Chapter 4:

AUDITORY INTERFACES

Introduction

Although scientific data representation has led the way in studies of non-speech audio interfaces, the use of non-speech audio for more general information is also being explored. In data representation, all the examples thus far have depended on the parameters of sound to represent the data information. We call this a reference to the sound source. Similarly, one can easily imagine informational cues being encoded in parameters of sound. There are many examples in the everyday world, the ticking of a clock for example. In general, these audio cues don't tell the listener so much about the source of the sound as about the fact that some event is taking place to which the listener might want to give attention.

An early example of using audio cues in interfaces was again scientific data. Note that the data values were not encoded in sound but rather the events surrounding the data. In 1975, Lee and Riviello used audio cues in a film portraying a two-step laser isotope separation process in 1975. Their work was about representing the values of the data variables in sounds but about using an audio interface to draw attention to isotope excitation and attraction. A gas of mixed U-235 and U-238 was exposed to a laser that excited the U-235. A beep corresponded to each isotope excitation. Subsequently, a second laser ionized those isotpies in the excited state so that they were attracted to a negative plate. For each ionization, a tone was heard which lasted until the ionized isotope reached the negative plate. If more than one isotope was ionized, then more than one tone was heard. Graphics displayed the motion of all isotopes. An isotope raised to an excited state was enlarged; an ionized isotope immediately fell towared the negative plate. However, the many events occurring on the display often distracted the observer's attention away from other critical events. The audio cues were valuable in drawing the observer's attention to the isotope excitiation and the subsequent ionization. Furthermore, the observer heard sequences of events without having to scan the display for rapidly changing situations.

Case Studies

As in scientific data perceptualization, the use of audio as interface cues is being explored in a variety of areas. Three case studies point to techniques and issues in auditory interfaces. First, Blattner and others have looked at a wide range of applications for sound in workstation environments with a particular focus on providing a basis for determining a set of sounds and the relationships among those sounds. Edwards, like Lunney and Morrison, is concerned with using sound to aid the visually handicapped in their use of computer workstations. Finally, somebody has taken seriously the notion of an experiment to test the effectiveness/use of audio cues.

Case Study 0: Alarms and Warning Systems

> [Certainly a familiar set of audio cues are those that provide alarms and warnings. [Why are we including this stuff anyway? It isn't really audio in the interface. Why not include stuff about the rings of telephones and the bells on clocks? And aren't these everyday sounds? i.e. Wouldn't I recognize a fire engine siren as a fire engine and not as a high pitched wail? On the other hand, what are we putting in this chapter? Not much....] Anyway, I think that Patterson's stuff isn't so much a case study as a precursor to thinking about audio cues in interfaces. Whadda ya say?]

Alarms and warnings are by far the most common nonspeech audio messages. Audio alarms range from ambulance sirens to computer error beeps, from stall alarms on aircraft to foghorns, from car horns to buzzers on clothes dryers. What all these uses of sound have in common is that they are meant to override ongoing processing and attract attention to themselves. In contrast, one of the largest challenges to auditory interfaces is to design sounds that can inform users without distracting them.

There is a danger that alarm systems can be *too* alarming. This is a problem with many existing systems. The severity of this problem is obvious if you consider, the next time that you fly in a commercial airliner, that if the plane should lose power the captain is more likely to start turning off alarms than to start correcting the situation. The reading we include in this section reports on the work of Roy Patterson and his colleagues to overcome this problem by systematically designing sets of alarms that are tailored to their sonic environment, and that are informative without being overwhelming.

There is a great deal we can learn by examining audio alarms and warnings. First, the design of traditional alarms reflects the kinds of psychoacoustic principles we discussed in Section 2. Second, some of the problems with existing systems of alarms can help make us aware of the kinds of traps we should avoid in designing our own auditory interfaces. Finally, newer systems of alarms (such as those presented in the reading) are beginning to resemble in their aims and concerns the kinds of audio outputs that are the subject of this tutorial.

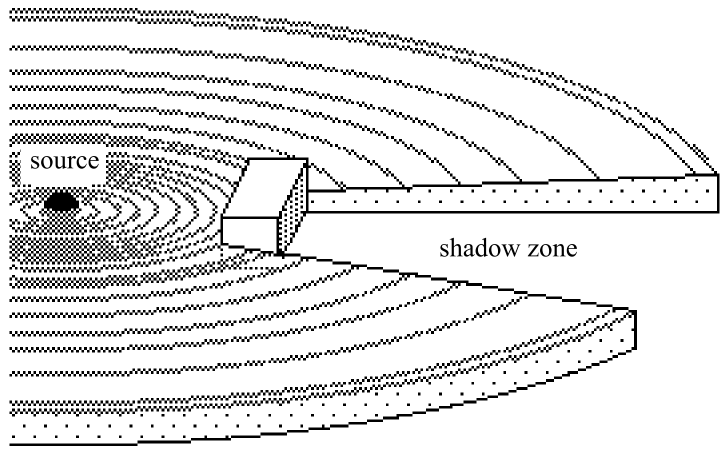
Alarms as Applied Psychoacoustics

Most of the principles of psychoacoustics we discussed in section 2 can be found reflected in the design of auditory alerts. For instance, think of the loud electronic drone that is used as a fire alarm in many buildings. Why is it so loud? Obviously so you can hear it in any part of the building, despite any ongoing activities – very loud sounds like these are likely to be immune to masking by virtually any other sound. In fact, they are so loud that they prevent very effective communication, which is good in that it convinces people to suspend their activities even if it *is* just another fire drill, but bad if it prevents conversations necessary to deal with the problem. Notice that many fire alarms do not change much in pitch or amplitude. Why is this? Apparently, the mere fact of a very loud sound is enough to alert people in office environments.

But think of the rising and falling pitch of the typical (at least to Americans) siren. Again, sirens tend to be very loud, and again this is to overcome possible masking noises. Why do they rise and fall? Because changing sounds tend to attract attention to a much higher degree than static sounds. In fact, many police cars now have a rather wide repertory of different sirens (one can count 10 distinct police alert sounds in New York City police cars), ranging from the traditional rising and falling pitch to the "do dah do dah" more familiar to Europeans to a sound like a massively overamplified cricket. Why all the different kinds of sounds? Two reasons. First, changing sounds reduces further the possibility that an alert will be masked by other noise (like a loud car audio system), since different sirens have different spectra. Second, and probably more important, changing sound patterns alert listeners even more than one repetitively changing pattern. Switching among different sirens introduces attention getting changes at a higher level than the variation of sound in any single siren.

In a very different domain, consider the low booming sound made by a foghorn. Why is this sound so different from the high-pitched wailing of a siren? Well, there are several reasons. Perhaps the most important is that foghorns must be audible for long distances, so that sailors far out to sea can use them for navigational purposes. Low frequency sounds lose much less energy as they propagate in air than high frequency sounds do – which is why you hear the bass thump thump of your neighbor's stereo much more than you do the highs.

A. High Frequency



B. Low Frequency

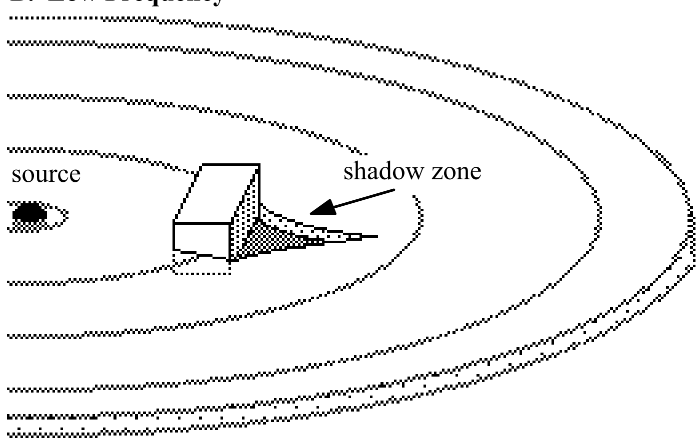


Figure 1: (A) High frequency sounds are greatly attenuated behind obstacles whose width are greater than the wavelength of the sounds; the area of attenuation is called the sound shadow. *(B) Sounds that have a long wavelength relative to the width of an obstacle cast relatively small sound shadows. This is one reason why foghorns make very low pitched sounds.*

In addition, the sounds made by foghorns must not be obscured by obstacles, such as a hill or point of land. As discussed in Chapter 2, sounds whose wavelengths are shorter than the width of an obstacle (i.e., high frequency sounds) are greatly attenuated behind it; the zone in which these sounds are muffled is called the *shadow zone* (see Figure 1). If a sound should be heard behind a hill, then its wavelength should be as long as possible – it should be a very low frequency sound.

Auditory alarms and warnings have evolved over long periods of time to attract and hold attention. Listening to them and considering the ways they are constructed can be quite valuable in understanding basic psychoacoustic principles such as those discussed in Section 2.

Problems With Traditional Alarms and Convergences with Audio Interfaces

At the most basic level, alarms are supposed to instantly draw attention to themselves, thus indicating that something is wrong with whatever triggered them. To be effective an alarm must not be masked by ambient sounds and must not be fade into the perceptual background.

Now creating a sound that draws attention and is not easily ignored is a simple task – ask any parent of a two–year–old. But there is a fundamental problem with the idea of an alarm as simply a loud, penetrating sound. Alarms usually signal emergencies, and the ability to communicate during an emergency is crucial. On the one hand, alarms must be heard, on the other hand, they must not prevent communication. As Patterson points out in the reading, most alarm designers opt for the "better safe than sorry" approach of extreme loudness. His aim has been to develop alarms that permit communication, but that will not be masked. He uses several techniques to achieve this goal. First, he tailors the alarms acoustic properties to the environments in which they will be used, so that they will not be masked even when played at fairly low levels. Second, alarms are not continuous, but rather are presented in "bursts" with silence in between. Finally, the relative urgency of the bursts are varied over the time course of the alarm, so that the initial burst has fairly high urgency, following bursts are more subdued, and a high-urgency burst is played if the problem is not corrected.

It may seem that in developing auditory interfaces problems of excessive loudness will not be relevant. After all, most of us are quite aware that auditory interfaces should not be annoying, and so they should probably be fairly quiet. But this leads us into the exact converse of the problem facing alarm designers: How can we create auditory interfaces that are quiet enough to be unobtrusive, without making them so quiet that they can't be heard? Patterson's approach of predicting how masking will affect hearing thresholds at different frequencies suggests that the answer to this question can be systematically approached.

Another problem with the idea of alarms being loud, attention–attracting sounds comes when an environment has several different alarms. Most of us are used to hearing (at most) one alarm in a given context – a siren as we drive, perhaps a fire alarm at work, and of course the error beeps made by our computer. Given that context constrains what we expect an alarm to indicate, many sounds can suffice in these situations. So it is that though the proliferation of error beeps made possible on the newer Macintoshes may be annoying, their functionality is pretty much unimpaired. Similarly, if we hear a new kind of siren while driving, we may marvel at its novelty, but we still know to pull to the side of the road.

But there are a many environments in which many alarms may be present simultaneously. For instance, in the Three Mile Island power plant crisis, the operators had to contend with over 60 different auditory warning systems (Sanders & McCormick, 1987). English hospital workers must contend with 33 possible warning sounds (Patterson et al., 1986). Most of these alarms are not designed with other possible alarms in mind, but instead are introduced to the environment with new equipment. In too many cases, one alarm (a repetitive beep, for instance) sounds pretty much the same as another (a slightly different repetitive beep). The problem, clearly, is in distinguishing between different alarms so that the particular problem can be recognized.

To give an idea of what this problem is like, think of being in a hallway near a series of offices, having a discussion with fellow workers. Suddenly a phone rings, and what happens? All the people with an office nearby disappear. The problem is that in such an environment, all phones tend to sound the same, so it is not clear what a particular phone ring means. And so it is with many complex environments with multiple alarms. In an emergency, it is important that alarms be distinguishable.

Note that when we start to be concerned about the ability to recognize the meanings of alarms we are beginning to address issues that are obviously important for informative audio interfaces. Clearly if one wants to convey several different messages using sounds, they should not be confusing, whether the messages concern emergency states or multidimensional data. Examining the ways developers of multiple alarms have addressed this problem can be very informative to developers of audio interfaces.

Work on alarms has converged with that on auditory interfaces even more than this, as the relative importance of different warnings is considered. For instance, Patterson et al. (1986) distinguished three kinds of hospital warnings: emergency, cautionary, and information available. What are information warnings but earcons, or auditory icons, or auditory cues?

It should be clear that work on auditory warnings converges with the research we are presenting here. Moreover, because such work is done from a different starting point, the contrasts between alarms and audio cues can be quite informative. For instance, whereas Patterson is concerned with characterizing the perceived urgency of sounds, we might want rather to understand how to make sounds less urgent and more unobtrusive. Alarms are meant to be instantly recognizable to listeners, while perhaps audio cues should work more like the sounds our cars make – general users, like Sunday drivers, might not know what all the sounds mean, but they know when something is wrong; and expert users, like mechanics, could have much more information available to them. In any case, it should be clear from this case study that alarms are not simply loud annoying sounds meant to attract attention.

Alarms in Critical Environments

Roy Patterson

Medical Research Council
Applied Psychology Unit
Cambridge, England

The following describes some recent work undertaken by Roy Patterson at the Medical Research Council Applied Psychology Unit in Cambridge. This work represents a "new wave" in auditory alarms and warning systems. It is a departure from the all too common brute force approach to alarms, where the first reaction in an emergency is to turn off the alarm rather than attend to the problem at hand. See Hopkins (1986) for a more detailed description.

- Aircraft, intensive care units, nuclear power plants
- Usually simple devices designed piecemeal with economy in mind
- The "better safe than sorry" approach
- W.W.II bomber alarms are still found in some airliners.

Problems with traditional alarms

- The sudden onset provokes startle reactions
- Loud alarms prevent effective communication
- The first reaction is often to turn off the #@$*! alarm.

• Difficult to remember and identify a range of alarms.

NOTES

1) Threshold plots for alarm sounds:

- can predict thresholds, optimal range for alarms.
- allows analysis of existing alarms, design of new ones.
- alarm on Boeing 727 much too loud.
- alarm on 747 too soft

2) Temporal confusion of alarms

Solution:

Design systems of alarms:

- Match spectra to environment
- Use intermittent alarms
- Make alarms easily recognizable
- Match sound urgency to situation

Attensons

1) Pulse

- acoustically tailored
- urgency controlled (attack, harmonicity, pitch/key)

2) Burst

- like a short tune
- imitates intonation of phrase
- long enough to be memorable

3) Alarm

- series of bursts with gaps
- bursts range in urgency
- gaps allow communication

Designed alarms

- These are the basis for an alarms standard
- Have been tested extensively, and designed with feedback.
- Less annoying, more easily recognizable than traditional alarms.
- Voice alarms are being introduced, but speech is slow, easily masked, easy to ignore

Implications for auditory interfaces

- Mapping
  - attensons mimic voice intonation, or use musical metaphor
  - design sounds to convey appropriate urgency

- Discriminability
  - spectral properties minimize masking
  - rhythms designed for maximum discriminability

- Annoyance
  - intermittent sounds reduce obtrusiveness
  - loudness, envelopes of sounds are tuned to acoustic environment

Case Study 1: Earcons and Icons

To validate the notion of auralization for exploratory data analysis, Bly ran a series of experiments. Something about the mapping, the first trials, and the final experiment with results. Show the graph from my dissertation.

M. Blattner, D. Sumikawa, R. Greenberg

Sound Example 4.1: Illustrate Meera's work with some of the usual stuff we use in the tutorial.

Case Study 2: Soundtrack, An Auditory Interface for Blind Users

The work of Mezrich, Frysinger, and Slivjanovski was important both in terms of the presentation offered, an integration of animated graphics and sound for time-varying data, and in the experimental data to test their approach. They also discovered informally that mapping the frequencies to the chromatic scale was more effective than using pure frequencies. Furthermore, they offered a system that was user-tailorable. The real-time interactive capabilities provided an important opportunity to highlight, or tease out, patterns in the the data that might convey information.

Somewhere I want to point specifically to Steve's SPIE paper and the discussion of "method".

Sound Example 4.2: Probably worth running through some various positions on the screen and a menu; if this is worth including at all.

Case Study 3: The Use of Auditory Cues to Reduce Visual Workload

M. Brown, S. Newsome & E. Glinert

Sound Example 4.3: I doubt that there is any example to use here. Again, I'm not even sure I think this is worth including as a case study.

Musical Sounds: Case Studies

[We have a bit of problem in calling this stuff "musical" sounds since not all examples are based on the musical scales. Some just map straight to any old pitch. I'll use musical sounds here but we should think about this. I really like Bill G.'s distinction of properties of sounds versus sources of sounds.]

Talk about the types of applications--data analysis and environmental cues: All work with non-speech audio for computer interfaces has appeared in two applications: sounds as dimensions for multivariate data presentation and sounds as cues for events or information in the computing environment. [Do I really believe there are only these two application types?] In the former case, sounds are used to aid scientists in identifying data patterns. Typically data variables or samples are mapped into sounds; the resulting notes are then played to the user for analysis. In using sounds as environmental cues, information such as error messages or actions such as locating windows are accompanied by sounds that provide feedback to the user.

Summary:

For all applications, the issues revolve around the perception of the sounds, the information the various sounds convey, and what information is best presented in the different sounds. However, most of the work has concentrated on applying sounds, on the use and effectiveness of sounds in relation to visual displays. In the domain of musical sounds, only Blattner has really addressed the issue of what sounds to use.

- Sounds are pressure variations that propagate as waves in the atmosphere.
- The waveform of a sound can be displayed in the time domain as amplitude by time.
- The time for one cycle of a repetitive wave is called the period, and is the inverse of the waves frequency (the number of repetitions per second).