Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653648

Auditory Icons: Using Sound in Computer Interfaces

William W. Gaver [a]
[a] University of California, San Diego.

Online Publication Date: 01 June 1986

HUMAN-COMPUTER INTERACTION, 1986, Volume 2, pp. 167-177

Auditory Icons: Using Sound in Computer Interfaces

William W. Gaver
University of California, San Diego

ABSTRACT

There is growing interest in the use of sound to convey information in computer interfaces. The strategies employed thus far have been based on an understanding of sound that leads to either an arbitrary or metaphorical relation between the sounds used and the data to be represented. In this article, an alternative approach to the use of sound in computer interfaces is outlined, one that emphasizes the role of sound in conveying information about the world to the listener. According to this approach, auditory icons, caricatures of naturally occurring sounds, could be used to provide information about sources of data. Auditory icons provide a natural way to represent dimensional data as well as conceptual objects in a computer system. They allow categorization of data into distinct families, using a single sound. Perhaps the most important advantage of this strategy is that it is based on the way people listen to the world in their everyday lives.

Author's present address: William W. Gaver, Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, CA 92093.

CONTENTS

1. INTRODUCTION

As digital sound-producing software and hardware become increasingly available, interest is growing in using sounds more complex than the ubiquitous interrupt beep to provide information to computer users. In particular, it seems desirable to use sound in a way that is analogous to the use of visual icons to provide information. Researchers have shown that sound can be used successfully to present time-varying, multidimensional, and logarithmic data, and that people can readily recognize certain kinds of patterns in data represented by sound (Bly, 1982; Mansur, Blattner, & Joy, 1985; Mezrich, Frysinger, & Slivjanovski, 1984; Morrison & Lunney, 1985). These results are encouraging, but the strategies employed so far have been overly constrained by a traditional psychophysical understanding of sound and hearing. Typically, dimensions of sound itself, such as pitch, loudness, or duration, have been used to represent dimensions of data. This has led to the use of artificial-sounding tones with no analogs in the everyday world. The result is a mapping between sound and data that is often more like an unlabeled graph than an icon.

I have been developing an alternative approach to the use of sound in computer interfaces, one that emphasizes the role of sound in conveying information about the world to the listener. The basic idea is that people listen to the world to find out the source and environment of a given sound. From this perspective, an auditory icon is a sound that provides information about an event that represents desired data. Instead of using dimensions of sound to stand for dimensions of the data, dimensions of the sound's source are used.

2. AUDITORY ICONS

One can imagine how a single sound could be used to give information about a file arriving in a message system. The file hits the mailbox, causing it to emit a characteristic sound. Because it is a large message, it makes a rather weighty sound. The crackle of paper indicates a text file—if it had been a compiled program, it would have clanged like metal. The sound comes from the left and is muffled: The mailbox must be in the window behind the one that is currently on the left side of the screen. And the echoes sound like a large empty room, so the load on the system must be fairly low. All this information from one sound!

2.1. Hearing the World, not the Sound

I refer to this kind of informative sound as an *auditory icon*. It depends on an understanding of sound that emphasizes the way listeners use sound to gain information about the world. When a file hits a mailbox, for example, you might pay attention to the sound itself: its duration, the changing pitch, or how the loudness varied with time. This experience is a product of attending to the *proximal stimulus,* the variations of air pressure near the ear. It is this kind of experience with which psychophysicists and musicians are concerned. But it is more likely that you will concentrate on the *source* of the sound—that the file is large and heavy, for example, and that the mailbox is metal. This is a very different kind of experience from the first. One doesn't seem to notice a sound's pitch, for instance, when identifying the source as paper hitting a metal enclosure; conversely, one doesn't notice the heaviness of the file when concentrating on the loudness of the sound.

Our normal mode of hearing is to listen to sounds to identify the events that cause them. From this perspective, sound provides information about materials interacting at a location in an environment. This perspective is related to Gibson's ecological approach to perception (Gibson, 1979), and may also be seen as exploring the computational level of auditory processing (Marr, 1982). According to this view, sound cannot be understood solely in terms of the physics of sound waves (the proximal stimulus) and the experience engendered by them. The physics of the source events themselves (the *distal stimulus*) must also be understood and related both to the proximal stimulus and experience.

If we consider sound as providing information about the world, we can use a source of sound to stand for a source of information. This is the strategy behind auditory icons. Auditory icons are caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes.

One advantage of this approach is that auditory icons may represent conceptual objects in a computer system more clearly than other sounds. Currently, some aspect of the timbre of sounds is usually varied to indicate separate channels of data. But because timbre is conceived in terms of the proximal stimulus, the relation to the source of data is wholly synthetic, and must be learned and relearned. If a good mapping between a source of sound and a source of data can be found, the meaning of an auditory icon should be easily learned and remembered.

Auditory icons should also be well suited to representing dimensional data. For instance, the magnitude of some value could be represented by the size of a virtual sound-producing object. The mapping between the values of the dimensions should be obvious: A large object stands for a large value. From this perspective, there are as many dimensions of sound that can be used to represent dimensions of data as there are physical dimensions of sound-producing events that can be distinguished by a listener.

Finally, considering sound as providing information about materials, interactions, and environments allows an organization of data into distinct categories. In the mailbox example, for instance, one of the objects interacting to produce the auditory icon was a file, while another was the mailbox. But both the size and the type of the file were communicated as well as the identity and location of the mailbox. Similarly, the type of interaction can be used to provide another sort of information: Scraping, for example, might indicate that a message had been moved within the mail file, and the force of the interaction might provide information about the processing time needed for the job. Not only can multidimensional data be presented using auditory icons, but the dimensions can be grouped into conceptual "families."

3. MAPPING INFORMATION TO REPRESENTATIONS

It seems useful at this point to consider how information can be mapped to representations. By many standards, this will be a woefully inadequate discussion of representation types, but it should be helpful in comparing the use of natural sounds as auditory icons to the more traditional manipulations of sound. For more rigorous discussions of the relations between signs and their meanings, texts such as Bates (1979) or Peirce (1932) may be consulted; Hemenway (1982) and Bertin (1981) considered such issues more explicitly in terms of the graphical representation of information.

The dimension of concern here is the kind of mapping between the data to be represented and the means used to represent it (see Figure 1). At one extreme, the mapping between the data and the representation is *symbolic*. Symbolic mappings are essentially arbitrary, relying on social convention for their meaning. Telephone bells, sirens, and stop signs are examples of symbols. At the other extreme of this dimension are representations which have a *nomic* relation to the information they convey—their meaning depends on the physics of the situation (see Heil, 1983). The relation between a sound and its source or a photograph and the scene it depicts are examples of nomic mappings: The representations are images of the information. Finally, some kinds of mappings are best considered *metaphorical*. Metaphorical mappings make use of similarities between the thing to be represented and the representing system: They are not wholly arbitrary, yet they do not depend on physical causation. Metaphorical mappings include structure-mapping (Gentner, 1983), in which similarities between the structures of two things are exploited, as well as metonymic mappings, in which a feature is used to indicate a whole. For example, the mapping between genealogy and a tree is an example of structure-mapping, whereas the

Figure 1. **Visual and auditory representation systems may be characterized by the kind of mapping between the data to be represented and the means used to represent it.**

Mapping	Visual	Auditory
Symbolic	 for stop	**Sirens** for approaching ambulance
	 for peace	**Applause** for approval
Metaphorical	 for organization of relationships	**Hiss** for snake
	 for horses	**Pitch** for falling
Nomic	 for scissors	**Mailbox sound** for arriving mail
	 for file	**Hit wood or metal sound** for size of object

use of a hiss to stand for a snake is a metonymic mapping. Thematic motives, which include Wagner's "Leitmotifs" and the use of sound in video games such as Pac Man, use metaphors based on similarities between the temporal progressions of the sounds and the events they are to represent. Other examples, such as the use of changing pitch to stand for changing heights, involve the use of a dimensional metaphor, in which one ordered dimension is used to represent another.

Physically, a nomic mapping is a necessary mapping: A given event produces a certain sound, and a given sound has one certain cause. However, psychologically there may be a many-to-many mapping between source and sound. Thus the same source (or rather, what we are willing to call the same source) may make different sounds, and a given sound may be perceived as having been produced by a number of possible sources. So it is that we can mimic the sound a trumpet makes, for instance, by causing a loudspeaker cone to vibrate in the right way. It may be best to speak of necessary constraints when talking about a nomic mapping, rather than a necessary, one-to-one mapping. Nonetheless, it should be clear that the relation between a natural sound and information about its source is very different from a symbolic mapping from sound to the information to be represented.

It is important to note that a given mapping may fall between the three categories outlined above. If a weak metaphor is used, or if the metaphor is not understood, a metaphorical mapping becomes increasingly symbolic. Similarly, nomic mappings depend on models of the source events that produce an image. As these models become more approximate, the result becomes more like a metaphor than a true image. Finally, even a nomically mapped auditory icon depends in some sense on a metaphor: The mapping will be nomic to some event in the model world presented to the user, not to underlying events in the computer itself.

3.1. Learnability of Mappings

Hutchins, Hollan, and Norman (1986) discussed the relation between the form and meaning of a statement in terms of *articulatory directness*. There is no obvious link between a beep, for instance, and the information that you are in the wrong mode to execute a given command; there is little articulatory directness in this representation. If, however, the action of a paintbrush in a graphics program were accompanied by the appropriate swishing sound, the representation would be an articulatory direct one.

From this point of view, it is clear that nomic mappings have more articulatory directness than metaphorical ones, which have more articulatory directness than symbolic ones. An increase in articulatory directness should be accompanied by an increased ease of learning (once the mapping has been pointed out; see Bates, 1979). In other words, the more a representation's form depends on its meaning, the easier it should be to learn. Thus nomic mappings should be relatively simple to learn, metaphorical mappings somewhat harder, and symbolic mappings the most difficult. Note that articulatory directness does not necessarily affect performance once a mapping has been well learned; it seems trivially easy to recognize the meaning of a stop sign or a telephone bell if one is from the appropriate culture. Still, the increased use of international road signs indicates that behind this ease is a great deal of learning, and if one is concerned with developing new representational systems, learnability is of course a crucial issue.

Traditional manipulations of sound, in terms of the proximal stimuli, by definition will not produce a nomic mapping between a sound and the information it is to represent. Thus using parameters of the sound itself (e.g., pitch or loudness) constrains possible mappings to be either arbitrary or, at best, metaphorical. A nomic mapping can only be produced by manipulating the sound in terms of its source—by using natural sounds. Of course, natural sounds may also be metaphorically or arbitrarily related to the information they are to represent. But in general, using natural sounds should make possible a higher degree of articulatory directness than manipulating dimensions of the sound itself.

This is not to say that using auditory icons will always be better than the more traditional, symbolic uses of sounds. A symbolic representation system that already exists in the culture and that is appropriate to the information (i.e., a telephone bell to signal a call on an interactive "talk" program) should probably be chosen over a nomic but unfamiliar mapping. Still, all things being equal, using auditory icons—natural sounds with a nomic or metaphorical mapping to the information to be represented—seems likely to be more powerful than creating auditory symbol systems out of whole cloth.

Natural sounds are related to events in a principled, systematic way (described by physics), and people learn this mapping from early childhood in their interactions with the world. Thus we can tell the sound of something hitting a piece of wood from that of something hitting metal, though the sounds themselves are physically quite similar. There is an underlying system to the similarities and differences between the sounds which depends on their sources; thus the relation between a natural sound and its meaning is articulatorily direct. Such a system does not exist ready-made for sonic variables based on the physics of the proximal stimulus. It is necessary to invent one, and many of the mappings of such an invented system are liable to be unintuitive, contradictory, and difficult to learn.

4. ARE AUDITORY ICONS FEASIBLE?

Obviously, the mailbox example is an exaggeration of what auditory icons might be like in the near future. We simply do not yet know how much of the information about size, weight, and other factors is conveyed acoustically. Moreover, it is not clear how much computing power will be needed to specify and produce this information. One of the advantages of using dimensions of sensation, such as pitch and loudness, to represent data is that systems with the capability for manipulating these qualities are widely available.

Yet there is reason to believe that the potential benefits of auditory icons may be realized. Remember that auditory icons are meant to be caricatures of natural sounds. They need be no more realistic a representation of the sounds they stand for than visual icons are of the things they look like. All they must do is capture the essential features of the sonic events they portray. It is not necessary that they even approach the acoustic resolution required to produce high-quality music on computers. So it is reasonable to suppose that they could be relatively inexpensive in terms of computing resources. Moreover, with the advent of MIDI, a standardized communication scheme for digital music devices (see Loy, 1985), as well as the sophisticated and cost-effective sound-synthesis and sampling capabilities of the new wave of personal computers, more powerful digital sound-production tools should be readily available in the near future.

A fairly simplistic model of the physics of a source event may suffice to model the sounds produced. Warren and Verbrugge (1984) showed that the difference between a bouncing object and a breaking one may be characterized by manipulating higher-order temporal properties of spectrally identical sounds. Bouncing may be described as a series of repeating, spectrally consistent sounds with diminishing time between repetitions. Breaking consists of an initial burst of noise followed by several overlapping series of sounds, each having different spectral qualities and different repetition rates (basically a noise burst followed by fast bouncing objects).

Similarly, I have been examining the information available in the sounds made by struck materials. In particular, I am interested in what can be heard about the identity – wood versus metal – and the length of a struck bar. Figure 2 shows the results of time-varying Fourier analyses of sounds produced by hitting three lengths of wood and metal bars. A study of the physics of the vibrations of struck wood and metal bars has suggested a number of physical dimensions of the sound sources that account for their acoustic differences. For instance, the most salient acoustic correlate to the length of a struck object is the component frequencies of the sound made (longer bars produce lower frequencies). Because the metal used in this study is harder and denser than the wood, the same length of metal produces higher-frequency sounds than that of wood. The forms of energy loss during vibration are also different between these two materials, which cause high-frequency vibrations to last longer in metal than those at lower frequencies, and conversely with wood. Finally, wood is an inhomogeneous material, so the overall amplitude of its sounds do not decay smoothly; instead there is a pronounced steady-state in its amplitude which seems important for its recognition.

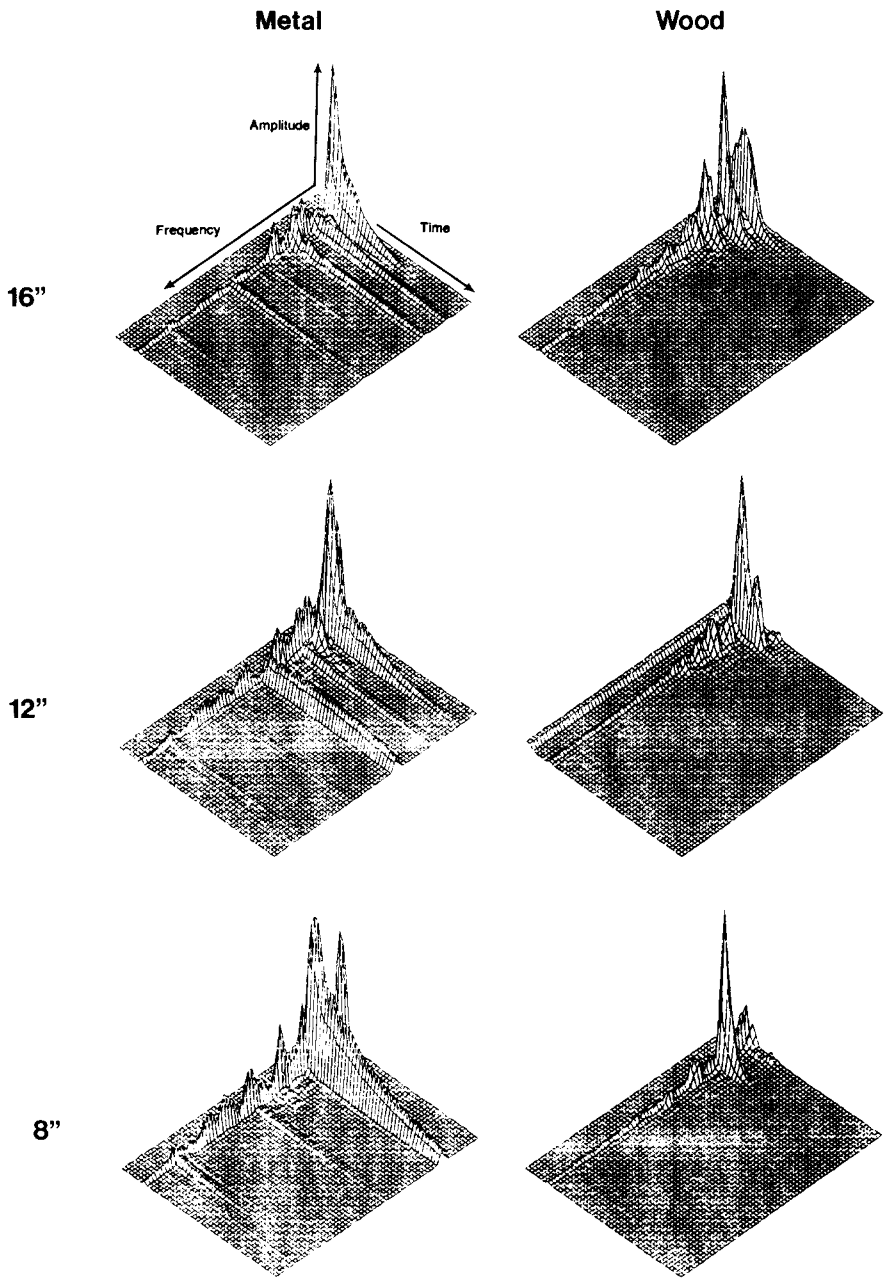
These studies suggest that it may be possible to characterize seemingly complex natural sounds fairly simply. The breaking and bouncing sounds studied by Warren and Verbrugge (1984) could be used to convey categorical information, whereas manipulating the physical dimensions of the sound-producing objects I studied could encode dimensional information. Think of these as cartoon sounds: caricatures that don't really look like (sound like) the objects they represent, but that capture their essential features.

5. USING SOUND IN COMPUTER INTERFACES

What I am suggesting, then, is that we should base sonic representations neither on the sensations of sound we hear when listening to music, nor on psychophysical experiments, but rather on the way we listen to the world in everyday life. Only in this way will we utilize the full potential of sound to give us information about the world.

There are three aspects of this approach that need to be addressed in future research. First, we need to know more precisely what information sound can actually convey about the world. For instance, different interactions between objects seem to give different information about the materials involved. To tell about the size of an object, it should be hit, whereas if we care about its texture,

Figure 2. **Time × Frequency × Amplitude plots for six sounds made by hitting objects. The plots on the left represent the sounds made by hit metal pieces; those on the right are the sounds of hit wood. The length of the struck objects decreases from top to bottom: The top row shows sounds made by 16-in. pieces, the middle are 12 in., and the bottom are 8 in. long.**



the object should be scraped. We also need to know how much of the available information in sound people actually use. In some cases, people seem to focus exclusively on information obtained visually, ignoring what they hear (see Massaro & Cohen, 1983, for an example of this in phonemic perception). In other instances, people may attend to only a limited amount of the total information available in a sound.

The second task is to determine how to create effective acoustic caricatures of the sonic events in which we are interested. For instance, it would be desirable to know how to characterize sound qualities that specify properties of materials, such as their size or weight, separately from those that provide information about interactions between them. The research by Warren and Verbrugge (1984) on the sounds produced by breaking and bouncing objects, as well as my work on hit objects, are examples of this kind of endeavor.

Finally, when a nomic mapping between an auditory icon and the information it is to represent is impractical, the creation of metaphors between sound-producing events and the processes and data to be represented must be addressed. In this event, the information of interest conveyed by an auditory icon is not the source event itself, but instead what that source event represents. Thus a system of mapping between source events and the information to be represented is necessary. The solution to this problem may follow from the creation of overall metaphors from the computer interface to other environments (e.g., the desk-top model). But it is important to remember that such metaphors were created with predominantly visually oriented systems in mind; the addition of sound is likely to shape those metaphors in new ways.

In spite of these difficulties, exploiting our ability to gain information about the world through sound seems worthwhile. Using auditory icons should prove a valuable supplement to already existing symbolic uses of sound. Moreover, such an approach has several advantages over using traditional manipulations of sound to create new representations. Using natural sounds allows a nomic mapping between an auditory icon and the information to be conveyed. Auditory icons provide a natural and intuitive way to represent dimensional data and to represent conceptual objects in a computer system. They allow categorization of data into distinct families, using a single sound. Perhaps the most important advantage of this strategy is that it is based upon the way people listen to the world in their everyday lives.

Acknowledgments. I thank Don Norman, Jim Hollan, and the members of the Intelligent Systems Group for their many helpful comments on this article. Many of the ideas in Section 3 are the result of a series of conversations with Don Norman. Anne Schlottman also gave many valuable suggestions for Section 3.

Support. This research was supported by a grant from the System Development Foundation and by contract N00014-79-C-0323, NR 667-437 with the Personnel and Training Research Programs of the Office of Naval Research. Support was also provided by the Naval Personnel Research and Development Center, and by Grant PHS

MH 14268 to the Center for Human Information Processing from the National Institute of Mental Health.

REFERENCES

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic.

Bertin, J. (1981). *Graphics and graphic information-processing.* Berlin: Walter de Gruyter.

Bly, S. (1982). *Sound and computer information presentation* (UCRL-53282). Doctoral dissertation, Lawrence Livermore National Laboratory and University of California, Davis.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gibson, J. J. (1979). *The ecological approach to visual perception.* New York: Houghton Mifflin.

Heil, J. (1983). *Perception and cognition.* Berkeley: University of California Press.

Hemenway, K. (1982). Psychological issues in the use of icons in command menus. *Proceedings of the CHI '82 Conference on Human Factors in Computer Systems,* 21-24. New York: ACM.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1986). Direct manipulation interfaces. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 87-124). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Mansur, D. L., Blattner, M. M., & Joy, K. I. (1985, January). Sound-graphs: A numerical data analysis method for the blind. *Proceedings of the 18th Hawaii International Conference on System Sciences, 18,* 63-174.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753-771.

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

Morrison, R., & Lunney, D. (1985). [contribution to panel on communicating with sound]. *CHI '85 Proceedings,* 118-119.

Peirce, C. S. (1932). In C. Jartshorne & P. Weiss (Eds.), *Collected Papers of Charles Sanders Peirce.* Cambridge, MA: Harvard University Press.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704-712.

HCI Editorial Record. First manuscript received December 12, 1985. Revisions received April 3, 1986, and July 16, 1986. Accepted by William Buxton. —*Editor*