not very well understood, though psychoacoustical factors concerning masking, for instance, are clearly important, as well as higher-level ones concerning congruency, place, and narrative.

In sum, there are many ways to use musical and everyday sounds, and this is one of the primary dimensions of the space of design for auditory interfaces. Not surprisingly, most systems have tended to use relatively simple sounds—whether musical or everyday—rather than the more complex structures that are possible. One way in which auditory interface design may be expected to expand, then, is towards more complex vocabularies of sound, whether musical or everyday. This will allow more complex mappings to be made from sound to meaning, and permit new functionality for auditory interfaces as well.

Mapping Sound to Meaning
The mappings between sounds and information is the second dimension along which auditory interfaces may be distinguished. Mappings vary in how systematic or determined they are. On one extreme, many mappings are completely arbitrary, as when pitch is used to stand for some data dimension which has nothing to do with pitch, or frequency. On the other hand, they may be very literal, or iconic, as when the sound of crumpling paper is used to indicate that a text file has been deleted in an appropriate model world (see Gaver, 1989, 1986).

Arbitrary mappings are used in many auditory interfaces. Most auralisations set up arbitrary mappings between data and sound, just as most graphing techniques arbitrarily assign data to vertical and horizontal dimensions, symbols, colour, and so forth. This works reasonably well for auralisation. In many cases, the mappings are specified by their users, and they are almost always introduced shortly before the auralisation is actually played. Arbitrary mappings are more problematic for alarms, especially in environments such as aircraft cockpits or intensive care units where many alarms are used. A similar problem faces systems that use motives: the more mappings required in a particular interface, the more careful design is needed to make them clear and unambiguous. Earcons often use arbitrary mappings, (in fact, the ease of establishing such mappings is one of the appeals of the strategy). For this reason, rules for generating families of motives are important because basic motives may be arbitrary, but their variations are rule-based. Finally, even everyday sounds may

be arbitrarily mapped to their referents (e.g., Brown et al., 1989), though the results do not really fit the definition of auditory icons. In general, arbitrary mappings are seductive to designers because they are easy to establish. They are problematic for users, however, because they are often difficult to learn and remember.

The clearest mappings are those that are iconic, that is, those formed when the relation between a sound and event is causal. This is possible to achieve in computer interfaces when a deep metaphor is used to establish a model world, and the sounds created by particular entities in that world are those that would be created by their physical counterparts (Gaver, 1989). Iconic mappings, along with the use of everyday sounds, are the defining features of auditory icons. Such mappings have seldom been used in other sorts of auditory interfaces, though they have appeared occasionally in auralisations. Iconic mappings are obviously desirable when they can be found. Even if they are not immediately guessable, they are often learned and remembered after a single explanation. However, sounds that map iconically to a given event or message may be difficult to find, not least because not all events make sound. In addition, the psychoacoustical requirements for sounds that are clearly meaningful may conflict with those that make sounds unobtrusive.

Taken together, these properties gives iconic mappings contrasting strengths compared to arbitrary mappings. On the one hand, iconic mappings may not be very useful for auralisations, because their memorability is relatively unimportant for temporary

mappings, and they might constrain designers or confuse listeners. For instance, if voices were used to represent population variables, how would occupation be represented? Would it matter what the voices were saying? On the other hand, iconic mappings may be more useful for alarms, because memorability becomes an important issue. Many different alarms may need to be interpreted, and yet they are sounded fairly rarely (one hopes). However, the sheer naturalness of iconic mappings to everyday sounds may undercut alarms' functions by failing to convey the essential fact that the alarm sound is an intentional message with a human author. Finally, iconic mappings, in the form of auditory icons, have proved to be very useful for more general interfaces, not only because they tend to be easily learned and remembered, but also because they tend to integrate well with graphical interfaces and the overall context of use.

Between arbitrary and iconic mappings are a large variety of mappings that are more systematic than purely arbitrary mappings, yet less literal than iconic ones. These are commonly glossed as "metaphorical" mappings, though this single term tends to mask the various kinds of analogy that are used. For instance, the use of high pitches to stand for something near the top of the screen, and low pitches for things near the bottom, relies on a cross-modal sensory correspondence which, though powerful, is little understood (see Marks, 1978). The use of walking, jogging, and running sounds to stand for CPU usage in ShareMon (Cohen, 1994a), on the other hand, relies on a higher-level semantic analogy between speed and effort

of walking and the processing load. Finally, the use of a whooshing sound to indicate opening and closing windows in the SonicFinder (1989) is a sound effect, a sound that is iconically mapped to a nonexistent event.

Metaphorical mappings are both flexible and powerful. The traditional conflict between difficult-to-learn arbitrary mappings and difficult-to-develop iconic ones increasingly seems to imply an unreasonable choice between two extremes. Suitable metaphors may be very difficult to find or develop, and poorly designed ones may be misleading. Nonetheless, metaphors share the best features of arbitrary and iconic mappings, being more flexible than literal mappings, yet more easily learned and remembered than arbitrary ones. Because of this, it seems likely (and desirable) that the use of metaphorical sounds will increase, ideally accompanied by better analyses of how they work, and by heuristics for their design. This development, in turn, should allow new functionality to be addressed, as the tools for mapping sound to meaning become simultaneously more powerful and more intuitive.

Functions for Auditory Interfaces

A final perspective for distinguishing auditory interfaces focuses on the functions they are meant to perform. As the systems I have discussed suggest, there are many uses for auditory interfaces: from alarms indicating some event, to auralisations allowing data patterns to be discerned, to motives that summarise equations, to auditory icons that support collaboration. The range of uses to which sounds may be put do not themselves

define a dimension, of course, but there are dimensions along which they may be considered.

One way to think about the functions that sounds perform is in terms of the complexity of information that is conveyed. This is not the same thing as the complexity of the sounds themselves. Even complex sounds often signal only a single bit of information; namely, that some event has occurred. For example, most interrupt beeps—whether simple tones or complex samples—indicate only that a condition has been reached in which somebody decided that an interrupt beep should be played. A step up from this are sounds that basically serve as labels, such as Patterson's (1982) alarms, simple versions of Blattner et al.'s (1989) earcons, or unparameterised auditory icons. Finally, some sounds convey rich systems of information, such as multidimensional auralisations, hierarchical earcons, or parameterised auditory icons. In general, the amount of information conveyed by a given sound is a trade-off between simplicity and ease of learning on the one hand, and efficiency on the other. Simple systems are easy to learn, while multi-layered systems can provide large amounts of information. Strategies such as parameterisation of everyday sounds seek to gain the benefits of very efficient, rich messages, while making them as easy to learn as simpler cues.

Another perspective to take on the functions served by auditory interfaces concerns their use over time. How an interface is to function over time affects how well various mappings will work. At one extreme, many auralisations are extremely ephemeral, with a given mapping between sound and data

established only for a few exploratory trials. At the other extreme, genre sounds such as telephone rings, ambulance sirens, and cuckoo clocks have such a long history that the mappings between sound and meaning they use seem almost necessary (Cohen, 1993). In between the extremes of fleeting use and historical stability, most alarms, earcons, and auditory icons are designed for relatively long use, without having the benefits of a preexisting history. It is for this reason that metaphorical or iconic mappings are valuable, in allowing the benefits of a related history to be applied to new sounds.

The functions performed by auditory interfaces can also be considered in terms of the prospective audience for the sounds. Data auralisations again represent an extreme, this time one in which the designer and audience are often the same, at least until some interesting pattern is determined. Even when auralisations are played for third parties, they are buttressed by their designer's explanations. On the other hand, most other auditory interfaces are at least nominally designed for use by an audience who has never met the designer. Some of these interfaces, such as most alarms, must be designed for a non-specialised audience with little desire to learn new mappings, and thus must be immediately distinctive, communicative, and motivating. Others, such as those using earcons and auditory icons, may be designed initially for audiences more likely to learn novel mappings. It is dangerous to overestimate this kind of tolerance, however, and again techniques for improving mappings are probably appropriate for any general audience. Finally, at the

opposite extreme from auralisations are systems meant to support collaboration. To mediate social interaction flexibly and subtly, such systems should ideally use auditory mappings as rich as those used for earcons and auditory icons, but their meaning should not rely on learning or the presence and authority of the designer. In fact, the meaning of given sounds may change as their role in social interaction is negotiated by their users.

Expanding the Space

There are many possibilities for new work on auditory interfaces. Here it is useful to consider the space in reverse order from the previous discussion. After all, the extension of existing strategies to new domains has characterised the field, and we can expect auditory interfaces to follow wherever digital technologies may go (and in some cases to lead the way). In doing this, new kinds of mappings become possible and desirable. These mappings, in turn, lead to new possibilities for thinking about and designing sounds. So while sounds may be the most obvious attributes of existing interfaces, it is their functionality that drives development. As new functionality is explored, the boundaries of existing research—the distinctions among auralisations, musical messages, and auditory icons—become blurred. In the end, new dimensions may appear as well, as aesthetic and design sensibilities are included in work on auditory interfaces. This is a move that could dramatically expand the space of auditory interfaces.

For instance, there has been a great deal of research on the spatialisation of sound, but little work on the systematic design of

sounds for virtual reality systems. As I described earlier, basic research on localisation has led to powerful enabling technologies and techniques for creating virtual auditory realities. This has been used to spatialise overlapping streams of speech, improving comprehension in aircraft cockpits (Begault, 1994). In addition, Wenzel and her colleagues (Wenzel et al., 1988) explored a system of nonspeech cues that helped replace tactile cues in a virtual reality manipulation task. Finally, many virtual reality systems have included sound, sometimes to striking effect (e.g., Laurel et al., 1994). Nonetheless, more systematic work addressing the new functions sound could perform in these environments would be helpful. This is a major opportunity for new research.

Similarly, auditory interfaces have not really kept up with the emergence of portable hand-held devices. Stifelman (e.g., Stifelman et al., 1993) has pursued innovative research on auditory-only PDAs, but though her designs have included auditory icons, her systems rely mainly on speech recognition and output. An alternative would be to focus on using nonspeech audio to convey information about events and colleagues, allowing such a device to be used as a portable awareness server (see Hindus et al., 1995). More generally, collaborative systems offer a rich domain for exploring auditory interfaces, and may be especially suited to the use of sound to maintain group awareness while allowing individual activity (Gaver, 1991). Finally, the explosion of the World Wide Web offers scope for interface design of all sorts. Albers and Bergman (1995) designed a system that produced a variety of auditory icons

to provide feedback about the state of a web browser. This is a start in the right direction, but more work could be done on the use of sound to orient oneself in the strangely homogenous space offered by the web.

Approaching new functionality will encourage explorations of new forms of mapping. An integral part of expanding the range of mappings that can be used—and, by implication, the range of acceptable sounds and approachable functionality—is the provision of tailorable and interactive interfaces to users. This is not a trivial problem: Customisation must be constrained to ensure that appropriate sonic attributes and dimensions are used, to guard against psychoacoustical problems like masking and interactions, and (not least) to avoid aesthetically horrifying results. Nonetheless, interactivity may enable greatly expanded uses of sound. It is one of the key elements of systems like *Kyma* (Scalletti, 1994) that allow users to quickly set up their own data auralisations. Tailorability is also likely to help systems using earcons and auditory icons avoid confusions in mappings, and moreover would allow users to tune auditory interfaces to fit their working environments.

Finally, work in new domains, and the use of more flexible mappings, has implications for the sounds that are used as well. There is a vast array of sounds, for instance, that defy easy categorisation as either musical or everyday. Some are familiar, such as the whistling of wind, the drone of an engine, or the melody of running water. Others may be constructed, either by synthesis or manipulated recordings (e.g., Eno, 1982). The potential of such sounds

is to combine the iconic possibilities of everyday sounds with the ease of manipulation of musical ones; an added value is greater freedom in aesthetically controlling the results.

Overall, there are many ways to extend the possibilities of auditory interfaces; to use a greater and more structured lexicon of sounds, to apply sounds to new functional domains, and to explore new and more responsive mappings among sounds and information. There are also rich opportunities for research meant to consolidate existing strategies. Each of the three major strategies for using sounds—auralisation, musical messages, and auditory icons—need more and better examples of real world applications. For each, the appearance of projects focused around their applications to real problems, which report the difficulties, failures and successes of the process, and the benefits of the resulting system, would be extremely valuable. From this perspective, the best way to improve the space of auditory interfaces would be to help colonise it.

Blurring the Boundaries

As auditory interfaces have been used in new domains, the boundaries that have traditionally defined the research are beginning to blur, so that the design space is beginning to be more evenly filled. Several examples help illustrate how newer systems have started to merge strategies for creating auditory interfaces.

OutToLunch

An excellent example of new developments in sound design comes from a system called *OutToLunch* (Cohen, 1994b). *OutToLunch* was designed to

recreate the aural experience of working in a shared space, in which sounds naturally provide awareness of group activity. The system counted each user's keystrokes, mouse clicks, and total mouse movement, and then represented this information to users using recordings of real keystrokes, mouse clicks, and mouse rolling noise. Unfortunately, the original prototype proved too annoying for continuous daily use. Cohen (1994b) speculates that this was because the sounds conveyed relatively little information (no information about who was active, or where), and that they were not aesthetically well designed (the sounds were too loud and had low resolution). Thus in the second iteration, *OutToLunch* was redesigned to simply indicate whether or not each member's machine was being used. This gave people a sense of who was around and what they were doing, without distracting them with detailed mechanical activity.

What moved the new version onto new ground, however, was the sound design it employed. The new version used sounds designed by Michael Brook, a professional musician and producer (e.g., Brook, 1992). Brook started by creating a continuous drone by looping a low-pitched, quiet guitar sound that faded in and out when activity started or stopped and played continuously during any group activity. Then he fashioned a set of concordant musical themes from short guitar or synthesiser melodies which had slow attacks, long decays, and a great deal of reverberation and echo, allowing them to merge with the underlying themes. Group members chose their own themes to represent them in the resulting display. In the final

system, the drone would fade up as the group became active, and each theme would be added to the mix at random offsets within a thirty second period: The overall effect was of a very atmospheric, non-repeating musical texture much like the sort of "ambient" music pioneered by Brian Eno (1975).

The iterated *OutToLunch* was not used very long by the group (which disbanded—for unrelated reasons—soon after it was deployed). Nonetheless, the new sounds represent a qualitative step forward in sound design for auditory interfaces. This reflects the contribution of a professional musician to the crafting of a set of expressive, aesthetically focused sounds. In part, this meant that Cohen and Brook were unconstrained by the strategies for using sounds discussed earlier. The focus on peripheral monitoring is typical of work on auditory icons (e.g., Gaver, 1991), but the musical themes are suggestive of the earcon approach, and the use of sounds to represent essentially numerical data is strongly reminiscent of auralisation work. *OutToLunch* mixed aspects of auralisations, musical messages, and auditory icons in an eclectic way, drawing on each to support unobtrusive group awareness. It is a cliché in auditory interface design to recommend that a sound designer be used in creating auditory interfaces, but *OutToLunch* represents one of the few, and best, examples of how this should be done.

Audio Debugging

Recently several systems have been developed to help programmers debug sequential and parallel programs. These systems resemble auditory icons and earcons in that

their overall purpose is to allow programmers to monitor events in the computer, but they employ strategies more reminiscent of auralisation in producing their sounds.

For instant, Sonnet (Jameson, 1994) is a visual programming interface that allows sounds and sound modifications to be “attached” to lines of code in a sequential program. Typically sounds are triggered after some line of code, and stopped sometime later. In addition, parameters of the sounds may be modified to indicate, for instance, the number of times the sound has been triggered, or the value of some variable. These sounds and modifications are then played in realtime as the program is run. Jameson (1994) describes several examples of how Sonnet might be used. For instance, a note might be triggered just before, and stopped just after, an iterative loop is declared. If the note doesn’t stop when the program is run, then the loop is not exiting, either because it is endless or because some subroutine is defective. More information can be obtained by modifying the volume of the note (usually in a triangle form, so over time a tremolo effect is produced) just before the first statement within the loop is reached. If the resulting sound doesn’t change in volume, then there is probably a bad subroutine; if the volume changes but doesn’t end, then the exit condition is probably wrong. Similarly, note parameters can be used to track the values of variables, or to indicate access to variables. Simply by “labelling” various lines of code with sounds, the details of code execution may be listened to, and problems identified by unexpected sounds.

A similar strategy has been developed by Jackson and Francioni (1994) for debugging parallel program behaviour. Like Jameson (1994), their system starts and stops notes to indicate various events within a program. But it is much more difficult to debug and optimise parallel programs than sequential ones, because overall system behaviour depends crucially on the interactions among a number of processors as well as the sequential behaviour of each of them. Thus Jackson and Francioni (1994) focus on using sound to track the communications and timing among processors. For instance, each processor may be assigned a different pitch, and notes triggered to indicate that it has sent or received a message. Sustained notes may be used to indicate when processors are idle or busy, or started when a message is sent and stopped when it is received. Finally, notes may be used to indicate specific events within each processor.

Using these sorts of mappings, patterns of behaviour can be heard emerging from a group of processors running in parallel. Jackson and Francioni (1994) suggest that the sounds are useful in helping to focus attention on relevant aspects of visual representations, or on details of timing that are difficult to perceive from visual representations. For instance, lost messages—those that are sent but not received—can be heard if a note triggered by a send is not stopped by a receive. The relative activity of processors can be heard by assigning different pitches to each, or by playing sounds while processors are idle. Finally, the sounds that result are often strikingly musical in nature, creating

a variety of auditory textures, melodies, or harmonies.

Sound is an appealing medium for debugging, as these systems have shown. Perhaps most importantly, sounds can be played as a program is running, without interfering with it. There is no need to step through a program, nor to set breakpoints at which variables can be examined. Of course, in many cases the same kinds of information could be printed out as text, but sounds have the same sort of advantage over text that graphs do: by integrating data, they allow the perception of patterns that would be more difficult to apprehend using lists or tables—particularly information about timing, as Jackson and Francioni (1994) point out. Sound allows processes to be monitored while visual attention is directed elsewhere, for instance at screen behaviour. Finally, triggering or modifying sounds for debugging is relatively lightweight. Many systems offer at least basic functionality for playing and modifying sounds, and MIDI equipment can be used to play a greater range of sounds at much less cost to the processor.

The tactic used by both Jackson and Francioni (1994) and Jameson (1994) of playing sounds that depend on events in the computer is similar to the strategies used in creating auditory icons and earcons, and the resulting ability to monitor several simultaneous dynamic processes is also similar. But there are striking similarities to auralisation work as well. Because these debugging tools are meant to be used by programmers themselves, rather than distributed to a large audience, relatively little importance is placed on making the sounds map to events in intuitive or

memorable ways. Instead, the emphasis is on allowing programmers to set up highly discriminable and flexible sound mappings to be played soon afterwards. In addition, both debugging systems tend to rely on triggering simple notes which are then modified to indicate various parameters (such as processor number or the value of a variable), just as many auralisations trigger sounds for each new data point, then vary them to indicate values along several dimensions. In the end, this work is not only valuable in its own right, but also because it emphasises the potential of merging the various techniques and strategies that have already been developed.

Throwing Open the Gates: Multimedia and Games

While traditional interfaces have yet to incorporate sound as a standard element of interaction, multimedia and games systems routinely employ very sophisticated music and sounds. Unfortunately, there has been little conversation between the research community and the ever-growing community of content developers. At the least, a survey of how sounds are used in a wide variety of multimedia products would be a valuable addition to the literature. It is clearly time for research on auditory interfaces to look to the established practices of these fields for inspiration and new ideas. Conversely, it should be realised that research on auditory interfaces has a great deal to offer games and multimedia work in return.

What multimedia and games have to offer auditory interface designers is a routine adherence to standards of sound design, the

aesthetic crafting of sounds, that have seldom been met in auditory interface research (barring Cohen and Brook's OutToLunch system). It is difficult to overstate the importance of good sound design to auditory interfaces. Bad sound design can make information-rich interfaces seem distracting, irritating, and trivial. Good sound design can reinforce a sound's message, allow it to fit into its environment of use, help avoid auditory fatigue, and allow and interface to be accepted and enjoyed. In general, research on auditory interfaces could learn a great deal about effective sound design from multimedia and games work.

Sound's expressive capabilities are often more important to games and multimedia designers than the drier goal of systematically transmitting information. This reflects the traditions of sound design for radio, television and movies. In such fields, sound's ability to provide information is almost tangential to its ability to produce tension, convey mood, and communicate and evoke emotion. As with these fields, so with multimedia and games: Sounds are used to engage users, to heighten the excitement and pace of games, to create an atmosphere appropriate to a given topic in multimedia treatments.

The aesthetic and expressive dimensions of sound have largely been overlooked in auditory interface research. Occasionally an auralisation will make appropriate emotional mappings between sound and data (e.g., Scalletti 1994 mapped pollution levels to coughing sounds). Work on urgency (Patterson et al., 1986; Edworthy et al., 1991) can be seen as an

example of trying to study a simple parameter of mood using psychoacoustical methodologies. Finally, work on auditory icons has been concerned with crafting sounds for their environments, and with sound's ability to engage users with systems (see, e.g., Gaver et al., 1991). Nonetheless, crafting sounds for their aesthetic qualities, and designing them to create mood and emotion, has seldom been anything but incidental to more functional goals. A greater concern for these issues could lead to auditory interfaces that better reflect the importance and implications of various events, not just their dry informational values.

On the other hand, there is much that auditory interface research can offer multimedia and games designers. The ability to systematically map sounds and their dimensions to information, to create families of related musical messages, and to vary everyday sounds along source dimensions could complement the sophisticated aesthetic and expressive sound design already employed, making multimedia and games more engaging, easier to use, and more challenging as well. As it is, it is not clear that most examples use sound to convey important information. Though Buxton (1989) speculated that turning off the sound on a video game would lead to lower scores, I am sceptical that this is generally true (the relevant data has never been collected). If multimedia and games designers drew more on auditory interface work, we would be much more likely to see sound's influence reflected in the scores.

Conclusion

There are three major lines of endeavour that should help the field

truly become mature. First, we need to generate more examples of systems that fulfil real user needs, whether these be to analyse complex data, comprehend complex messages, orient to real and virtual events, or simply to have fun. Second, we need to understand and explore the richer structural possibilities of sound, whether these be musical, metaphorical, or environmental. Finally, we need to focus our attention on sounds that are aesthetically controlled, as subtle and beautiful as those we hear in the orchestra hall, or on a walk through the woods.

ACKNOWLEDGEMENTS

I am very grateful to Beth Mynatt, Anne Schlottmann, Jayne Roderick, and Sara Bly for their comments on earlier drafts of this chapter.

REFERENCES

- Albers, M. C. (1994). The Varese system, hybrid auditory interfaces, and satellite-ground control: Using auditory icons and sonification in a complex, supervisory control system. In G. Kramer and S. Smith (eds.), *Proceedings of the Second International Conference on Auditory Display* (Santa Fe, N.M., 7-9 November, 1994), 3 - 14.
- Albers, M. C., & Bergman, E. (1995). The audible web: Auditory enhancements for Mosaic. CHI'95 conference companion, ACM Conference on Human Factors in Computing Systems, 318-319.
- Ballas, J. A. (1994a). Common factors in the identification of an assortment of brief everyday sound. *Journal of Experimental Psychology: Human Perception and Performance*. 19: 250 - 267.

Ballas, J. A. (1994b). Delivery of information through sound. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 79 - 94.

Ballas, J. A., & Mullin, T. (1991). Effects of context on the identification of everyday sounds. *Human Performance*. 4: 199 - 219.

Begault, D. R. (1994). 3-D sound for virtual reality. New York: Academic Press.

Blattner, M. M., Papp, A. L. III, & Glinert, E. P. (1994). Sonic enhancement of two-dimensional graphic displays. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 447 - 470.

Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction* 4(1), Spring 1989.

Bly, S. (1982a). Sound and computer information presentation. Unpublished doctoral thesis (UCRL-53282), Lawrence Livermore National Laboratory and University of California, Davis, CA.

Bly, S. (1982b). Presenting information in sound, *Proceedings of the CHI '82 Conference on Human Factors in Computer Systems*, 371-375. New York: ACM.

Borin, G., De Poli, G., & Sarti, A. (1993). Algorithms and structures for synthesis using physical models. *Computer Music Journal*, 16(4), 30 - 42.

Boyd, L. H., Boyd, W. L., & Vanderheiden, G. C. (1990). The graphical user interface: Crisis,

- danger, and opportunity. *Journal Visual Impair. & Blind.* 496 - 502.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA.: MIT Press.
- Brewster, S.A., Wright, P. C, & Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. In Kramer, G. (ed.), *Auditory Display.* New York: Addison-Wesley, 471 - 498.
- Brook, M. (1992). *Cobalt blue.* CAD 2007 CD, 4AD.
- Brown, M., Newsome, S. & Glinert, E. (1989). An experiment into the use of auditory cues to reduce visual workload. *Proceedings of the CHI '89.* New York: ACM, 339-346.
- Burgess, D. A. (1992). Real-time audio spatialization with inexpensive hardware. (Report No. GIT-GVU-92-20). Graphics Visualization and Usability Center, Georgia Institute of Technology.
- Buxton, W. (1986). Human interface design and the handicapped user. *Proceedings of CHI'86,* 291 - 297.
- Buxton, W. (1989). Introduction to this special issue on non-speech audio. *Human-Computer Interaction* 4(1), Spring 1989.
- Cabot, R.C., Mino, M.G., Dorans, D.A., Tackel, I.S. & Breed, H.E. (1976). Detection of phase shifts in harmonically related tones. *Journal of the Audio Engineering Society,* 24, 568-571.
- Cohen, J. (1993). "Kirk here:" Using genre sounds to monitor background activity. *INTERCHI'93 Adjunct Proceedings (Amsterdam, April 24 - 29, 1993),* 63 - 64.
- Cohen, J. (1994a). Monitoring background activities. In Kramer, G. (ed.), *Auditory Display.* New York: Addison-Wesley, 499 - 531.
- Cohen, J. (1994b). Out to lunch: Further adventures monitoring background activity. In G. Kramer and S. Smith (eds.), *Proceedings of the Second International Conference on Auditory Display (Santa Fe, N.M., 7-9 November, 1994),* 15 - 20.
- Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors.* 33(2), 205-231.
- Eno, B. (1975). *Discreet music.* EEGCD 23, E.G. Records Ltd.
- Eno, B. (1982). *Ambient 4 / On land.* EEGCD 20, E.G. Records Ltd.
- Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory display over a visual display in a complex, multivariate system. In Kramer, G. (ed.), *Auditory Display.* New York: Addison-Wesley, 307 - 326.
- Fletcher, H.F. & Munson, W.A. (1933). Loudness, its definition measurement and calculation. *Journal of the Acoustic Society of America,* 5, 82-108.
- Freed, D. J. (1988). Perceptual control over timbre in musical applications using psychophysical functions. Unpublished Masters Thesis, Northwestern University.
- Freed, D. J., & Martins, W. L. (1986). Deriving psychophysical relations for timbre. *Proceedings of the International Computer Music Conference, Oct. 20-24, 1986, The Hague, The Netherlands,* 393-405.
- Gaver, W. W. & Smith, R. (1990). Auditory icons in large-scale collaborative environments. In D. Diaper et al. (Eds.), *Human-Computer Interaction - INTERACT '90,* Elsevier Science Publishers B.V. (North-Holland), 735-740.
- Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction.* 2, 167-177.
- Gaver, W. W. (1988). *Everyday listening and auditory icons.* Doctoral Dissertation, University of California, San Diego.
- Gaver, W. W. (1991). Sound support for collaboration. *Proceedings of ECSCW'91 (Amsterdam, September 24 - 27, 1991).* Kluwer, Dordrecht. Reprinted in Baecker, R. (ed.), *Readings in groupware and CSCW: Assisting human-human collaboration.* Morgan Kaufmann, San Mateo, CA, 1993.
- Gaver, W. W. (1993a). How do we hear in the world? Explorations of ecological acoustics. *Ecological Psychology,* 5(4): 285 - 313.
- Gaver, W. W. (1993b). What in the world do we hear? An ecological approach to auditory source perception. *Ecological Psychology,* 5 (1): 1-29.
- Gaver, W. W., & Mandler, G. (1987). *Play it again, Sam: On liking music.* Cognition and Emotion, 1 (3) 259-282.
- Gaver, W. W., (1989). The SonicFinder, a prototype interface that uses auditory icons. *Human Computer Interaction,* 4, 67 - 94.
- Gaver, W. W., Smith, R. & O'Shea, T. (1991). Effective sounds in complex systems: the ARKola simulation. *Proceedings of CHI '91, ACM Conference on Human Factors in Software,* 85-90.
- Gibson, J. J. (1979). *The ecological approach to visual perception.* New York: Houghton Mifflin. (Also published by Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986).
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America.* 61, 1270-1277.
- Grinstein, G. & Smith, S. (1990). The perceptualization of scientific data. In E. Farrell (Ed.). *Extracting meaning from complex data: processing, display, interaction.* *Proceedings of the SPIE, Vol. 1259,* 190-199.
- Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: The MIT Press.
- Helmholtz, H. L. F. (1885/1954). *On the sensations of tone as a physiological basis for the theory of music.* New York: Dover.
- Hindus, D., Arons, B., Stifelman, L., Gaver, W., Mynatt, E., and Beck, M. (1995). Designing auditory interactions for PDAs. *Proceedings*

of the ACM Symposium on User Interface Software and Technology, 143-146

Holmes, N. (1993). The Best in Diagrammatic Graphics.

IMA (1983). MIDI musical instrument digital interface specification 1.0. North Hollywood, CA: IMA. (Available from IMA, 11857 Hartsook Street, North Hollywood, CA, 91607, USA.)

Jackson, J. A., & Francioni, J. M. (1994). Synchronization of visual and aural parallel performance data. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 291 - 306.

Jameson, D. H. (1994). Sonnet: Audio-enhanced monitoring and debugging. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 253 - 266.

Laurel, B., Strickland, R., & Tow, R. (1994). Placeholder: Landscape and narrative in virtual environments. *ACM Computer Graphics Quarterly*, 28(2).

Lövstrand, L. (1991). Being selectively aware with the Khronika system. Proceedings of ECSCW'91 (Amsterdam, September 24 - 27, 1991). Kluwer, Dordrecht. Reprinted in Baecker, R. (ed.), *Readings in groupware and CSCW: Assisting human-human collaboration*. Morgan Kaufmann, San Mateo, CA, 1993.

Loy, G. (1985). Musicians make a standard: The MIDI phenomenon. *Computer Music Journal*, 9(4), 8-26.

Lucas, P. (1994). An evaluation of the communicative ability of auditory

icons and earcons. Proceedings of the Second International Conference on Auditory Display (Santa Fe, N.M., 7-9 November, 1994).

Paul A. Lucas

Ludwig, L. F., & Cohen, M. (1991). Multi-dimensional audio window management. *International Journal of Man-Machine Studies*, 34(3): 319 - 336.

Lunney, D. & Morrison, R.C. (1990). Auditory presentation of experimental data. In E. Farrell (Ed.). *Extracting meaning from complex data: processing, display, interaction*. Proceedings of the SPIE, Vol 1259, 140-146.

Lunney, D., & Morrison, R. C. (1981). High technology laboratory aids for visually Handicapped chemistry Students. *Journal of Chemical Education*, 58, 3, 228-231.

Lunney, D., Morrison, R.C., Cetera, M.M., Hartness, R.V., Mills, R.T., Salt, A.D. & Sowell, D.C. (1983). A microcomputer-based laboratory aid for visually impaired students. *IEEE Micro*, 3(4), 19-31.

Marks, L. E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.

McAdams, S. & Bregman, A. (1979). Hearing musical streams. *Computer Music Journal*, 3 (4), 26-43, 60, 63. Also appear in Roads, C. & Strawn, J. (1985). *Foundations of Computer Music*. Cambridge MA: MIT Press, 658-698.

McIntyre, M. E. , Schumacher, R. T., & Woodhouse, J. (1983). On the oscillations of instruments. *Journal of the Acoustical Society of America*, 74 S52.

Mezrich, J. J., Frysinger, S., & Slivjanovski, R. (1984). Dynamic representation of multivariate time series data. *Journal of the American Statistical Association*. 79, 34-40.

Mynatt, E. D. (1994a). Auditory presentation of graphical user interfaces. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 533 - 555.

Mynatt, E. D. (1994b). Designing with auditory icons. In G. Kramer and S. Smith (eds.), *Proceedings of the Second International Conference on Auditory Display (Santa Fe, N.M., 7-9 November, 1994)*, 21 - 30.

Patterson, R. D., Edworthy, J., Shailer, M. J., Lower, M. C., & Wheeler, P. D. (1986). Alarm sounds for medical equipment in intensive care areas and operating theatres. Report AC598. University of Southampton Auditory Communication & Hearing Unit.

Patterson, R.D. (1982). Guidelines for auditory warning systems on civil aircraft. CAA Paper 82017. London: Civil Aviation Authority.

Pierce, John R. (1983). *The science of musical sounds*. New York: W. H. Freeman and Company.

Risset, J. C., & Wessel, D. (1982). Exploring timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press.

Sanders, M. S. & McCormick, E. J. (1987). *Human factors in engineering and design (6th ed.)*. New York: McGraw-Hill.

Scaletti, C. & Craig, A. (1991). Using sound to extract meaning from complex data. Proceedings of the 1991 SPIE/SPSE Symposium on Electronic Imaging Science and Technology.

Scalietti, C. (1994). Sound synthesis algorithms for auditory data representations. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 79 - 94.

Scharf, B., & Houtsma, A. J. M. (1986). Audition II: Loudness, pitch, localization, aural distortion, p pathology, In K. R. Boff, L., Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley, Vol. I, 15.1-15.60.

Shepard, R.N. (1964). Circularity in judgements of relative pitch. *Journal of the Acoustical Society of America*, 36, 2346-2353.

Sikora, C., Roberts, L., & Murray, L. T. (1995). Musical vs. real world feedback signals. CHI'95 Conference Companion, New York: ACM, 220 - 221.

Smith, S., Bergeron, R.D. & Grinstein, G.G. (1990). Stereoscopic and surface sound generation for exploratory data analysis. Proceedings of CHI'90, ACM Conference on Human Factors of Computing Systems, 125-132.

Smith, S., Pickett, R. M., & Williams, M. G. (1994). Environments for exploring auditory representations of multidimensional data. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 79 - 94.

Stevens, R., Brewster, S., Wright, P., & Edwards, A. (1994). Design and

evaluation of an auditory glance at algebra for blind readers. In G. Kramer and S. Smith (eds.), *Proceedings of the Second International Conference on Auditory Display* (Santa Fe, N.M., 7-9 November, 1994), 21 - 30.

Stifelman, L., Arons, B., Schmandt, C., and Hulteen, E. (1993). *VoiceNotes: A speech interface for a hand-held voice notetaker*. In *Proceedings of INTERCHI'93*, pp.179-186.

Swift, C. G., Flindell, I. H., & Rice, C. G. (1989). *Annoyance and implusivity judgments of*

environmental noises. *Proceedings of the Institute of Acoustics 1989 Spring Conference*. Vol. 11, Part 5, 551 - 555.

Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Wenzel, E. M. (1994). *Spatial sound and sonification*. In Kramer, G. (ed.), *Auditory Display*. New York: Addison-Wesley, 127 - 150.

Wenzel, E. M., Wightman, F.L. & Foster, S.H. (1988). *A virtual display for conveying three-dimensional acoustic information*. *Proceedings of the Human-Factors Society*, 32, 86-90.

Wenzel, E.M., Wightman, F.L. & Kistler, D.J. (1991). *Localization with non-individualized virtual acoustic display cues*. *Proceedings of CHI '91, ACM Conference on Human Factors in Software*, 351-359.

Wessel, D. (1979). *Timbre space as a musical control structure*. *Computer Music Journal*, 3(2), 45-52. Also appear in Roads, C. & Strawn, J.

(1985). *Foundations of Computer Music*. Cambridge MA: MIT Press.

Wightman F. L., and Kistler, D. J. (1989). *Headphone simulation of free-field listening. I: Stimulus synthesis*. *Journal of the Acoustical Society of America*, 85, 858 - 867.

Wildes, R., & Richards, W. (1988). *Recovering material properties from sound*. Richards, W. (ed.), *Natural computation*. Cambridge, MA: MIT Press.