THE MANY FACETS OF AUDITORY DISPLAY

Meera M. Blattner*
University of California, Davis
Lawrence Livermore National Laboratory
Livermore, CA 94550

INTRODUCTION

It would be difficult to imagine a virtual world without sound. Sound surrounds us constantly; the ability to hear sound is one of our basic senses. Why hasn't sound become an integral part of the human-computer interface? There are many historical reasons why this is the case: the letters of the alphabet when typed into a keyboard were easily interpreted into binary form for textual displays upon a screen or printed page. Voice input has many difficulties when used as an input medium and corresponds more to the difficulties of using handwritten input, where errors occur because input is not precise. In the past, nonspeech audio has been associated with music and has not been used for conveying information, with certain exceptions such as bugle calls, fog horns, talking drums, etc., which were not universally known and limited in scope (Ref. 1). Interfaces of the future will be designed for human expression with all its subtleties and complexities (Ref. 2). One such example is the use of facial expressions used by interface agents that act as guides to assist their human users in making decisions. The range of expressiveness in audio is just starting to be appreciated by interface designers. Audio can be used to create intense emotion through music or to enhance our perception of real-word phenomena through auditory display.

In this presentation we will examine some of the ways sound can be used. We make the case that many different types of audio experiences are available to us. We should not limit our use of audio to one type of sound or even several types. A full range of audio experiences include: music, speech, real-world sounds, auditory displays, and auditory cues or messages. The technology of recreating real-world sounds through physical modeling has advanced in the past few years allowing better simulation of virtual worlds. Three-dimensional audio has further enriched our sensory experiences.

SOUNDSCAPES

The computational limitations of real-time interactive computing do not meet our requirements for producing realistic images for virtual reality in a convincing manner. Regardless of these restrictions the representations can be no better than the graphics. Computer graphics is still limited in its ability to generate complex objects such as landscapes and humans. Nevertheless, useful and convincing visualizations can be made through a variety of techniques.

A similar situation is true for sound for virtual reality. It is beyond our ability to create interactive soundscapes that create a faithful reproduction of real world sound; however, by choosing one's application carefully and using sound to enhance a display rather than only mimic real-world scenes, a very effective use of sound can be made.

SOUNDSCAPES

■ **We cannot create interactive soundscapes that are .faithful reproductions of real world sounds.**

■ **We should use sound to enhance a display rather than only mimic real-world sounds.**

Figure 1

OBJECTS

What we hear is very different from what we see. We do not always hear objects in the real world. Some objects are only heard because they are out of sight. Some objects make sounds at some times, but not others. Objects only make sounds when they set up vibrations in a medium that surrounds us, such as air or water. This means that they must create a movement to make a sound. Even movement does not always create sounds we can hear.

To associate sounds with an object that does not ordinarily make sound is artificial and there may not be natural associations with these sounds. Sound can be used very effectively to indicate the presence of objects that are not seen. Film sometimes does this through the use of music or other sound effects.

OBJECTS

■ **Objects are usually seen.**

■ **Some objects are only heard.**

■ **Sometimes the presence of an object is detected by sound and interaction with another object**
  ◆ footsteps
  ◆ coughing
  ◆ doorbells
  ◆ electrical equipment

■ **Sound may tell us how an object is constructed.**

Figure 2

MUSIC AND SPEECH

A powerful use of music is found in film scores. Music comes to bear in helping to realize the meaning of the film, in stimulating and guiding the emotional response to the visuals. Music serves as a kind of cohesive, filling in empty spaces in the action or dialogue, and the color and tone of music can give a picture more richness and vitality and pinpoint emotions and actions. It is the ability of music to influence an audience subconsciously that makes it truly valuable to the cinema. Music specific to particular cultures is used in the study of history, geography and anthropology. A scene placed in a geographical context may be enhanced by local music.

Speech is required for detailed and specific information. It is through speech (rather than through other sounds) that we communicate precise and abstract ideas. Speech may be used as input as well as output in the computer interface. Very little is known about building successful speech interfaces for two-dimensional displays, let alone three-dimensional interfaces.

TYPES OF SOUNDS

✔ **Music**
✔ **Speech**
  Real-world sounds
  Auditory displays
  Cues and auditory messages

  3D Auditory displays

Figure 3

REAL-WORLD SOUND AND AUDITORY DISPLAYS

Real-world sounds are the natural sounds of the world around us, such as leaves rustling or birds singing, or man-made sounds such as machine noises or even a band playing in the background. What about sound in our everyday life? Real-world sounds are essential to our sense of presence in a scene that depicts our world around us. R. Murray Schafer (Ref. 3) describes "soundscapes," as historical reconstructions of the sound that surrounds people in various environments. Examples are street criers, automobiles, the crackling of candles, church bells, etc.

Auditory displays include the interpretation of data into sound, such as the association of tones with charts, graphs, algorithms or sound in scientific visualization. These auditory display techniques are used to enable the listener to picture in his or her mind real-world objects or data. An example of auditory display is the work done by Mansur, Blattner and Joy (Ref. 4), in which points on an x-y graph were translated into sonic equivalents with pitch as the x-axis and time on the y-axis (a nonlinear correction factor was used). Recently, Blattner, Greenberg, and Kamegai (Ref. 5) enhanced the turbulence of fluids with sound, where audio was tied to the various aspects of fluid flow and vortices.

TYPES OF SOUNDS

Music
Speech
✔ **Real-world sounds**
✔ **Auditory displays**
Cues and auditory messages

3D Auditory displays

Figure 4

Generally audio cues will be considered auditory icons or *earcons,* to provide information to the user. This information tends to be more abstract than that received through auditory displays. Auditory signals are detected more quickly than visual signals and produce an alerting or orienting effect (Ref. 6). Nonspeech signals are used in warning systems and aircraft cockpits. Alarms and sirens fall into this category, but these have been used throughout history, long before the advent of electricity. Examples are military bugle calls, post-horns, church bells that pealed out time and the announcements of important events.

Work on *auditory icons* was done by Gaver (Ref. 7) and *earcons* by Blattner, Sumikawa, and Greenberg (Ref. 8). Gaver uses sampled real-world sounds of objects hitting, breaking, and tearing as described in real-world sounds above. However, Gaver's auditory icons are meant to convey information of a more abstract nature, such as disk errors, etc. Gaver used the term "everyday listening" to explain our familiarity with the sounds of common objects around us. Blattner, Sumikawa, and Greenberg took musical fragments, called motives, and varied their musical parameters to obtain a variety of related sounds. We describe the construction of earcons below.

TYPES OF SOUNDS

✔ **Cues and Auditory Messages**

■ **Auditory messages or signals were used by people before the discovery of electricity.**

■ **Bells, bugles, trumpets and drums sent information to the countryside or announced the arrival of an important person.**

Figure 5

VIRTUAL REALITY, TELEPRESENCE
AND TELECONFERENCING

The types of sounds described above, speech, music, audio cues, and real-world sounds, can all be located in a three-dimensional audio environment. Sound localization by NASA has shown the effectiveness of separating voices in space to improve their clarity (Ref. 6). Cohen and Wenzel (Ref. 9) are studying the three-dimensional acoustic properties of teleconferencing systems to filter out extraneous sounds by the use of "audio windows." The general idea is to permit multiple simultaneous audio sources, such as in a teleconference, to coexist in a user-controlled display to easily move through the display and separate the channels while retaining the clarity and purity of the sounds.

TYPES OF SOUNDS

- ✔ **3D Audio**
- ■ **Virtual Reality**
- ■ **Telepresence**
- ■ **Teleconferencing**

Figure 6

THREE-DIMENSIONAL SOUND

Three-dimensional (localized) sound truly immerses the listener in his or her auditory environment. The basis of the work in three-dimensional acoustic displays is psychoacoustics. The virtual acoustic environment is part of the NASA Ames View System (Ref. 6). The technology of simulating three-dimensional sound depends on reconstructing the sounds as they enter the ears. The acoustic signals are affected by the pinnae (outer ear) and the distance and direction of the ears. Microphones were placed in the ears of humans or mannequins to measure this effect, called the head-related transfer function. A real-time system, the Convolvotron, is used to filter incoming sounds using a head-related transfer function (Ref. 6).

3-D AUDITORY DISPLAY
Synthesis Technique

PINNAE (OUTER EAR) RESPONSES MEASURED WITH PROBE MICROPHONES

PINNAE TRANSFORMS DIGITIZED AS FINITE IMPULSE RESPONSE (FIR) FILTERS
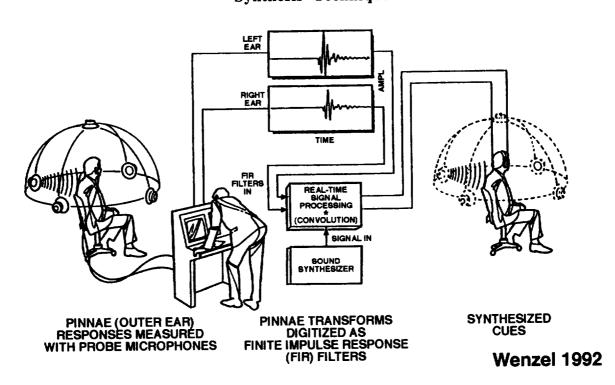
SYNTHESIZED CUES

Wenzel 1992

Figure 7

PARAMETERS OF SOUND

Audio has dimensions or parameters. In nonspeech audio these parameters are manipulated to provide the symbols or syntax of messages. The dimensions of sound are (Ref. 9):

- harmonic content
  - pitch and register (tone, melody, harmony)
  - wave shape (sawtooth, square, ...)
  - timbre, filters, vibrato, and equalization
- dynamics
  - intensity/volume/loudness
  - envelope (attack, decay, sustain, release)
- timing
  - duration, tempo, repetition rate, duty cycle, rhythm, syncopation
- spatial location
  - direction (azimuth, elevation)
  - distance/range
- ambiance: presence, resonance reverberance, spaciousness
- representationalism: literal, abstract, mixed.

THE PARAMETERS OF SOUND

■ **Graphical parameters are (Ref. 10):**
  - ➤ size
  - ➤ saturation
  - ➤ texture
  - ➤ orientation
  - ➤ shape
  - ➤ color

■ **Sound parameters are (Ref. 9):**
  - ➤ harmonic content
  - ➤ dynamics
  - ➤ timing
  - ➤ spatial location
  - ➤ ambiance
  - ➤ representationalism

Figure 8

SAMPLED VERSUS DIGITIZED SOUND

Sampled sounds are digital recordings of sounds which we can hear. These sounds have the advantage of immediate recognizability and ease of implementation into computer interfaces. Synthesized sounds are those sounds which are created algorithmically on a computer. They can be made to sound similar to real-world sounds through sound analysis (such as Fourier analysis) and trial-and-error methods. Since synthesized sounds are created algorithmically, it is easy to modify such a sound in real time by altering attributes like amplitude (volume), frequency (pitch), or the basic waveform function (timbre). Furthermore, it is easy to add modulation of amplitude or frequency in real time to create the effects of vibrato or tremolo without changing the basic sound. It is for these reasons that sound synthesis is so popular in music creation today. Synthesized sounds offer a high degree of flexibility with a reasonable amount of ease. A drawback of synthesized sound is that each algorithm used typically mimics some sounds very well and others not as well.

Since sampled sounds are digital recordings, they can reproduce with extremely high accuracy any sound which can be heard. However, the amount of work required to attain equal flexibility in modification, compared with synthesized sounds, is very high. Typically, sampled sounds are modified only in amplitude (volume) and frequency (pitch).

SOUNDS ON COMPUTERS

- **Sampled sounds are digital recordings.**
- **Synthesized sounds are sounds which are created algorithmically.**
- **Synthesized sounds may be modified in real time by altering attributes like amplitude (volume), frequency (pitch), or the basic waveform function (timbre).**
- **Typically, sampled sounds are modified only in amplitude (volume) and frequency (pitch).**

Figure 9

THE STRUCTURE OF AUDIO MESSAGES

Earcons are short, distinctive audio patterns to which arbitrary definitions are assigned. They can be modified in various ways to assume different but related meanings. The building blocks for earcons are short sequences of tones called motives. From motives we can build larger units by varying musical parameters. The advantage of these constructions is that the musical parameters of rhythm, pitch, timbre, dynamics (loudness) and register can be easily manipulated. The motives can be combined, transformed, or inherited to form more complex structures. The motives and their compounded forms are called earcons. However, earcons can be any auditory message, such as real-world sounds, single notes, or sampled sounds of musical instruments.

EARCONS - SOUND MESSAGES

- **Motives: The basic melodic and rhythmic units**
- **A motive is either a single pitch or a sequence of two to four pitches**
- **The family motive is the specific durational sequence (rhythm) associated with the motive**
- **A motive has variable parameters of**
  - **timbre (tone color)**
  - **dynamics (loudness) and**
  - **register (high/low pitches)**

Figure 10

EARCON CONSTRUCTION

A motive may be an earcon or it may be part of a compounded earcon. Let A and B be earcons that represent different messages. A and B can be combined by juxtaposing A and B to form a third earcon AB. Earcon A may be transformed into earcon B by a modification in the construction of A. For example, if A is an earcon, a new earcon can be formed by changing some parameter in A to obtain B, such as the pitch in one of its notes. A family of earcons may have an inherited structure, where a family motive, A, is an unpitched rhythm of not more than five notes and is used to define a family of messages. The family motive is elaborated by the addition of a musical parameter, such as pitch (A+p = B) and then preceded by the family motive to form a new earcon, AB. Hence, the earcon has two distinct components, an unpitched motive followed by a pitched motive with the same rhythm. A third earcon, ABC, can be constructed by adding a third motive, C, with both the pitch and rhythm of the second motive, but now has an easily recognizable timbre (A + p + t = B + t = C).

CONSTRUCTING EARCONS

- **Combining**
  - ➤ The process of combining to create an earcon means linking different motives together in a chain-like sequence.

- **Transforming**
  - ➤ The process of transformation cosmetically alters a motive by changing its timbre, register, and/or tempo.

- **Inheriting**
  - ➤ The process of inheriting is one in which a single earcon is heard in an increasing complex chain.

Figure 11

84

Inheriting earcons

Error messages for novice users
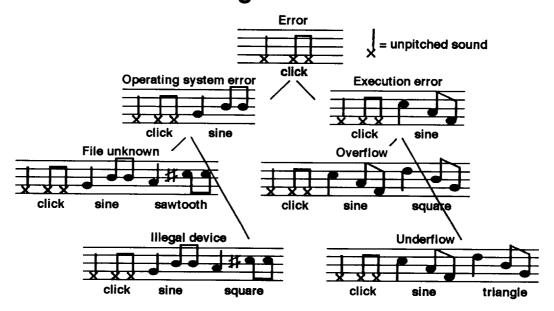

Figure 12

To display more than one earcon their temporal locations with respect to each other have to be identified. Two primary methods are used: overlaying one earcon on top of another and the sequencing of earcons (Ref. 11). Some sort of merging or melding in to new sound could be considered; for example, the pitch of two notes can be combined into a third pitch. Programs typically play audio without regard to the overall auditory system state. As a result, voices may be played simultaneously or occur with several nonspeech messages making the auditory display incoherent. An audio server is being constructed that blends the sounds of voice, earcons, music, and real-world sounds in a way that will make each auditory output intelligible (Ref. 12).

MULTIPLE EARCONS

How to combine earcons with each other?
Two primary methods are considered here:

■ **Overlaying one earcon on another.**

■ **The sequencing of earcons.**

■ **Some sort of merging into a new sound.**

Figure 13

WILL AUDIO MESSAGES BE USED?

Will sounds as abstract as earcons be accepted by the majority of users? The advantages are very clear: they are easily constructed on almost any type of workstation or personal computer. The sounds do not have to correspond to the objects they represent, so objects that either make no sound or an unpleasant sound still can be represented by earcons without further explanation. Auditory icons that make real-world sounds usually can be recognized quickly; however, most messages do not have appropriate iconic images.

Brewster, Wright, and Edwards (Ref. 13) found earcons to be an effective form of auditory communication. They recommended six basic changes in earcon form to make them more easily recognizable by users. These changes were: 1) use synthesized musical timbres, 2) pitch changes are most effective when used with rhythm changes, 3) changes in register should be several octaves, 4) rhythm changes must be as different as possible, 5) intensity levels must be kept close, and 6) successive earcons should have a gap between them. Earcons are necessarily short because they must be learned and understood quickly. Earcons were designed to take advantage of chunking mechanisms and hierarchical structures that favor retention in human memory. Furthermore, they use recognition rather than recall. If earcons are to be used by the majority of computer users, they must be learned and understood as quickly as possible taking advantage of all techniques that may help the user recognize them.

EFFECTIVENESS

■ **Will sounds that convey information in a form as abstract as earcons be accepted by the majority of users?**

■ **The advantages:**
✔ **Easily constructed**
✔ **Do not have to correspond to the objects they audify**

■ **Brewster, Wright and Edwards found earcons to be an effective form of auditory communication.**

Figure 14

MAPS

To test our theories, we chose to combine auditory display techniques with two-dimensional maps. Maps are primarily used for orientation, navigation within, and analysis of, geographic terrain. However, a broad range of additional information may be of interest in some cases: average annual rainfall, soil composition, location of mineral deposits and other natural resources, location of rail lines, location of historical sights, various economic factors, elevation, etc.

MAPS

■ **Maps are used for orientation and navigation.**

■ **Other information of interest:**
  ➤ average annual rainfall
  ➤ soil composition
  ➤ location of mineral deposits
  ➤ location of rail lines
  ➤ location of historical sights, elevation
  ➤ ownership
  ➤ utilities

Figure 15

Because they need not be static, computerized maps can take advantage of many more methods for displaying information than can traditional paper maps. Enlarged windows can appear at a point of interest; numerical data can pop up on demand and disappear when no longer needed. Animation and pseudocolor can be used to track or call attention to specific information. Nevertheless, because the addition of visual data requires that space be allocated for it, a saturation point will eventually be reached beyond which interference with text and graphics already on display cancels without any possible benefit. In such cases (and others), it may be advantageous to present some of the data in a sonic representation. Auditory maps were used by Kramer to enhance Magellan's view of Venus. Auditory output to convey information such as the emissivity (i.e., radiation) and gravity of the area being viewed. The auditory output did not disrupt the view of the underlying landscape.

COMPUTERIZED MAPS

- **Computerized maps can take advantage of many more methods for displaying information.**

- **Enlarged windows can appear at a point of interest.**

- **Numerical data can pop up on demand and disappear when no longer needed.**

- **A saturation point will eventually be reached with text and graphics.**

Figure 16

MAP IMPLEMENTATION

The floor plans are visible, as are geographical data such as roads, parking lots, etc. Information associated with each building includes sewer lines, water lines, power lines, number of computers and people housed within, the department or administrative unit in charge, construction type, level of security clearance required to work there, job titles of those in the building, etc. As the cursor is dragged over the image, relevant information is presented to the user. The mouse may remain still and information requested by the user clicking on an appropriate button. Sonic information must be presented in a "short form" when the mouse is in motion but can be presented in a "long form" when it is stationary. The short form cannot encode sufficient information to distinguish between items within a family, whereas the long form can easily do so. The functionality required involved retrieval not only of the location of particular items, but also of area information. We also needed a way to provide summary data to users. It would be too slow and inefficient to scan an entire scene with a mouse! So we had to develop methods for scanning areas and presenting multiple data. Summary data is used to indicate that there were many items of a certain type in an area. We chose a simple method to handle summary data - by a linear mapping of earcon volume (loudness) to the magnitude of the numeric value.

AUDITORY MAPS

■ An experimental system was implemented on a Silicon Graphics INDIGO workstation, "visual" cartographic data were changed into sonic representations.

■ The data used was a map of Lawrence Livermore National Laboratory.

■ This particular map was selected because it was in the form of a machine-readable data base of buildings.

Figure 17

EARCON IMPLEMENTATION OF MAPS

We used timbres of various musical instruments to create earcons to help differentiate the sounds, as suggested by Brewster, Wright, and Edwards (Ref. 13). The sampled waveforms are altered by varying the frequency multiplier and the amplitude factor, to create different pitches and volumes. The frequencies range from 100 Hz to 2,000 Hz. None of the earcons varied in loudness within themselves. Instead, we interpreted summary data using dynamics; that is, sets of data with more like items are louder than sets with smaller amounts of like data.

AUDITORY MAPS

■ **Timbres of musical instruments for earcons help to differentiate the sounds.**

■ **TRANSFORMATIONS:**
➤ **Knocking earcons are administrative access.**
➤ **The computer earcon are earcons transformed over the x-axis, y-axis, and both axes with a timbre change.**

Figure 18

Simple sounds as well as complex ones built by combination, transformation, and inheritance from motives and other earcons were used in auditory maps. There is a simple earcon of a tom-tom drum pounding, which represents a building restriction. The pounding is transformed in pitch and frequency to indicate different access restriction levels of buildings. Higher restrictions are represented by faster, higher-pitched knocking. A simple three note earcon in one pitch using a saxophone timbre indicates all properties of the first earcon except that they have different pitch changes. It is important to note that each family of earcons shares the same timbre. Since timbre is one of the most easily recognized attributes of sound, one can immediately identify the family of an earcon just by recognizing what instrument is used in playing that earcon. A combination of these different earcons can be used to build new, more complex, earcons. For instance, a combination of three earcons can indicate a physics buildings with a clearance level of confidential which houses Sun computers.

AUDITORY MAPS

■ INHERITANCE:

A pitchless earcon indicates an administrative building. The second level motives inherit all the properties of the first earcon except that they have pitch and timbre. The pitch is the same, but the timbre varies.

Figure 19

TEMPORAL QUALITIES

Sequential combinations are sounds which are heard one after another. Concurrent sounds are those which are either played simultaneously or which partially overlap in time. Combined sounds are those whose attributes are combined into a single new sonic item. An approach that is used to combine data is to consider every mapping from a data item to a dimensional coordinate system, where the coordinates are sonic attributes. The user may choose to listen concurrently or sequentially. However, if more than four earcons need to be sounded in concurrent mode, then the first four will play concurrently after which as each earcon ends another will begin. We have not implemented combined data at this time.

TEMPORAL COMBINATIONS

- **The user has a choice of listening concurrently or sequentially.**

- **If more than four earcons are displayed in the concurrent mode then four will play, and as each one ends another begins.**

Figure 20

SPEED SELECTION

Audio information can be displayed in two modes: moving and stationary. When the user moves the cursor over a region there is not sufficient time to play its earcon. A series of short, truncated sounds inform the user that there are items of interest in that location. Stationary mode is indicated when the mouse is clicked and the long form of the data under the cursor can be displayed. Selection while moving can be turned off if the user wishes.

SPEED SELECTION

■ MOVING

When the cursor is moving there is not sufficient time to play its earcon. Short sounds inform the user that there are items of interest in that location.

■ STATIONARY

Stationary mode is indicated when the cursor is clicked and then the long form can be displayed. Moving has two modes: on and off.

Figure 21

94

A CENTRALIZED AUDIO PRESENTATION SYSTEM

User interfaces which support concurrent program executions have little, if any, audio management. Typically, a number of audio channels exist and programs request the number of audio channels required. The operating system either grants or denies the request. Therefore, in environments where multiple programs output sound, each individual program has no overall context of the auditory systems state with the possible exception of how many audio channels have been allocated. Programs typically play audio without regard for the overall auditory environment which can cause sound masking and perceptual unintelligibility.

A CENTRALIZED AUDIO PRESENTATION SYSTEM

- **Motivation**
  - ➤ **Maintain the intended informational encoding**
  - ➤ **Perceptual issues can then be addressed.**
  - ➤ **Simultaneous speech and/or non-speech audio presentation**
  - ➤ **Maximize clarity of each request**

- **Multiple representations in audio requests**
  - ➤ **Abstract earcons**
  - ➤ **Representational earcons**
  - ➤ **Voice**
  - ➤ **Sampled sound (no semantic content)**
  - ➤ **Other representations**

Figure 22

The presentation manager receives descriptive messages which contain information about system activities and program states, as specified by the user or application programmer. The sonic output of the set of running programs and the overall auditory system state is controlled by the presentation manager. It chooses how the information is to be presented in sound, within the constraints of the descriptive message. The presentation manager must choose the form with consideration for other current output.

MOTIVATION

■ **Maintain the intended informational encoding**

■ **Perceptual issues can then be addressed**
➤ **Simultaneous speech and/or non-speech audio presentation**
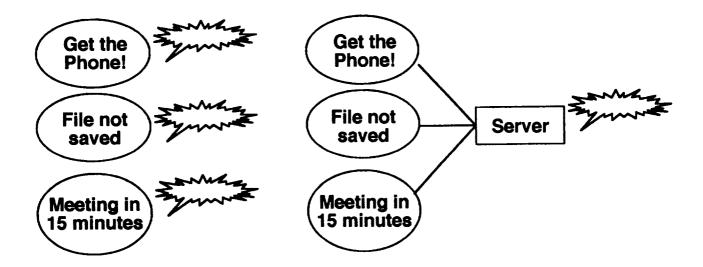➤ **Maximize clarity of each request**

Figure 23

PRESENTATION MANAGER DESIGN

The audio presentation manager is composed of three distinct parts: the descriptive message server, the medial selector, and the scheduler. A message passing paradigm serves as the underlying model for communication between application and the various parts of the presentation manager.

Whenever an application is to represent some information in sound, it sends a message to the presentation manager. This message includes a high level description of the information to be displayed. The presentation manager then decides how the message is to be displayed.

SYNTHESIZER MODULES

■ **Currently available:**
➤ Earcon synthesizer
➤ Voice synthesizer
➤ Sine wave synthesizer
➤ "Sample" player

■ **In production:**
➤ Algorithmic music synthesizer

Figure 24

REQUIREMENTS FOR A SYNTHESIZER MODULE

In order to give the presentation manager as much flexibility as possible, applications need not send raw auditory information. Rather, many common forms of data can be sent to the presentation manager along with more general information about that data. Depending upon the information received and the current auditory state of the system, the most appropriate auditory representation for the data will be used in its presentation.

REQUIREMENTS FOR A SYNTHESIZER MODULE

- A set of variables which constitutes its "state"
- An initialization routine for the state variables
- Must be able to compute the next n samples from the current state, and this computation must occur within $n/sample\_rate$ seconds
- The ability to algorithmically encode its impact on the other forms
  - Do sounds produced by this synthesizer interfere with the perception of other sounds?
  - Can this form be sounded simultaneously with itself?

Figure 25

Audio request contains many parameters to guide the decisions of the server
    Priority, latency; Is it interruptible?
    Semantics upon interruption
        Re-play the whole request
        Continue from where the request was interrupted
        Remove the request completely
    Desirability of each audio form
    Information specific to the different synthesizer modules

How to determine which requests to play now:
    Function of priority, latency, current audio system state, and forms for presentation
    Must be above a minimum threshold or else that request is postponed

How to determine which form for each sounded request:
    What forms are already playing? Penalties associated with multiple forms (i.e., two or more voices)
    Which forms are preferable to the application?
    Which forms are preferable to the user in general?

SERVER DECISIONS

■ Audio request contains many parameters to guide the decisions of the server.

■ How to determine which requests to play now.

■ How to determine which form for each sounded request.

Figure 26

AUDIO EXAMPLE

We implemented a simple navigation system that
Saxophone--crossed intersection
Rising trumpet--north
Drums--south

Speed indicated by beeps

Approaching intersection
      Hear choices
      Suggests which way to go
      Request when crossing intersection
      Tone indicates distance from the finish -- low to high

EXAMPLE APPLICATION

■ **With a strong voice user bias**

■ **With a strong abstract earcon bias**

■ **With no strong biases**

Figure 27

SUMMARY

■ Audio is richer in three-dimensions.
■ Sound sources are clearer when separated in space.
■ The sense of immersion in an artificial world is greater when sound surrounds the listener.
■ Sound can be used to replace touch.
■ Sound can impart abstract information.

Figure 28

REFERENCES

1. Blattner, M. M., and Greenberg, R. M., Communicating and Learning Through Nonspeech Audio, Multimedia Interface Design in Education, Springer Verlag, NATO ASI, Series F, 1992.

2. Blattner, M. M., "In Our Image: Interface Design in the 1990s," IEEE Multimedia, Vol. 1, No. 1, 1994, pp. 25-36.

3. Schafer, R. M., The Tuning of the World, Alfred Knopf, New York, 1977.

4. Mansur, D.L., Blattner, M.M. and Joy, K.I., "Sound Graphs: A Numerical Data Analysis Method for the Blind," Journal of Medical Systems, Vol. 9, No. 3, 1985, pp. 163-174.

5. Blattner, M. M., Greenberg, R. M. and Kamegai, M., "Listening to Turbulence: An Example of Scientific Audiolization," Multimedia Interface Design, Blattner and Dannenberg (eds.), ACM Press/Addison-Wesley, Reading, MA, 1992, pp. 87-104.

6. Wenzel, E. M., "Three-Dimensional Virtual Acoustic Displays," Multimedia Interface Design, Blattner and Dannenberg (eds.), ACM Press/Addison-Wesley, Reading, MA, 1992, pp. 257-288.

7. Gaver, W. W., "Auditory Icons: Using Sound in Computer Interfaces," Human-Computer Interaction, Vol. 2, 1986, pp. 167-177.

8. Blattner, M. M., Sumikawa, D. A. and Greenberg, R.M., "Earcons and Icons: Their Structure and Common Design Principles," Human-Computer Interaction, Vol. 4, No. 1, 1989, pp. 11-44.

9. Cohen, M. and Wenzel, E. A., "The Design of Multidimensional Sound Interfaces," Virtual Environments and Advanced Interface Design, Barfield and Furness (eds.), Oxford University Press (to appear).

10. Bertin, J., Semiology of Graphics, University of Wisconsin Press, Madison, WI, 1983.

11. Blattner, M. M., Glinert, E. P. and Papp, A. L., III, "Sonic Enhancement of Two-Dimensional Graphic Displays, " Auditory Display: Sonification, Audification and Auditory Interfaces, G. Kramer (ed.), Addison-Wesley, Santa Fe Institute, Reading, MA, 1994, pp. 447-460.

12. Blattner, M. M. and Papp, A. L., III, "A Centralized Audio Presentation Manager," Proceedings of the International Conference on Auditory Displays, Santa Fe Institute, 1995, pp. 231-240.

13. Brewster, S. A., Wright, P.C. and Edwards, A. D. N., "An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces," Proceedings of INTERCHI '93, Amsterdam, 1993.