HUMAN-COMPUTER INTERACTION, 1989, Volume 4, pp. 11-44
Copyright © 1989, Lawrence Erlbaum Associates, Inc.

Earcons and Icons: Their Structure and Common Design Principles

Meera M. Blattner, Denise A. Sumikawa, and **Robert M. Greenberg**
Lawrence Livermore National Laboratory and The University of California, Davis

ABSTRACT

In this article we examine *earcons*, which are audio messages used in the user-computer interface to provide information and feedback to the user about computer entities. (Earcons include messages and functions, as well as states and labels.) We identify some design principles that are common to both visual symbols and auditory messages, and discuss the use of representational and abstract icons and earcons. We give some examples of audio patterns that may be used to design modules for earcons, which then may be assembled into larger groupings called *families*. The modules are single pitches or rhythmicized sequences of pitches called *motives*. The families are constructed about related motives that serve to identify a family of related messages. Issues concerned with learning and remembering earcons are discussed.

A preliminary version of this article was presented at the Nineteenth Annual Hawaii International Conference on System Sciences.

Meera M. Blattner is also with the Department of Biomathematics, M.D. Anderson Cancer Research Hospital, University of Texas Medical Center, Houston, TX 77030.

Authors' present addresses: Meera Blattner, L-540, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550; Denise Sumikawa, L-262, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550; Robert Greenberg, 4241 Edge Drive, Oakland, CA 94602.

Samsung et al. v. Techiya
IPR2024-01031, Exhibit 1025

CONTENTS

1. INTRODUCTION
2. SOME WORK RELATED TO THE STUDY OF AUDIO INFORMATION
3. ICONS: THE GRAPHICAL COUNTERPART OF EARCONS
    3.1. The Use of Icons in Existing Systems
    3.2. Types of Icons
    3.3. Iconic Families
    3.4. The Design of Icons
4. EARCON DESIGN PRINCIPLES
    4.1. Representational Earcons
    4.2. Abstract Earcons
5. MOTIVES
    5.1. Rhythm
    5.2. Pitch
    5.3. The Variable Parameters of Motives
        Timbre
        Register
        Dynamics
    5.4. Length of Earcons
6. EARCON CONSTRUCTION
    6.1. One-Element Earcons
    6.2. Compound Earcons
    6.3. Inherited Earcons
    6.4. Transformed Earcons
7. LEARNING, REMEMBERING, AND UNDERSTANDING EARCONS
    7.1. Learning and Remembering Earcons
    7.2. Classes, Types, and Hierarchies
    7.3. Space Complexity and Memory
    7.4. A Summary of Learning and Memory Issues for Earcons
8. CONCLUDING REMARKS

1. INTRODUCTION

One of the most successful improvements to the user–computer interface made within the last decade is the inclusion of icons—graphical symbols that visually represent information in the computer display. Icons can present a great deal of information concisely. However, overly dense visual displays can lead to a "cognitive overload" that adversely affects a user's decision-making performance (Ramsey & Atwood, 1980). Therefore, alternate channels of communication should be explored.

Sound is such an alternative communication channel. Sounds can be grouped or structured along principles similar to those of icons. We call such

structured sounds *earcons*,[1] which are defined as nonverbal audio messages used in the user–computer interface to provide information to the user about some computer object, operation, or interaction. Examples of computer objects are files, menus, and prompts. Editing, compiling, and executing are examples of operations. An example of an interaction between an object and an operation is editing a file.

Earcons, then, are the aural counterparts of icons. As one might expect, many icon design techniques also apply to earcon design, so it would be profitable to illuminate the principles of icon design and to draw appropriate parallels. However, there is little, if any, published material about aspects of icon design such as hierarchical and transformational organization, particularly in the formal or syntactic aspects of design. Therefore, a second goal is to examine icon types and construction methods, particularly as these methods apply to syntactic issues, that is, the formal arrangement of symbols. In addition, for earcons to be useful, they must be easy to comprehend and remember. For this reason, we delve briefly into learning theory to illustrate how listeners acquire and retain sound information that is to be associated with some external event.

In this article, we propose methods for constructing audio messages with meanings derived through modular, transformational, and hierarchical structures. The methods make use of *motives,* which are single pitches or rhythmicized sequences of pitches. These motives may be assembled into larger groupings called *families.* This approach is not meant to exclude alternative designs; the main utility of our scheme is that it is easy to understand, use, and modify. The guidelines are written for those who are not experts in sound or music theory. The guidelines lay out some simple rules and a limited range of choices. Once the basic choices and constraints are laid out, the nonexpert can follow this methodology to produce a useful set of earcons.

In the next section, we discuss icons as the visual counterpart of earcons. There is one striking difference in the applications of icons and earcons at the present time. Icons are both selectable and informational, whereas earcons are informational only. There are other differences that must be considered as well. Earcons are transient but, while they are played, they demand our attention. Icons may be, and usually are, presented simultaneously. The presentation of simultaneous earcons is more complex. Some examples of possible applications of audio cues are:

1. The arrival of a mail message.
2. The system going down in a few minutes.

[1] An unknown referee pointed out that "earcon" has been used in Buxton, Baecker, and Arnott (1985).

3. The echo of a mode of an operation.
4. The indication of a misspelled word in user input.
5. A sound equivalent of a file name or available tool.
6. A reminder that a file in temporary workspace has not been written to disk.
7. An out-of-bounds variable in a structured editor.

2. SOME WORK RELATED TO THE STUDY OF AUDIO INFORMATION

The development of user interfaces has been linked to the technology of cathode ray tubes. For this reason, sound input/output has been slow in developing; the use of sound has not been sufficiently explored. Voice is a primary means of communication for sound, just as text is a natural first choice for graphical presentations. Voice and text have many similar properties as a means of communicating information. Both are sequential, highly dependent on a natural or programming language, and generally require the user's full attention for comprehension. Hence, it is natural that voice became the first to be explored as a means of communication in the incorporation of sound in the user interface.

Some of the previous work in nonverbal audio communication in user–computer interfaces is in the area of data analysis. Sound can be used successfully for the analysis of multivariate, time-varying, and logarithmic data (Bly, 1982). Auditory displays of data have been used as an additional avenue of information transfer (Lunney et al., 1983; Mansur, Blattner, & Joy, 1985; Mezrich, Frysinger, & Slivjanovski, 1984). Audio alarms and signals of various types were with us long before there were computers. They are related to earcons as examples of audio messages. The characteristics and features of audio alarms in use as warning signals have been studied by Deatherage (1972) and Adams and Trucks (1976). In their studies, the alarms were horns, whistles, sirens, bells, buzzers, chimes, and oscillators with varying intensity and frequency. It has long been known that messages presented using voice seem to be assimilated with less effort than the same messages presented through visual media (Elliot, 1937; Sticht, 1969). Studies made by Buxton (personal communication, July 27, 1987) in video arcades show that scores drop when the sound accompanying video games is turned off.

Gaver (1986) studied auditory icons, a subject similar to the one we examine. We construct families of sounds by the use of compound audio elements, whereas Gaver examined auditory icons as caricatures of naturally occurring sounds. Gaver's work is discussed in Section 4.1. Portions of this material appeared in another study (Sumikawa, 1985) that presented guidelines for the use of earcons in the user–computer interface.

Art is to icons as music is to earcons. Even though earcons are not music, a knowledge of music is important in understanding the problems of designing audio messages. There is a great deal of literature relevant to the study of audio messages in the area of the psychology of music. A good selection of papers on this subject can be found in Deutsch (1982c). The *Computer Music Journal* has been a valuable resource for us. One of the recent significant advances in computer music is the development of the Musical Instrument Digital Interface (MIDI), a standard interface between computers and music synthesizers. MIDI is both a hardware standard and a software protocol. Now, sound interfaces of great musical complexity may be written to be played on a variety of computer and synthesizer systems (Loy, 1985; Moog, 1986).

3. ICONS: THE GRAPHICAL COUNTERPART OF EARCONS

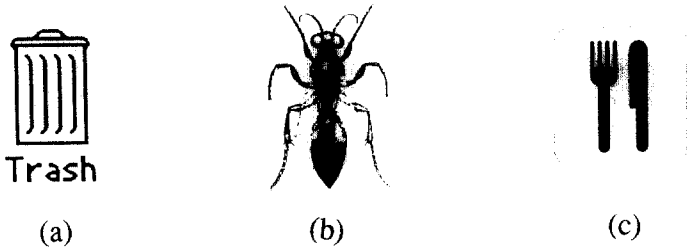
Valuable insight on the design of earcons can be obtained by investigating the wealth of information on their graphical counterpart, icons. As with any new area, the most effective applications of earcons are undoubtedly still unknown. At the present time, the use of earcons can best be understood through the similarities with icons. This is not to say that earcons are identical to icons in both composition and usage. We draw the analogy primarily to model a new concept. This section examines icons: what they are, the principles behind their structure and design, and how individuals learn and use them. We believe the organizational division of modular, hierarchical, and transformational icons is a new insight into their formal (not semantic) design.

3.1. The Use of Icons in Existing Systems

Early computer interfaces typically displayed information as text. This trend has changed considerably since the early 1970s as graphical symbols, known as *icons*,[2] have come to convey more information. Icons are versatile because they can represent a variety of computer entities in the user interface. Examples include system state information, utilities, processes, programs,

[2] The term *icon* in its ordinary English usage means a highly representational image. The term *visual symbol* is often used for abstract or semi-abstract symbols. In computer science the term *icon* has assumed the broader meaning that includes both representational images and visual symbols. *Representational icons* is not redundant in this terminology. Iconic representation is also used for musical images as well as for graphical images, although this use is less well known. One reason why we believe it is important not to distinguish between icons, in their traditional meaning, and visual symbols is that many computer interfaces intermix these types and a single name is required for both. The term *icon* has assumed this meaning among computer users.

Figure 1. **Representational icons are immediately recognizable pictures of objects. Icon (a) is from the Apple Macintosh computer, © Apple Computer, Inc., Cupertino, CA. Used with permission. Icon (b) is from *Graphic Communication* (p. 69) by W. J. Bowman, 1968, New York: John Wiley and Sons. Copyright © 1968 by John Wiley and Sons. Reprinted by permission. Icon (c) is from *Symbol Signs* (p. 160) by the AIGA staff, 1981, New York: Hastings House. Copyright 1981 by AIGA. Reprinted by permission.**



commands, cursors, menus, menu items, windows, screen selection buttons, and status objects such as files.

Iconic communication is successful because its imagery relies on the human ability to quickly perceive natural form and shape (Huggins & Entwisle, 1974). In addition, icons can represent much information in a small amount of space (Hemenway, 1982). Other advantages are less apparent. For example, research indicates that users spot, recognize, and process graphical images faster than they do words (Hemenway, 1982; Shneiderman, 1986). Users also make selections from iconic menus faster than they do word- or phrase-based menus (Marcus, 1984). Icons often have a universality about them that is not possessed by written language (Kolers, 1969). To facilitate and maximize the international exchange of computers, displays must be as language free as possible. Similar advantages hold for holistic information in general—earcons as well as icons.

3.2. Types of Icons

Our study showed that different types of icons are in use. Some are pictures of familiar objects, whereas others are stylized compositions of geometric shapes and marks. To simplify our investigation, we classify and distinguish between three different types of icons. Borrowing terminology describing sign types in general (Marcus, 1984), we label these icons as (a) *representational,* (b) *abstract,* and (c) *semi-abstract* (a combination of representational and abstract).

Representational icons are typically simple pictures of familiar objects or operations. Many of the icons found in the Star (Smith, Irby, Kimball, Verplank, & Harslem, 1982), Lisa (Williams, 1983), and Macintosh systems are representational icons (see Figure 1). However, representational icons do have their limitations. Many objects and operations do not have a familiar or obvious pictorial representation. Even those that do may be difficult to draw

Figure 2. Abstract icons are composed of geometric shapes and other nonrecognizable figures. *Note.* From "Corporate Identity for Iconic Interface Design: The Graphic Design Perspective" by A. Marcus, 1984, *Computer Graphics and Applications, 4*, p. 31. Copyright © 1984 by the Institute of Electrical and Electronics Engineers, Inc. Reprinted by permission.

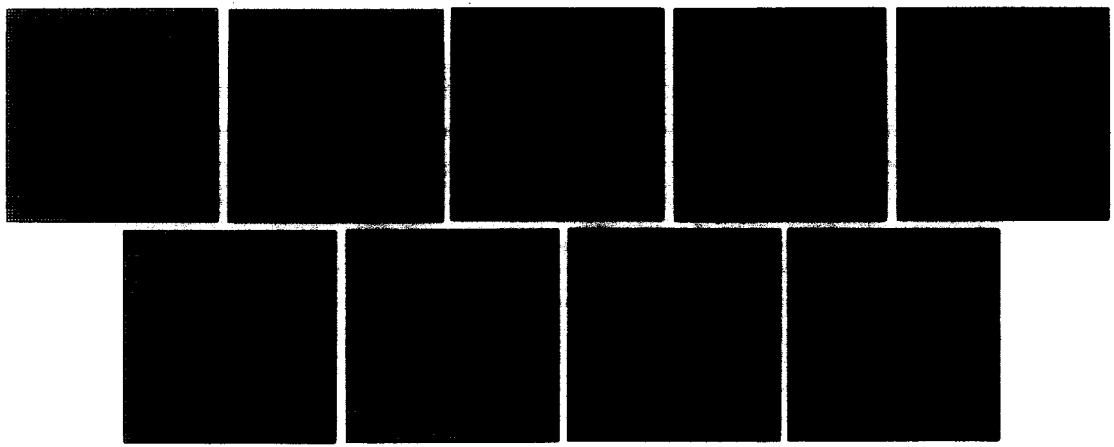
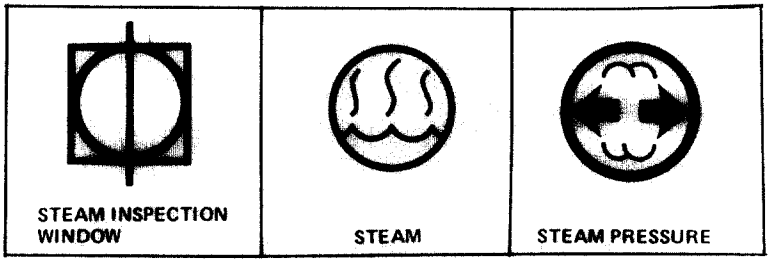


Exhibit 1025
Page 07 of 34

Figure 3. **Semi-abstract icons are combinations of abstract and representational icons. *Note.* From *Symbols Sourcebook* (p. 107) by H. Dreyfuss, 1972, New York: McGraw-Hill. Copyright 1972 by H. Dreyfuss, Inc. Reprinted by permission.**



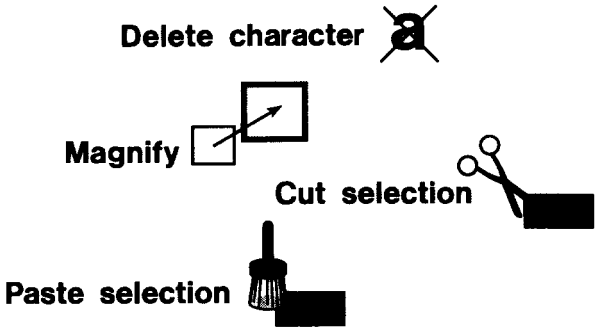
and implement, because some objects and operations are detailed or complicated. Not every implementer is an artist! Representational images have a more basic problem: Suppose we wish to convey the message "John is tall." A picture of John may also convey the information that John is wearing a checked shirt or that John has red hair (Fodor, 1981).

Abstract icons (Figure 2) are formed by combining geometric marks and shapes to depict a specific computer object or operation that is not easily or optimally depicted by a picture. Semi-abstract icons (Figure 3) may be either composed of both representational and abstract images or may be a representational image so simplified that it can only be considered an "abstracted" form. Most visual signs used in international communication fall into this category of highly abstracted representations (American Institute of Graphic Arts [AIGA], 1981). Arnheim (1969) believed that any picture, other than a photograph, must be considered partly as an abstraction. We do not dispute this statement, but we classify the symbols in Figure 1 as representational because of their resemblance to the object depicted.

Visual symbols of all types have been studied in the context of graphic art. A good deal of time and effort has gone into the development of an international set of symbol signs for public areas whose meaning can be grasped immediately and unambiguously (AIGA, 1981; Dreyfuss, 1972). Graphic designers are more concerned with the message communicated by the symbol rather than its classification as representational or abstract; it is often difficult to classify icons as clearly being of one type or another (AIGA, 1981). It is best to consider representational and abstract as the endpoints in a continuous scale of abstraction, but coarsely broken into the three classifications previously discussed.

The general structure of icons, regardless of type, can be broken into two components: elements and compounds (Kolers, 1969). An element is a graphical image, either representational or geometric, whose decomposition

Figure 4. **Icons can be related by combining elements.**



would result in uninterpretable parts. Countless pictographic images, geometric shapes, and marks can be used as elements. A compound is any combination of these elements.

3.3. Iconic Families

Many computer entities representable by icons have a variety of features in common. Computer commands, for example, are highly structured sets, in which many commands share features (Hemenway, 1982). We can create icons sharing common features to represent relationships between entities. Therefore, similarly shaped elements represent similar classes of information (Kolers, 1969). The advantages of shared elements are that users have less to learn and that they may make predictions about important characteristics of an unfamiliar command.

We have identified at least three ways of creating compound icons: (a) combining, (b) transforming, and (c) inheriting. Figure 4 shows icons produced by combining symbols. In this scheme, some elements represent objects and others actions. The icon that shares an object and action represents the performance of an action upon an object (Hemenway, 1982).

Compound icons created through transformation often represent objects with an associated state. A transformation is depicted in Figure 5, which shows a symbolic representation of heating elements on a stove. Korfhage and Korfhage (1984) described methods of creating compound objects in pictolinguistic languages. Our own examples, taken from Japanese, are shown in Figure 6. Here we have an example of compound objects with inheritance. The symbol for "fire" is the root of the transformation and when combined with the symbol for "flood" represents "disaster." "Ashes," "inflammation," and "pale" are also created in this manner. The hierarchy, in this case, is not generic in meaning; ideas move from simple to complex. The

Figure 5. Icons can be related with transformations. *Note.* From *Symbols Sourcebook* (p. 102) by H. Dreyfuss, 1972, New York: McGraw-Hill. Copyright 1972 by H. Dreyfuss, Inc. Reprinted by permission.

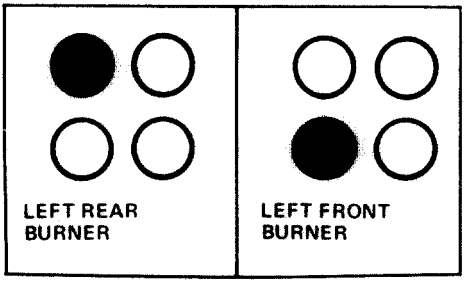
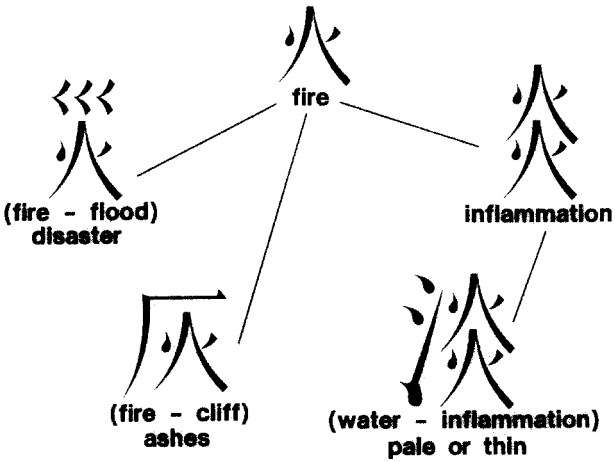


Figure 6. Icons can be related by inheriting elements from other icons.



scheme we suggest for earcons will be hierarchical in both syntax and semantics. In all three methods—combining, transforming, and inheriting—we use symbols to define a class or family of related objects.

3.4. The Design of Icons

The concept of "figural goodness" in Gestalt psychology helps to determine the general characteristics of a "good" icon. Easterby (1970) used figural goodness to describe five characteristics possessed by a good icon: closure, continuity, symmetry, simplicity, and unity. These guidelines apply to musical figures as well (Davies, 1978; Deutsch, 1982a; Meyer, 1967). Gestalt psychologists proposed that stimuli are grouped into configurations on the

basis of various simple principles, which Deutsch characterized as proximity, similarity, and continuity.

Principles of good graphic design also apply to icon design. Marcus (1984) described graphic design principles in his concept of "corporate graphics":

> The concept of corporate graphics implies that all images are designed to meet unique communication needs, while being adjusted to produce a visual consistency with the system. This combined approach can be achieved by the use of a constant scale, limited size variations, the orientation of figures with respect to text, limited use of colors, limited variations in line weights, and the treatment of borders for figure or pictograms. These visual themes help to establish recognizability, clarity, and consistency. (p. 26)

The aforementioned principles provide valuable information on the general design of icons, but they are somewhat vague for the actual design process. Easterby (1970) advocated an experimental design approach in which the designer first defines a prototype for each icon in the interface while considering the general principles, the function to be represented, and the notions developed from research on pattern perception and discrimination. The designer then does human-factor tests on the intended user population. Finally, the designer manipulates these prototypes according to the results obtained in the human-factor experiments. The Easterby approach to icon design would work equally well with earcons. The AIGA has used a similar approach to evaluate passenger/pedestrian oriented symbols for the U.S. Department of Transportation (AIGA, 1981). Evaluations were made in two ways: through a private rating of each symbol on a symbol concept evaluation sheet and through committee discussions. Each symbol was rated on a scale from 1 to 5 on its semantic, syntactic, and pragmatic dimensions.

4. EARCON DESIGN PRINCIPLES

Even though they utilize different senses, icons and earcons involve similar communication needs and design problems. Knowledge obtained by examining theories of icon design can be used to model equally successful approaches for earcons. Just as in icon design, we divide earcons into three classes: representational, abstract, and semi-abstract.

4.1. Representational Earcons

Many different schemes for designing earcons are possible. One such scheme might involve digitizing natural sounds from the surrounding

environment. The first inclination with icons is to use highly representational images of existing objects. In the same way, natural sounds could be used as audio messages. We denote these representational earcons. The advantages and disadvantages of representational earcons are similar to those of pictorial icons. Gaver (1986) investigated representational earcons, which he called *auditory icons*. His auditory icons are caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes. He divided the mappings between data and their auditory representation into three different types: symbolic, nomic, and metaphorical. *Symbolic* mappings rely on social convention for meaning, such as applause for approval; *nomic* (the most concrete) representations are physical, such as the sound of a closing metal cabinet for closing a file; and *metaphorical* mappings that are similarities, such as falling pitch for falling object. Gaver noted that auditory icons need not be realistic representations of the objects they portray but should capture their essential features, as do representational icons.

4.2. Abstract Earcons

Marcus described a set of "elements"—geometric shapes and marks used as building blocks of sets of icons. The power behind his method is the very existence of these well-defined elements, which can be put together in various ways to build large and highly customized sets of icons. In the same way, the earcon design approach advocated here used single pitches or groups of pitches as the elements or building blocks of earcons. Motives, described more fully in the following section, are sequences of pitches that create a short, distinctive audio pattern often characterized by the simplicity of its rhythm and pitch design (Christ, DeLone, Kliewer, Rowell, & Thomson, 1966). Their very brevity and distinctive manner make motives very powerful tools for composing earcons. Compound earcons composed of motives or single pitches can be used to express complex objects. Of the many possible design approaches, ours has the following advantages:

1. It is systematic, with well-defined building blocks that may be used to construct larger sets of earcons. A systematic approach is usually straightforward to understand and use. Modularity allows easy modification, future expansion, and tailorability.
2. Motives may be transformed, combined, or inherited, thereby creating families of related motives.
3. Motives may be grouped into families, where earcons with similar meanings have similar sounds; different families will sound dissimilar.
4. Our approach is easily implementable on most contemporary computer systems, large or small. Most computers today have the necessary

equipment to perform rhythm and simple pitch. This approach makes investments in expensive sound boards or other specialized equipment unnecessary.

Regardless of which approach is taken for the design and implementation of audio messages in a sight-and-sound user interface, including an expert in sound relationships in the design team to determine initial constants of a sight/sound user–computer interface is of utmost importance. Ideally, the expert should be a professional composer, preferably one with some experience in electronic (synthesized) music. The science of sound is a highly technical, diverse, and complicated discipline. An expert in this field understands the existence, importance, implications, and consequences of the many perceptual problems of sound/language structures. The guidelines, as presented here, are written for those who are not experts in sound. Experts may wish to use our syntactic methods, while embellishing and varying sound relationships.

5. MOTIVES

A motive is a brief succession of pitches arranged to produce a rhythmic and tonal pattern sufficiently distinct to allow it to function as an individual, recognizable entity (Bernstein & Picker, 1966). We use these larger structures for earcons. We define a *motive,* then, as a rhythmicized sequence of pitches. Rhythm and pitch are the *fixed parameters* of motives, and timbre, register, and dynamics are the *variable parameters* of motives. The eloquence of motives lies in their ability to be combined to create larger recognizable structures. The repetition of motives, either exact or varied, or the linking of several different motives produces larger, more self-sufficient patterns. Note, however, that an earcon is any audio message and need not be a motive. The material of rhythm, pitch, timbre, register, and dynamics may be found in textbooks on music appreciation, such as Christ et al. (1966), Bernstein and Picker (1966), and Kerman (1980).

5.1. Rhythm

Rhythm is the most prominent characteristic of a motive. Listeners respond more readily to rhythm than to any other musical parameter. Kerman (1980) stated that "No single feature of music, not even melody, determines the effect of music more crucially than the timing and weighting of notes" (p. 5). In music, duration is indicated with note values. Common time divisions and their iconic notations are listed in Figure 7. A whole note is generally given the longest time value. A half note is half as long as a whole note, a quarter note

Figure 7. **Musical notations that indicate note duration.**

Note Type	Notation
1/16	𝅘𝅥𝅯
1/8	𝅘𝅥𝅮
Dotted 1/8	𝅘𝅥𝅮.
1/4	𝅘𝅥
Dotted 1/4	𝅘𝅥.
1/2	𝅗𝅥
Dotted 1/2	𝅗𝅥.
Whole	𝅝

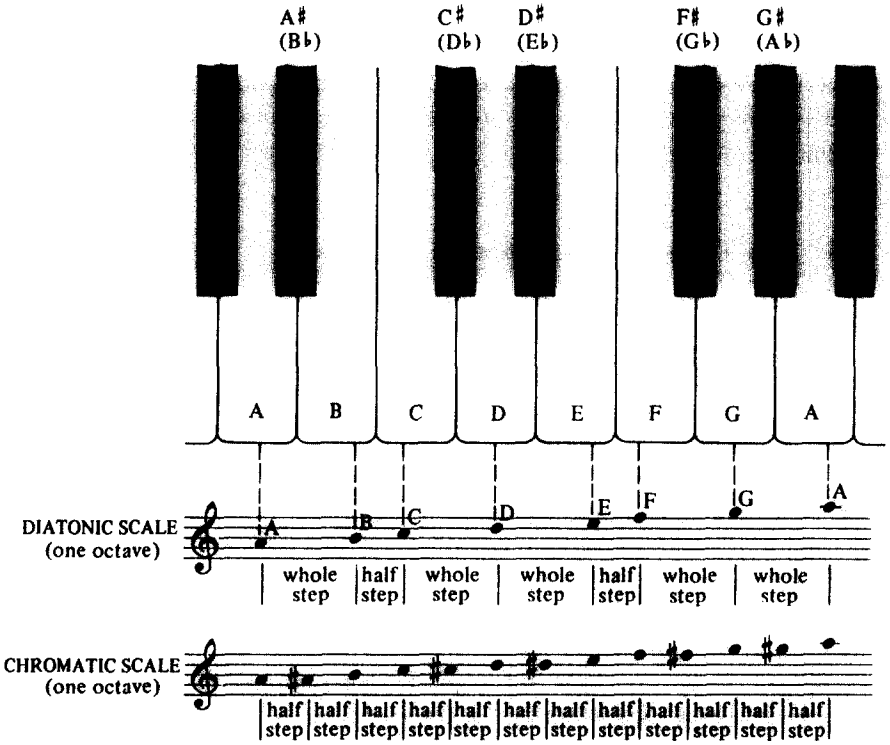
is half again as long, and so on. Short silences (called *rests*) as well as timed sounds can be used to create rhythms for earcons. The easiest way to produce two distinct motives is to vary the rhythmic structure of each. The same sequence of pitches can be made to sound different when different rhythmic structures are imposed on it.

5.2. Pitch

The pitch sequence of motives can take a variety of forms because so many pitches are available. Most computer users are familiar with the tonal system of eight octaves of 12 pitches each. In theory then, 96 unique one-pitch earcons, 9,216 two-pitch motives, 884,736 three-pitch motives, and 84,934,656 four-pitch motives are possible. In reality, however, it is difficult to distinguish the fundamental frequency of very low and very high pitches. This, along with the difficulty of reproducing such extreme frequencies, eliminates these outer octaves from consideration. In addition, motives constructed of random pitches from these 96 available are undesirable. The pitches chosen for a particular motive are important.

All pitches in a motive should be chosen from the same octave to avoid problems inherent in octave perception. This also facilitates changing the register of the pitch sequence (Section 5.3).

Figure 8. **The pitches that make up an octave.** ***Note.*** **From** ***Listen*** **(p. 18) by J. Kerman, 1980, New York: Worth Publishers, Inc. Copyright 1983 by Joseph Kerman. Reprinted by permission.**



Careful attention must be paid to the harmonic references and implications created by the sequence of pitches in an earcon. By controlling the harmonic relationships implied by certain pitch sequences, one avoids creating unwanted expectations on the part of users and avoids a subsequent need for tonal resolution.

To understand the following important information and guidelines on motive construction, a short digression into musical theory is necessary.

In Western tonal music, the octave (see Figure 8) is divided into 12 pitches, each equidistant from its neighbor. These 12 pitches collectively constitute the chromatic scale. The distance between any two adjacent pitches in the chromatic scale is called a half step. For example, C–C# and B–C are each a half step apart. Two half steps equal one whole step. C–D and F–G are each a whole step apart. The existence of half steps in a sequence of pitches can create musical tension, which in certain contexts should be avoided for earcon construction, unless, of course, this tension is desirable for its semantic and

pragmatic effects. The novice user will have the least difficulty if he uses preexisting pitch structures known as the *diatonic major/minor scales.* Major and minor scales are seven-note scales, the adjacent notes of which are variously half steps and whole steps apart. We also refer to *intervals,* the distance between two pitches as measured by the number of diatonic scale notes between them. Because of cultural familiarity with major and minor modes, motives using notes drawn from a single major or minor scale will be easily recognized and understood. (Learning and memory are discussed in Section 7.)

5.3. The Variable Parameters of Motives

We identify motives by rhythm and pitch. In some sense, a motive played with a different timbre is still the "same" motive. Think of some familiar tune such as "Twinkle, Twinkle Little Star." It will remain the same tune even if the octave varies or we change the instrument on which it is played. Dowling (1982) studied the identification of melodies with varying attributes. As an earcon, however, the meaning associated with a motive may change with a change in timbre, dynamics, or register.

Timbre

Timbre is a tonal attribute that refers to the quality or "color" of a sound. Musical instruments sound different because the physical makeup of their sound differs. For example, the same note played on a piano and clarinet sounds different because each instrument has its own peculiar set of waveforms and harmonics. The timbre of a sound is usually described with adjectives such as *bright, warm, harsh, hollow, twangy,* or *brassy.* Even though timbre is difficult to describe and notate precisely, it is one of the most immediate and easily recognizable characteristics of sound (Kerman, 1980).

Earcons assembled by novice users should be made from a set of basic, easily differentiated waveforms. The sinusoidal waveform will be "timbreless" because it is pure, unmodulated sound. The sine wave lacks timbre in the same sense that white lacks color. To alter our perception of a given earcon, we can simply change the earcon's waveform, thereby producing a different timbre.

Register

Register refers to the relative high/low of a pitch or a set of pitches. An earcon in the low register sounds "low"; one in the high register sounds "high." If we label the eight octaves in our tonal system 1 to 8, from low to high respectively, we can define an earcon in a low register, Octave 2, for example, as one with pitches entirely contained in one of the lower octaves. (As explained previously, an earcon in the lowest octave would be difficult to

Figure 9. **An earcon with dynamics that change from soft to loud.**

(soft ——→ loud)

distinguish and reproduce.) Likewise, a motive with medium register has pitches from Octaves 4 or 5, and one with high register has pitches from Octave 7. Register can be creatively manipulated to indicate vertical location or direction in a sight/sound interface. Earcons constructed with pitch sequences contained in the low, medium, or high registers are easily differentiated and, therefore, easily perceived. An earcon can be made to sound very different by changing its register. In Section 5.2, "Pitch," we suggest that an earcon be composed of pitches from a single octave because changing the register of that earcon, thereby producing a new sound, is very easy.

Dynamics

Dynamics refers to the relative loudness or softness of an earcon. One can differentiate the relative loudness or softness of an earcon by making its dynamics either constant or variable. Thus, an earcon can be loud for the duration of the motive, soft for the duration of the motive, graded from loud to soft, and graded from soft to loud. Any combination of these four is also possible.

Dynamics can be creatively used in an earcon to indicate direction from left to right. For example, the three-note earcon shown in Figure 9, whose dynamics changes from soft to loud, could be used to indicate a direction or movement, as in the nomic mappings suggested by Gaver (1986).

The musical term describing this feature of dynamics is the *crescendo.* Crescendo means "to grow" in Italian and represents a sound (or sounds) increasing from soft to loud. The opposite, a *decrescendo,* means "to get smaller" and describes from loud to soft. The musical notation for a crescendo is the wedge-shaped symbol shown under the earcon in Figure 9. A decrescendo is the same symbol reversed. Both crescendo and decrescendo can be used in motives to indicate location, for example, of a window on the computer screen. Also, an earcon representing the operations of enlarging, zooming, and shrinking could be constructed from motives in which the dynamics have been creatively varied.

5.4. Length of Earcons

The length of an earcon should be sufficient to convey the message effectively, but no longer. The earcon's meaning must come across as quickly

and as easily as possible since sound is transient. The optimal number of pitches in a motive is two to four: "A motive is as short as two or three notes—just long enough so that its rhythmic and/or melodic character is easily recognized and easily remembered when it comes back again" (Kerman, 1980, p. 22). We suggest an upper limit of four pitches for several reasons. Sound takes place in time, so unlike images, a listener cannot dwell on a sequence of pitches (Miller, 1956). Therefore, motives should be kept as short as possible so listeners can make the necessary audio connections between pitches. However, a long succession of pitches tends to create a melodic pattern. Motives consisting of more than four pitches can have undesirable melodic implications. As soon as an earcon sounds like a tune, the listener will associate that earcon with music. Hearing a simple tune 10 or more times a day potentially irritates users and could cause audio fatigue.[3]

6. EARCON CONSTRUCTION

Just as icons consist of different structures that coexist in the same user interface, so do earcons. Combining one or more audio elements to create an earcon leads to two basic types of structures: (a) earcons composed of only one element and (b) earcons composed of more than one element (compound earcons).

6.1. One-Element Earcons

One-element earcons may be digitized sounds, a sound created by a synthesizer, a single note, or a motive. An element may be compared to a word, whereas a note may be compared to a letter of the alphabet. Elements need not be purely of one type, for example, Kana, the Japanese written language. Kanji characters (Chinese pictograms) are used with two different phonetic alphabets, Katakana and Hiragana. Similarly, in our sound language we can use a variety of elements to create the basic sounds. We may even include voice in this collection; however, with the inclusion of voice the theory grows complex. We recommend that the user form audio messages from one simple scheme. Single-pitch and single-motive earcons are both composed of one element.

A *single-pitch earcon* is any audio message composed of one note with the attributes of pitch, duration, and dynamics. Because single-pitch earcons are

[3] Communicated by our expert in sound relationships, Robert M. Greenberg. This view is also supported by Davies (1978), who pointed out that studies show people dislike tunes that are low in complexity. A tune, already simple in structure, would become even more so through familiarity.

simplest, they can be used to represent simple, basic, or commonly occurring user-interface entities such as key clicks, cursors, or selection mechanisms.

A single pitch earcon that decreases in volume can be used as an audio cue to indicate that a computer is going down. Here, the single, decreasing-volume pitch becomes a metaphor for the application it depicts (Gaver, 1986). Another example of a single-pitch earcon is a key click representing an editing mode, say the delete mode. The user hears the click each time a character is deleted.

A *single-motive earcon* has the attributes of rhythm, pitch, timbre, register, and dynamics. Because single-motive earcons are relatively simple, they can represent basic, common computer entities such as certain error messages, system information, windows, and files.

A single-motive earcon might represent some system message in one of several windows displayed simultaneously on a screen. The earcon's register and dynamics can indicate the location of the window displaying the system message. Low register and soft dynamics, for example, might indicate that a system message is being displayed in the lower left window.

6.2. Compound Earcons

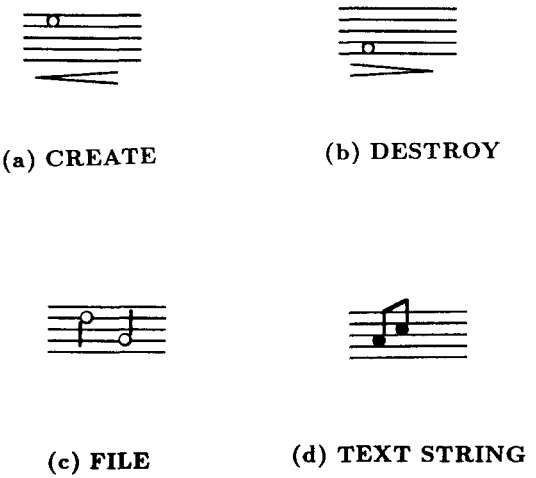
The three construction principles for compound earcons are combining, inheriting, and transforming. These three methods are used with both representational and abstract elements.

Combined earcons are formed by placing two or more audio elements in succession. The combined icons in Section 3.2 were object/operation icons. As with combined icons, combined earcons can represent computer entities sharing common features and are constructed by using similar audio elements to represent similar classes of information. We assume that a set of audio elements (one element for each feature shared by two or more system entities) has been constructed, and thus these elements can be combined into different earcons depicting various system functions. The elements of this set, called the *primary set,* are constructed according to the guidelines of Section 6.1.

We use the elements of the primary set to construct a unique earcon that represents a system function or computer entity. We can then systematically build up a large set of earcons for the entire system. Repeating of audio elements gives us the same advantages for earcons as repeating visual elements did for icons: ease of construction, set expansion, and ease of user identification and retention. Combining elements is a simple method of creating both icons and earcons because it requires little understanding of hierarchies or other abstractions.

As an example, we define a set of audio elements representing the computer

Figure 10. **The four audio elements: CREATE, DESTROY, FILE, and TEXT STRING.**



entities CREATE, represented by one crescendo pitch; DESTROY, by a different decrescendo pitch; FILE, by a two-note motive with a long, long rhythm; and TEXT STRING, by a two-note motive with the short, short rhythm (Figure 10).

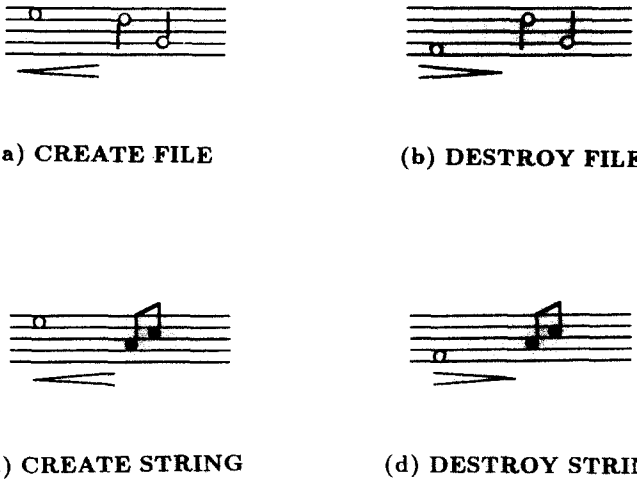
Using these audio elements, we can systematically construct earcons for the computer entities CREATE FILE (Figure 11a), DESTROY FILE (Figure 11b), CREATE TEXT STRING (Figure 11c), and DESTROY TEXT STRING (Figure 11d), by combining the audio elements for the common features.

6.3. Inherited Earcons

An earcon in a family of earcons related by inherited or transformed attributes is a *family earcon.* First we examine hierarchical families. Each earcon is like a node in a tree of earcons, where an earcon on the nth level is a sequence of n elements. The first n-1 elements are the preceding n-1 elements of that node. Figure 12 shows these sequences. Each level modifies an attribute of motives on that level. Because we have only five attributes (rhythm, pitch, timbre, dynamics, and register), only five levels are possible. Following are the rules for forming hierarchical families of earcons:

1. The hierarchy of messages pertaining to computer entities must be clustered into families, that is, the entity families and their descendents to be represented by audio messages must be identified. Examples of such families are the ERROR MESSAGE, WINDOW, PROMPT, and EDITOR. In this article, our examples of hierarchies are generic, so "error" is generic for "operating system error" in Figure 12.

Figure 11. **Combining audio elements:** CREATE FILE, DESTROY FILE, CREATE STRING, **and** DESTROY STRING.



2. Each family is assigned a different and distinct rhythmic structure. This rhythm is the level one or *family rhythm,* the audio version of a human family's last name. Family rhythms have no pitch structure; they are purely rhythmic sequences. Clicks or other sounds can articulate characteristic family rhythms. The family rhythm is technically not a motive because it is pitchless. Each entity at the highest level in a family hierarchy is represented by an earcon composed solely of the rhythmic structure that identifies the family.
3. All entities residing at the second highest level are represented by a two-part earcon, the first being the family rhythm and the second being a sequence of pitches rhythmicized by the family rhythm. This second part is a level-two motive. We suggest that all level-two motives be composed on a sine wave. Because sine waves are the most "colorless" waveforms, their use helps minimize any difficulty in differentiating between the timbre of level-two and level-three motives.
4. All entities residing at the third level of the hierarchy are represented by a three-part earcon. The first and second parts belong to the two higher levels, namely, the family rhythm followed by a level-two motive. The third part is a duplication of the level-two motive, but it is assigned a different timbre and a slightly higher pitch. This third element is called a level-three motive.
5. If more than three levels exist in the hierarchy, one can represent additional levels by varying an earcon's register and dynamics. For example, level-four entities could be depicted by the three-part earcon from level three, but with different register or dynamics.

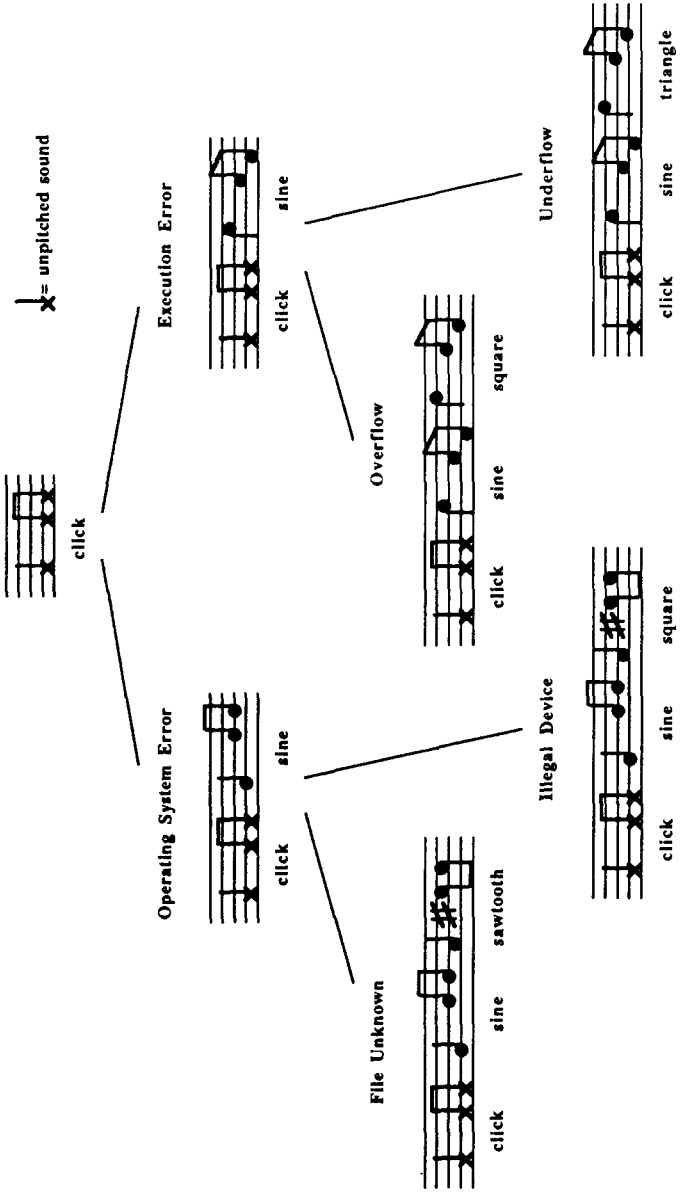


Figure 12. **The earcons contain all the elements in the hierarchy for expert users.**

Exhibit 1025
Page 22 of 34

Therefore, in three-part earcons, the first part (family rhythm) identifies the entity family (error, system information, menu, etc.), the second part identifies the type of error (operating system, execution, syntax, compilation, etc.), and the third part identifies the exact error message. If only three levels are used for family errors, say, rhythm, pitch and timbre, then the other two attributes, dynamics and register, may be varied throughout the levels. An audio message could accompany the textual error message appearing on the screen or have a voice message follow it. At the end of Section 6.4, we describe a shortened form of the family earcons just described.

A variety of different hierarchical schemes could be used. For reasons given in Section 7, we believe the hierarchical structure described here has many advantages. Another approach, however, would be to use sequences where the elements are not as closely related as in family earcons. Using the Japanese pictograms as an example, we might create sequences of representational sounds to form more complex objects. Using the example given in Figure 6, we could construct a disaster sound by following the sound of fire with the sound of water. Note that these are not combined earcons in the sense of Section 6.2.1. Here the two objects take on a more complex meaning when concatenated.

6.4. Transformed Earcons

A transformation of a musical segment is a modification of that segment by a change in its attributes. Transformations that result in perceptual equivalences that retain the identities of musical objects have been studied intensely by those interested in the psychology of music (Deutsch, 1982b). Researchers have proposed that melodies are similar to visual shapes (Deese & Grindley, 1947). The word *contour* describes musical shape. Preserving the contour of a sequence of notes seems to preserve the identification of the sequence, even if the interval size is changed (Dowling, 1982). Among the most interesting musical transformations were those used by the composer Arnold Schoenberg. Schoenberg (1951) proposed that a transformed row of tones may be recognized as equivalent to the original when all ascending intervals become descending and vice versa (inversion), when presented in retrograde order (retrogression), or when transformed by both these operations (retrograde-inversion). His transformations have particular appeal to scientists because they are reflections of the original tone row across an x-axis (inversion), a y-axis (retrogression), and both (retrograde-inversion). However, Deutsch (1982b) warned that "whether such transformations indeed result in perceptual equivalences is debatable" (p. 283).

We suggest that transformed earcons be modified in simple ways that clearly retain perceptual equivalences. Simple changes in timbre, dynamics,

and register pose no perceptual difficulties. Pitch changes change the contour of the earcon and should be administered with care (Deutsch, 1982b; Dowling, 1982). A set of transformations that may be used for earcons is described here.

The inherited family earcons, described in Section 6.3, may be thought of as a tree of motives (see Figure 12). The leaves of the tree contain the entire message, represented by a path from the root (family rhythm) to the leaf. Hence, a family earcon is composed of one or more similar motives and the family rhythm. Each level encompasses all information from the level above it. A multilevel earcon of this type provides novice users with repetitive, familiar information they may need to identify somewhat unfamiliar information. Intermediate level and expert users may not require repetition of motives describing the path through the trees. Therefore, earcons designed for expert computer users could consist of the last motive in an n-level motive (n is the number of levels). Intermediate users might find earcons consisting of the last two parts of an n-level motive useful. Continuing with our previous example, Figure 13 shows how the earcons of Figure 12 can be shortened for expert users. Likewise, Figure 14 shows the same motives reconstructed for intermediate-level users.

The leaves of the tree shown in Figure 13 are related by transformations of pitch, timbre, dynamics, and register. Rhythm is the one attribute that remains constant throughout the family of motives. Figure 14 is a mixture of hierarchical and transformational types.

7. LEARNING, REMEMBERING, AND UNDERSTANDING EARCONS

For an earcon to be useful, listeners must learn and remember it. Learning and remembering represent the process by which listeners acquire and retain sound information that is to be associated with some external event. Gaver (1986) said that "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" (p. 173). Not all computer entities can be captured by sound when representational earcons are used, so some mappings must be symbolic or metaphorical. Representational earcons do not solve our problems with learning, remembering, and understanding the structure and meaning of earcons. Also, a large number of representational earcons would create the problem of memorizing each as a distinct entity. This problem is shared with icons. Fodor (1981) pointed out that an icon cannot express the meanings of sentences. Instead, we must not only identify the object represented by an icon, but we must understand its mapping to the world. That is, we must understand the meaning and function of the icon.

Figure 13. **The earcons need not contain all the elements in the hierarchy for expert users.**

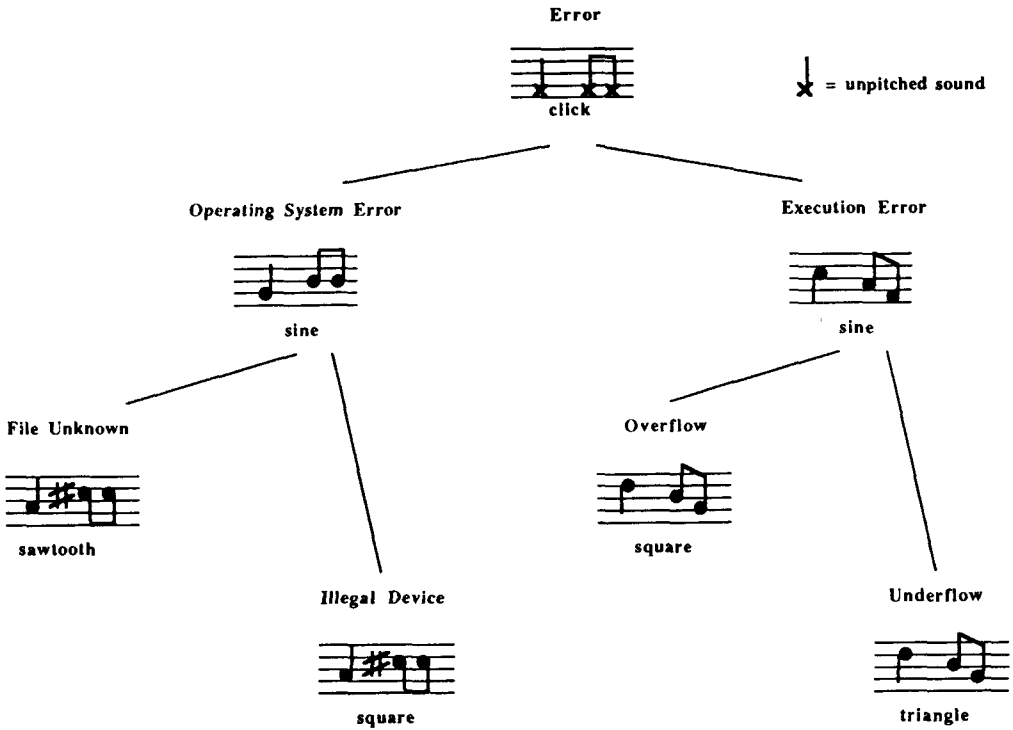
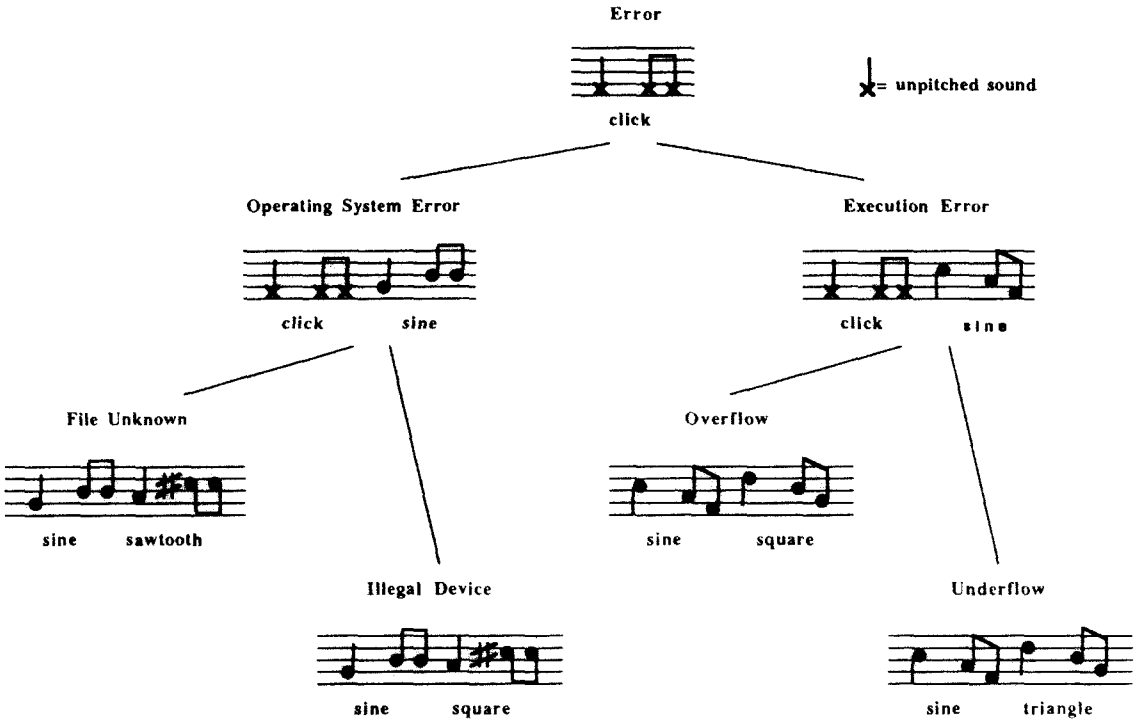


Figure 14. The earcons for intermediate users contain some of the elements in the hierarchy.



In the following sections, we discuss theory of learning and memory, specifically with respect to abstract earcons. The reason is that so much material is available from work done in the psychology of music and pure sound. A resolution of the difficulty of learning representational versus abstract earcons would depend on many factors such as the size of the systems, the nature of the messages, and so on. If we consider similar questions that have been raised for visual images, the semi-abstract seem to have satisfied the needs of an international sign language the greatest (AIGA, 1981; Dreyfuss, 1972). Here we examine some arguments for exploiting a good syntax for earcons and provide some justification of the construction scheme outlined previously.

7.1. Learning and Remembering Earcons

Section 5.2 recommended the Western tonal scales for motives. The question is always raised: With so many interesting sounds to choose from, why use the Western tonal scale? Tonal sequences appear to be easier for Western listeners. (This is likely to be true for the same reasons Gaver, 1986, gave, namely, a lifelong experience of listening to tonal music.) The pitches and intervals of atonal melodies appear especially hard to learn. When they are remembered, it is in a relatively specific way and not easily generalized to a new context; performance is even better if the sequences are familiar (Dowling, 1982). Attneave and Olson (1971) and Bartlett and Dowling (1980) showed that persons with little or no musical training could reproduce or recognize familiar melodies. Dowling (1982) also reported that musical training enhances the importance of the tonal scale system in information processing of melodies. However, the intervals of the scale system are firmly embedded in the minds of even untrained listeners. Deutsch (1982b) and Dowling (1982) both reported that distorted tunes can be difficult to recognize.

A mental structure for storing audio information facilitates memory recall. Davies (1978) agreed that listeners have an internal representation of the organized perceptual units composing an earcon. When learning an earcon, the listener codes the exact structure of this internal representation or template. Listeners then use the internal representation and compare it to the existing templates of known earcons to remember and recognize incoming earcons. If the sequence of sounds is sufficiently similar to an existing template, then the listener recognizes the sequence as an already learned earcon. If the sequence does not match any existing template, then it is unrecognized and can be considered as new. To remember this new earcon, listeners must go through the learning process. This evidence points to the superiority of constructing basic templates for earcons, such as templates for

the family earcons we have described and varying the structure and meaning with small, easily recognized differences.

The ability of the brain to group sound into its organizational units facilitates a listener's ability to remember sound information. Miller (1956) described the process of memorization as the internal formation of chunks or related tones. The process continues until there are few enough chunks so that we can recall all the tones in an earcon. A similar belief is held in Gestalt psychology (Davies, 1978), where researchers conjecture that people perceive general patterns in strings of separate tonal events, thereby organizing sound into perceptual units. Earcons composed of related and organized groups of sound elements are easier to recognize and remember than those composed of unrelated sequences of tones (Davies, 1978). Chunks of 5 to 6 unrelated tones or 10 to 12 related tones may be remembered without difficulty (Lundin, 1953). Earcons composed of a simple repeated configuration involving only three different note values are easiest to remember (Lundin, 1953).

7.2. Classes, Types, and Hierarchies

Rosner and Meyer (1982) studied musical archtypes and concluded that music is understood in terms of types and classes. Meyer (1967) explained musical hierarchies as follows:

> Complex musical works, like complex physical or biological structures, tend to be hierarchic. Low level events, made up of one or two "entities" (pitches), combine to form larger structural units; these in turn unite in various ways and produce still more extensive and complex organizations; and so on, until the highest level, that of the whole entity or the complete musical work, is reached. (p. 258)

This statement is rather remarkable because it shows that music and computer science share some similar "data structures." Deutsch (1982b) believed that music is indeed like other types of information in that it is organized by classes and hierarchies:

> The tonal music of our tradition is also composed of small segments that are systematically organized in hierarchical fashion. It is reasonable to suppose that such hierarchical organization reflects the ways in which musical information is abstracted and retained. (p. 288)

Deutsch also pointed out that we retain hierarchies of rules, syntactic hierarchies in language, and hierarchies of visual scenes. The musical hierarchies discussed by Deutsch are somewhat different than the hierarchical

earcons described in Section 7.3. Deutsch stated that, "The most important development in this field is that of Heinrich Schenker (1868–1935), who proposed a hierarchical system for tonal music that has points of similarity with the system proposed by Chomsky (Chomsky, 1963) for linguistics" (p. 291).

7.3. Space Complexity and Memory

Space complexity measures the amount of human memory required to retain an earcon. The goal is to minimize the space complexity for each audio message in the user–computer interface. The less a user must remember, the easier it is for him or her to later recall and identify audio messages. Donald Knuth (1984) humorously examined the complexity of songs and used the words of songs as the units of his complexity measures. Before translating his results to sequences of notes rather than sequences of words, we examine some of his song complexities.

A song of length n words without any type of repetitive structure has space complexity n. This complexity can be reduced with a refrain. A refrain imposes an organization or relationship on the entire song, thereby eliminating the amount of memory necessary for successful recall. Knuth showed that when a song has a refrain, its space complexity can be reduced to cn, where $c < 1$. Songs with space complexity $O(\sqrt{n})$ are not uncommon and use the trick of repeating both refrains and verses. "Old MacDonald" has an even lower complexity, accomplished by substituting the names of animals and animal sounds into similar verses as well as repeating the refrain and previous verses. "Partridge in a Pear Tree" has a still lower complexity of $O[\sqrt{(n/\log n)}]$, whereas "m Bottles of Beer on the Wall" has complexity of $O(\log m)$.

An earcon is designed for maximal information content only. The space complexity of earcons is, perhaps, more important than for many of the songs because earcons convey only information. An earcon should convey the necessary information while requiring the user to remember as little as possible, yet the result should not be unpleasant nor fatiguing. In other words, an earcon should have low space complexity. Earcons with low space complexity are easier for the user to learn and remember.

Consider the space complexity of an earcon as the number of elements that must be remembered in the earcon. Each note may be thought of as a five-tuple of rhythm, pitch, timbre, dynamics, and register. An earcon of length n notes with no inherent relationship or organizational structure has space complexity $5n$ or $O(n)$. In a three-motive family earcon, for example, the second motive is related to the first by a change in pitch, and the third motive is related by a change in timbre. A user has less to remember because the

second and third motives are only slight variations of the first. The space complexity of family earcons is the complexity of the family motive plus changes in pitch, timbre, register, and dynamics, which are each of complexity one. Each offspring motive in a family earcon requires remembering only one element more than its parent motive. The aspect of memory required for transformational or hierarchical systems of earcons is $O(\log n)$, where n is the number of earcons.

Compound earcons greatly reduce space complexity because they use common audio elements to represent similar features of different computer entities. A user has less than the full set of earcons to remember because a particular common feature is always represented by the same audio element. In the example given in Section 6.2, a user would only have to remember the four audio messages for create, destroy, file, and string. From these the user could easily recognize and identify the four earcons for create file, create string, destroy file, and destroy string. Hence, the growth complexity for object/operation earcons is $O(n + m)$, where n is the number of objects and m the number of operations, and the number of earcons is nm.

7.4. A Summary of Learning and Memory Issues for Earcons

From the previous material we can summarize the advantages and disadvantages of the various methods used to create earcons as follows:

Representational Earcons (Not Compounded). These are useful for systems that require few earcons. Users can quickly associate an earcon with its representation, but must still remember a mapping. Well-chosen representations can be easily discriminated. For n earcons we must remember n representations and mappings.

Combined Earcons. These may be combined abstract or representational earcons, and fall between the representational (uncompounded) and the family earcons in difficulty. Combined earcons are useful for intermediate size systems of earcons. We speculate that combined earcons are initially easier to learn than family earcons. If representational earcons are used for objects and operations, more complex messages may be conveyed than through compounded representational earcons. For n earcons, where $pq = n$, and p and q are unrelated elements, $p + q$ representations and mappings must be learned.

Family Earcons (Hierarchical and Transformational). Modular elements such as motives must be used for inherited or transformed earcons. This means that abstract earcons would most likely have to be used. Associations between the semantic interpretation and earcons could be made by rhythm,

dynamics, and so on, as mentioned by Gaver (1986) and others. The mapping between the motive and its interpretation must be learned. However, for a large number of earcons the message can become very sophisticated and the number of earcons that may be learned is much greater, although the initial learning period would be longer. Because transformations may be difficult to recognize, only simple transformations should be used.

8. CONCLUDING REMARKS

Not all aspects of earcons could be fully investigated or even addressed. For the sake of future research, however, a few are worth mentioning. Voice messages offer many interesting possibilities in a sight and sound user interface. Earcons could be used to reinforce voice output and/or graphics and nonverbal sounds. Also, voice recognition as sound input could result in complete two-way man–machine communication. Earcons for different types of systems could also be investigated. For example, would the structure and composition of earcons for office use be different than earcons for scientific use? The details of physical implementation of earcons in a user interface is an area in itself and requires substantial research. Because earcons are a communication language utilizing nonverbal sound messages, a formal grammar could be defined for their constructions. Formal grammars have been successfully established for music (Langston, 1986; Roads, 1979; Smoliar, 1980).

Additional research must be done into the differences between icons and earcons. This study suggests the possibility of using earcons to replace icons, but situations also arise where an earcon has no iconic equivalent. Because this would be a project in itself, it was not incorporated into this article. Our project initially was to include audio messages played simultaneously. We found that playing earcons concurrently created unanticipated problems that could not be resolved until the structure of individual earcons had been thoroughly studied. In addition, we would like to investigate how earcons might benefit the handicapped. Mansur et al. (1985) examined graphs in sound for the blind. Similarly, earcons could provide the visually handicapped with a variety of audio messages.

The utilization of sound in computer systems is still in its infancy, but its life expectancy is promising. Many contemporary systems already provide some type of sound-producing hardware and make minimal use of sound. As a secondary sensory stimulus, Pict, an interactive graphical programming environment (Glinert & Tanimoto, 1984), currently employs single audio cues in its user interface, where a bell confirms acceptance of user commands and two tones in rapid succession draw attention to an error. Human-factor testing of Pict on more than 60 subjects found that 78% of its users thought sound was important and useful in interactive computer sessions.

Finally, we believe the work done here may contribute to the study of a visual language for iconic displays. Korfhage and Korfhage (1984) pointed out that a better theory of visual languages is needed. Many principles used here can be carried out into iconic displays and may form part of a basis for a visual language of icons.

Acknowledgments. Kenneth Joy contributed many hours to this project and coauthored a preliminary version of this article. Sara Bly frequently participated in our meetings, and her suggestions and observations have been incorporated into this article. Eugene Schultz, our human-factor expert, taught us much about visual and auditory displays. Jonathan Hymer edited the manuscript. Finally, we thank our referees for their technical advice and suggestions.

Support. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

REFERENCES

Adams, S., & Trucks, L. (1976). A procedure for evaluating auditory warning signals. *Proceedings of the 6th Congress of the International Ergonomics Association and Technical Program for the 20th Annual Meeting of the Human Factors Society,* 366–374. Santa Monica, CA: Human Factors Society.

American Institute of Graphic Arts. (1981). *Symbol signs.* New York: Hastings House.

Arnheim, R. (1969). *Visual thinking.* Berkeley: University of California Press.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84,* 147–165.

Bartlett, J. C., & Dowling, W. J. (1980). The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 501–515.

Bernstein, M., & Picker, M. (1966). *An introduction to music* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

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

Buxton, W., Baecker, R., & Arnott, J. (1985). *A holistic approach to user interface design.* Unpublished manuscript.

Chomsky, N. (1963). *Syntactic structures.* The Hague, The Netherlands: Mouton & Co.

Christ, W., DeLone, R., Kliewer, V., Rowell, L., & Thomson, W. (1966). *Materials and structure of music I.* Englewood Cliffs, NJ: Prentice-Hall.

Davies, J. B. (1978). *The psychology of music.* Stanford, CA: Stanford University Press.

Deatherage, B. H. (1972). Auditory and other sensory forms of information presentation. In H. P. Van Cott & R. G. Kinkade (Eds.), *Human engineering guide to equipment design* (rev. ed., pp. 123–160). Washington, DC: U.S. Government Printing Office.

Deese, V., & Grindley, G. C. (1947). The transposition of visual patterns. *British Journal of Psychology, 37,* 152–163.

Deutsch, D. (1982a). Grouping mechanisms in music. In D. Deutsch (Ed.), *The*

psychology of music (pp. 99-134). New York: Academic.

Deutsch, D. (1982b). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (pp. 271-316). New York: Academic.

Deutsch, D. (1982c). *The psychology of music.* New York: Academic.

Dowling, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The psychology of music* (pp. 413-428). Orlando, FL: Academic.

Dreyfuss, H. (1972). *Symbols sourcebook.* New York: McGraw-Hill.

Easterby, R. S. (1970). Perception of symbols for machine displays. *Ergonomics, 13,* 149-158.

Elliott, F. R. (1937). Eye versus ear in moulding opinion. *Public Opinion Quarterly, 1,* 83-86.

Fodor, J. A. (1981). Imagistic representation. In N. Block (Ed.), *Imagery* (pp. 63-86). Cambridge, MA: MIT Press.

Gaver, W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2,* 167-177.

Glinert, E. P., & Tanimoto, S. L. (1984). PICT: An interactive graphical programming environment. *Computer, 17*(11), 7-25.

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.

Huggins, W. H., & Entwisle, D. R. (1974). *Iconic communication: An annotated bibliography.* Baltimore, MD: The Johns Hopkins University Press.

Kerman, J. (1980). *Listen.* New York: Worth.

Knuth, D. E. (1984). The complexity of songs. *Communications of the ACM, 27*(4), 344-348.

Kolers, P. (1969). Some formal characteristics of pictograms. *American Scientist, 57*(3), 348-363.

Korfhage, R. R., & Korfhage, M. A. (1984). The nature of visual languages. *1984 IEEE Computer Society Workshop on Visual Languages,* 177-183. Hiroshima, Japan: IEEE Computer Society Press.

Langston, P. S. (1986). Eedie & Eddie on the wire. *Proceedings: USENIX,* 13-27. Atlanta, GA: The USENIX Association.

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

Lundin, R. W. (1953). *An objective psychology of music.* New York: Ronald Press.

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

Mansur, D. L., Blattner, M. M., & Joy, K. I. (1985). Soundgraphs: A numerical data analysis method for the blind. *Proceedings: 18th Hawaii International Conference on System Sciences, 18,* 163-174. Honolulu, HI: IEEE Computer Society Press.

Marcus, A. (1984). Corporate identity for iconic interface design: The graphic design perspective. *Computer Graphics and Applications, 4*(12), 24-32.

Meyer, L. B. (1967). *Music, the arts, and ideas.* Chicago, IL: University of Chicago Press.

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

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our

capacity for processing information. *The Psychology Review, 63*(2), 81-97.

Moog, B. (1986). MIDI: Musical instrument digital interface. *Journal of the Audio Engineering Society, 34,* 394-404.

Ramsey, H., & Atwood, M. E. (1980). MMI design guidance: State of the art. *Proceedings: International Conference on Cybernetics and Society,* 579-582. IEEE Systems, Man and Cybernetics Society.

Roads, C. (1979). Grammars as representations for music. *Computer Music Journal, 3*(1), 48-55.

Rosner, B. S., & Meyer, L. B. (1982). Melodic process and the perception of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 317-341). London, England: Academic.

Schoenberg, A. (1951). *Style and idea.* London: Williams & Norgate.

Schneiderman, B. (1986). *Designing the user interface: Strategies for effective human-computer interaction.* Reading, MA: Addison-Wesley.

Smith, D., Irby, C., Kimball, R., Verplank, B., & Harslem, E. (1982, April). Designing the star user interface. *Byte Magazine,* pp. 242-282.

Smoliar, S. W. (1980). A computer aid for Schenkerian analysis. *Computer Music Journal, 4*(2), 41-44.

Sticht, T. G. (1969). *Learning by listening in relation to aptitude, reading, and controlled speech* (Tech. Rep. No. 69-23). Alexandria, VA: Human Resources Research Organization.

Sumikawa, D. A. (1985). *Guidelines for the integration of audio cues into computer interfaces.* Unpublished master's thesis, University of California, Davis, CA. (Also published as a Lawrence Livermore National Laboratory Technical Report, UCRL 53656.)

Williams, G. (1983, February). The Lisa computer system. *Byte Magazine,* pp. 33-50.

HCI Editorial Record. First manuscript received September 2, 1986. Revisions received August 10, 1987, and June 1, 1988. Final manuscript received August 8, 1988. Accepted by William Buxton. —*Editor*