1:

NONSPEECH AUDIO - An INTRODUCTION*

Introduction

This book is devoted to the thesis that human-computer interaction can be significantly enhanced through the improved use of the audio channel. Our focus is narrow and deals specifically with one especially neglected aspect of sound: the use of nonspeech audio to communicate information from the computer to the user.

Unless we are hearing impaired, nonspeech audio plays a significant role in helping us function in the everyday world. We use nonspeech audio cues in crossing streets, answering phones, diagnosing problems with our cars, and whistling for cabs. By virtue of such usage, we have built up a valuable set of everyday skills. Our thesis is that these skills have real potential as an aid in helping improve the quality of human interaction with complex systems. To date, however, these skills have been neglected and this rich mode of interaction has had little impact on how we interact with computers. Based on our own

---

* This introduction is a revised and expanded version of, Buxton, W. (1989). Introduction to this Special Issue on Nonspeech Audio, *Human-Computer Interaction* 4(1), Spring 1989. That, in turn, evolved from material that originally appeared in Chapter 9 of Baecker and Buxton (1987).

experience and pioneering studies by others, we feel that this should and can be changed. Helping to bring this about is our motivation in writing this book.

Video games illustrate the potential of nonspeech audio to effectively communicate useful messages. In games that use sound effectively, an expert player's score is lower with the audio turned off than it is when the audio is turned on. This is a clear indication that the audio conveys strategically critical information and is more than a nonessential frill.

As it is in play, so can it be in work. There are significant potential benefits to be reaped by developing our capabilities in the use of sound. Examples of where sound is already used extensively are process control, flight management systems, and user interfaces for the visually impaired. Even the sound of key clicks, disk drives and printers on personal computers convey useful feedback about system state.

There is an established literature in what the human-factors community call audio displays (Deatherage, 1972; Kantowitz & Sorkin, 1983; Patterson, 1982; Sanders & McCormick, 1987). We build upon this work and extend the range of applications for which nonspeech audio is used. For those interested in collected readings in the area, volume 4, number 1 (spring 1989) of the journal Human-Computer Interaction is a special issue devoted to the use of nonspeech audio at the interface. Farrell (1990) is another useful collection.

What About the Noise?

Let us address a question that inevitably comes up. It usually goes something like, "I work in a crowded office and the last thing I need is more noise to distract me," or, "When I'm thinking, what I really want is absolute silence."

Let us assume that we could create a perfectly silent work place. An anecdote by that 20th-century master of sound (and silence), composer John Cage, describes what that would be like. He recounted that after sitting quietly in an anechoic chamber for about half an hour, he was struck by the fact that he heard two sounds: a high sustained sound and a low pulsating one. On asking about them after the fact, he was told that one was his nervous system and the other was his circulatory system.

This anecdote serves to illustrate that there is no such thing as silence. In performing our day-to-day tasks, we are surrounded by sounds. Some help us, others impede us. The former are information; the latter are noise. However, despite their potential effect on our performance, we exercise little influence over the ambient sounds of our working, playing, and living environments.

An underlying thesis of the work presented in this tutorial is that we can benefit by exercising greater control over the sounds around us. Namely, by effective design, we can reduce the noise component and increase the information-providing potential of sound. Our ambition is to promote the acquisition of an understanding that will support the design of audio cues that will improve human performance in computer mediated tasks: "designer sound" for computer systems.

Figure and Ground in Audio

Audio brings some important and interesting properties to the repertoire of the user interface designer. One of the more interesting is the ability of most users to monitor simultaneously a number of nonspeech audio signals while performing a motor/visual task.

This can be seen in driving a car. Consider that driving a car at 80 m.p.h. on a motorway is a critical task in which error could result in death. Nevertheless, one can perform the task, with the radio on, while holding a conversation with the passenger. Despite concentrating on the conversation, one can still monitor what is on the radio and, if of sufficient interest, interrupt the conversation to point out a favorite melody. While all of this is going on, one could well be passing another car and, in the process, changing gears. A clicking sound confirms that the turn signal is working, and if the car has a manual transmission, audio cues (rather than the tachometer) will most likely determine when to shift. And throughout all of this, one is immediately aware if the engine starts to produce some strange noise, or if an ambulance siren is audible.

In contrast with the driving example, the use of audio with the computer is very much impoverished. Our belief, however, is that this need not be so. Although the example's task space was full of sound, most of it was functional and, therefore, not noise. It helped the driver and passenger achieve their agendas and performance potentials. What the research we discuss in this tutorial is aimed at is the use of sound in computing to achieve the same thing.

Sound and the Visually Impaired

The trend toward visually oriented direct-manipulation interfaces means that computer-based access to work and independent living were more accessible to the blind 10 years ago than they are today. Some universities in the United States, for example, require all first-year students to purchase a Macintosh computer. Bowe (1987) pointed out that this is tantamount to saying that "No blind people need apply." Despite (or because of) the progress of icons and windows, it is becoming increasingly difficult to adapt existing software for use by the blind. Using estimates for the United States for 1990, we can get a sense of the significance of this issue from the following: 3.4% of the population have visual impairments, .7% have severely impaired vision, and .2% are legally blind (Elkind, 1990) .

The applicability of this work extends beyond the visually impaired. There is a case to be made that if we saw the real world in the same restricted way which we see our computer displays, we would probably not be able to be certified for a driver's license. As our displays become more visually intensive, the visual channel becomes increasingly overloaded, and we are impaired in our ability to assimilate information through the eyes. In such instances, the same audio cues that help the totally blind user can be applied to help the normally sighted.

Some Examples

In the last ten years, several researchers from a variety of disciplines have begun using non-speech sounds as part of their user interfaces. This exploratory work divides along many dimensions: scientific data analysis and office workstation environmental cues, musical notes and everyday (or natural) sounds, enhancements to visual displays and replacements of visual displays, and application problem solutions and technological innovations.

Of course, the counter situation exists, and a reliance on audio without alternative modalities of communication places a barrier to access to those with hearing impairments. Using the same US estimates for 1990, 8.2% of the population are hearing impaired, 2.6% have mild to moderate bilateral hearing loss, and 1.0% have severe to profound bilateral hearing loss (Elkind, 1990).

In applications, existing work has appeared in two modes: sounds as dimensions for multiversity data presentation and sounds to provide feedback and other information to support interaction. Early examples of the former are Bly (1982a, 1982b), Mezrich Frysinger and Slivjanovski (1984) and Lunney and Morrison (1981). In these examples, data variables were mapped into sounds and the resulting notes were then played to the user for analysis. Examples of the other trend, using audio to support interaction, are Gaver's *SonicFinder* for the Macintosh (Gaver, 1989) and Edwards' text editor for the visually impaired (1989). Both use sounds as cues to events in their computing environments, although in very different ways; however, in each, actions such as selecting files, locating windows, or searching for text strings are accompanied by sounds that provide feedback to the user. These and other examples will be discussed in more detail in later chapters.

The kinds of sounds can be categorized as musical sounds, sounds created by specifying pitch, loudness, duration, and waveshape, and everyday sounds, sounds that are perceived in terms of events in the natural environment such as a door slamming or people applauding. Thus, the musical sounds focus on the properties of the sound itself, while the everyday sounds focus on the source of the sound. Bly, Mezrich et al, Morrison and Lunney, Blattner et al., and Edwards map their information into properties of sounds; Gaver use familiar environmental sounds. For all of the authors, issues revolve around the perception of sounds, the information the various sounds convey, and what information is best presented in the different sounds.

Much work is oriented toward reducing the visual workload by providing additional or redundant information in sound. Morrison and Lunney and Edwards have worked to extract the necessary information from visual displays and encode it into sounds so that visually impaired users can use computers effectively. Other examples do not necessarily attempt to replace the visual display but rather to augment it. A few results suggest that some information may be more readily accessible to all users when presented in sound than when presented visually.

Most of the work thus far has concentrated on the use and effectiveness of sounds in relation to visual displays. However, Gaver and Blattner et al. address the issue of what sounds to use. They devote much of their work to considering the capabilities of sounds, both in the perception of the sounds and the kinds of information their sounds might encode.

Chapter 1 Sound in Collaborative Work

Another area where sound has particular relevance is in computer supported collaborative work (CSCW). Of particular relevance are situations where people in remote locations are collaborating synchronously on some computer mediated activity. Here, the problems of human-computer interaction are compounded, since participants have to maintain an awareness of their collaborator's activities. This is in addition to the regular overhead of monitoring and directing their own. Since one can only visually attend to one thing at a time, this especially taxes the visual system - even more so when all participants are not seeing the same thing on their screen.

Recent work with the ARKola system (Gaver, Smith & O'Shea, 1991) has demonstrated how nonspeech audio can be used to create what might be called a "shared acoustic ecology" for the participants in the shared activity, and thereby enhanced the sense of telepresence, or shared space.

The nonspeech audio in this approach functions as an awareness server by allowing each user to hear what others are doing in the background, while concentrating on their own activities in the foreground. As the state of the art advances, we will be able to hear what is

happening, where it is happening and who caused it - all using the audio channel and existing skills that we have acquired from a life-time of living in the everyday world.

Nonspeech Audio and Multimedia

Multimedia is emerging field which is getting significant attention. To some, its absence as a specific topic in this book may be questioned. However, just because it is not highlighted as a specific topic does not mean that it is not important, or that it is not dealt with. From one perspective, the entire book is about multimedia. It all depends on what one means by the term.

We believe that the importance of "multimedia" lies in the modalities and channels of communication employed, not the media used. Viewed this way, the book can be seen as addressing some issues concerning one such channel: audio. Virtually everything that we say, therefore, has relevance to that aspect of multimedia systems and applications.

Hopefully, having read this book, the multimedia author will have gained a number of insights that will help in going beyond the use of audio for sound effects and sound tracks.

Function and Signal Type

Functionally, nonspeech audio messages can be thought of as providing one of three general types of information: alarms and warnings, status and monitoring indicators, and encoded messages. Typically, different types of audio cues are used for each.

Alarms and Warning Systems: These are signals that take priority over other information. Their purpose is to interrupt any ongoing task and alert the user to something that requires immediate attention. They normally only sound in an "exception" condition. They are usually loud, easily identifiable sounds with sharp transients. In the car driving example, this is illustrated by the ambulance siren. Doll and Folds (1986) provide an interesting discussion of the contrast between principles of ergonomics and current practice in the use of auditory signals in modern fighter aircraft.

Status and monitoring messages provide information about some ongoing task. The nature of such cues very much depends on the type of task being monitored.

The key-click produced when typing on a conventional keyboard is one example of how audio cues can provide status feedback for short discrete tasks. In typing, the sound cue only indicates whether the key has been pressed or not. However, Monk (1986) showed that one can go beyond this. In an experimental situation, he showed how mode errors could be reduced by a third by having the pitch of the sound associated with each keystroke depend on which of two modes the system was in. Likewise, Roe, Muto and Blake (1984) showed that audio feedback provided a powerful cue, complimenting tactile and kinesthetic feedback in operating a membrane switch.

For ongoing continuous tasks, sounds providing status information are usually sustained tones or repeating patterns that are audible for the duration of the process that they are monitoring. In such cases, unlike alarms, these messages are designed to fade rapidly into the background of the operator's consciousness, so that attention can be directed to some other foreground task. They are designed to come back into the foreground only when there is a significant change in the process being monitored. The design of this type of message

exploits the fact that the human perceptual system does not remain conscious of steady-state sounds. In contrast, it is very sensitive to change. Hence, if a steady-state sound representing an ongoing background task stops, then that transition will bring the fact of a change in state to the user's attention. The sound of a washing machine turning off is one such example. In the driving example, any change in the normal background sound of the car motor is another.

Humans are capable of monitoring more than one such signal in the background, providing that the sounds are appropriately differentiated. As with alarms, however, if more than one simultaneously requires attention, then it is likely that the user will become confused and performance will be affected. An actual case in which this was evident was the Three Mile Island power plant crisis. In this case, the operator had 60 different auditory warning systems to contend with (Sanders & McCormick, 1987, p. 155). This example illustrates that although we can recognize and simultaneously monitor a number of different audio cues, we can normally only respond to one or two at a time.

Encoded messages are used to present numerical (or quantitative) data, such as statistical information, in patterns of sound. The complex and varying sounds used in this type of application contrast with the penetrating one or two sounds used with alarms or with the steady-state tones or patterns used in status monitoring.

The design of this class of message often exploits our capabilities of pattern matching and recognition. In some cases, such messages are much like musical melodies. The usage has a lot in common with Wagner's use of leitmotiv, Procofiev's use of motives to represent the characters in *Peter and the Wolf* and the sounds in the video game PACMAN.

Audio Cues and Learning

Just as we do not know the meaning of the themes in Procofiev's *Peter and the Wolf* without being told, we are not born knowing the meaning of a fog horn, fire alarm, or police siren. They must be learned. Furthermore, the quality of their design with respect to human perceptual and cognitive capabilities affects how easy or hard this learning process will be. If audio cues are to be used in interactive systems, then the quality of their design is important. As graphic design is to effective icons (AIGA, 1982), so acoustic design is to effective auditory signs or *earcons*. If audio cues are to be employed, they must be clear and easily differentiated. To be effective, they require careful design and testing.

In his work, Gaver (1989, 1986) makes the point that we can accelerate the learning process by using everyday sounds in the interface. As pointed out by Blattner, et al. (1989), this use of everyday sounds is analogous to the use of representative (as opposed to abstract) graphic icons.

Gaver's work is directed at exploiting our skills built up over a life-time of everyday listening. His intention is to design user interfaces that use the same skills employed in everyday tasks such as crossing the street. An example of such a skill would be our built-up association of reverberation with empty space: all other things being equal, the more reverberant a room, the more space.

Gaver proposes that such cues be used to convey information about a computer's status and, because the cue is based on existing every day skills, they will be quick to learn and not easily forgotten. A way of using our sense of reverberation, for example, would be to have a reverberant "clunk" every time that we saved a file, and have the amount of

reverberation indicate how much free space is left on the disk. Similarly, on the Apple Macintosh, placing a file into the "trash can" could be accompanied by an appropriate "crash." These are concepts that Gaver develops in his article, which describes the rationale behind the prototype *SonicFinder* for the Apple Macintosh.

Much of the work in nonspeech audio interfaces has been based on mapping attributes of data onto the parameters of sound. These techniques depend on using the listener's knowledge of this mapping as the basis for communication. Doing so has its benefits and builds upon our skills acquired in listening to music, for example. It is important to note, however, that Gaver's approach is quite different. His use of sound is based on a theory of sources. That is, what is important is what you think made the sound, rather than the psycho-physical properties of the sound itself. This is a distinction that is developed in Chpters 5 and 6.

Perception and Psychoacoustics

In the preceding sections, we discussed the importance of design in the use of acoustic stimuli to communicate information. One of the main resources to be aware of in pursuing such design is the available literature on psychoacoustics and the psychology of music.

Psychoacoustics tells us a great deal about the relationship between perception and the physical properties of acoustic signals. Music and the psychology of music tell us a lot about the human's ability to compose and understand higher level sonic structures. In particular, the literature is quite extensive in addressing issues such as the perception of pitch, duration, and loudness. It is also fairly good at providing an understanding of masking, the phenomenon of one sound (e.g., noise) obscuring another (e.g., an alarm or a voice). These and other topics are covered in Chapter 2.

Under a different name, acoustic design has had a thriving life as music. Although music perception is not a part of main-stream human factors, it does have something to contribute. In particular, classic psychoacoustics has dealt primarily with simple stimuli. Music, on the other hand, is concerned with larger structures. Hence, melodic recognition and the perception and understanding of simultaneously sounding auditory streams (as in counterpoint) is of great relevance to audio's use in the human-computer interface. As a reference to this aspect of auditory perception, see Deutsch (1982, 1986) and Roederer (1975).

Emotion and the Affective Aspects of Audio

To be added. Build upon Tajadura-Jiménez & Västfjäll (2008).

The Logistics of Sound

In the past, one of the biggest problems in exploring the use of audio signals was a logistical one. For example, Bly (1982a, 1982b) had to build special hardware and a custom computer interface in order to undertake her work.

Some important recent developments have changed this situation dramatically. The change is due to the adoption of guidelines by the music industry for a common protocol for interfacing electronic sound synthesis and processing equipment to computers. This

standard is known as MIDI, the Musical Instrument Digital Interface (IMA, 1983). An excellent general introduction to MIDI can be found in Loy (1985).

The MIDI protocol specifies guidelines for the physical, logical, and electrical aspects of the interconnect. It uses an inexpensive serial protocol. Specialized interfaces are available for most personal computers. As a result of MIDI, there is a wide range of inexpensive and easily available computer controllable sound synthesis and processing equipment available to the interested researcher. MIDI, MIDI controllable devices, and the practicality of equipping a lab for studying nonspeech audio are discussed in detail in Chapter 6. (See also Buxton and Moran, 1990). For the first time, from a logistical perspective, the audio channel has become viable as an important mode of interaction.

Summary

We believe that the audio channel deserves more attention. This is shown in the papers cited in this introduction and in later chapters. This research is a beginning, and it is clear that there is still a long way to go before the channel is used to its full potential. Nonetheless, it demonstrates the potency of the approach and whets the appetite for more.