Designing with Auditory Icons

Elizabeth D. Mynatt
Graphics, Visualization and Usability Center
Georgia Institute of Technology
Atlanta, GA 30332-0280
Email: elizabeth@cc.gatech.edu

Abstract

Current research in auditory interfaces has produced a number of interesting interfaces without resulting in the compilation of design guidelines for creating descendent interfaces. This lack of a common design methodology may be partially responsible for the scarcity of commercial interfaces that make use of sound. Factors that affect the usability of auditory interfaces are identified, and are used as the basis for a simple design methodology for incorporating auditory icons in human-computer interfaces. Two steps in this methodology are evaluating the identifiability of auditory cues and evaluating the possible conceptual mappings between the cues and concepts in interfaces. Two experiments that explore these issues are summarized.

1 Introduction

Over the past few years, workshops, conferences and publications concerning human-computer interaction have witnessed an increased interest in building auditory interfaces. Nevertheless, the majority of current work presented describes systems built on auditory display techniques (auditory icons, earcons, sonification) introduced many years ago. One conclusion to be drawn from this observation is that these techniques have been firmly established and that we (as a community) are only developing different applications using these auditory display techniques. But, an examination of textbooks on general human-computer interaction does not support this conclusion. Generally, the portions of the text devoted to sound in computer interfaces (if they exist) provide a short, favorable summary of the SonicFinder and descendent systems and still sometimes conclude with a section on why sound should be used sparingly in computer interfaces [1]. Interface designers can locate numerous guidelines for using color in interfaces as well as guidelines for constructing common graphical interface objects such as menus, dialog boxes and so on, but are left little to no assistance for including sound in interfaces.

Another conclusion is that this field is still young and that, with time, such guidelines for the use of sound will eventually become commonplace. While this conclusion is undoubtedly partially true, acknowledging that time is a factor is not sufficient to explain the current dearth of design guidelines. It appears that the general motivation behind current work is to prove that sound is useful. This quest would be satisfied by the emergence of a commercial application in the mainstream market that effectively uses sound in a number of interesting ways. But technology transfer is difficult, and simply focusing of the creation of more demonstration auditory interfaces may not be enough to move auditory interfaces into the mainstream computer industry.

Starting with the assumption that sound is useful in computer interfaces, the next logical step would be to develop a design methodology for using sound. Since it is reasonable to conclude that this design methodology might be dependent on the general technique employed to create auditory interfaces, I have narrowed the scope of this work to addressing the use of auditory icons. Auditory icons are based on the concept of everyday listening that contends that people describe sounds in terms of the objects and events causing the sounds, not in psychoacoustic or musical terms. Like their graphical counterparts, auditory icons are designed to remind the user of an object or concept in the everyday world [2].

The use of auditory icons has already been demonstrated in a number of systems. The SonicFinder [2] first utilized auditory icons for objects (the Macintosh trash can sounded like a metal trash can), actions (copying sounded like pouring liquid), and attributes of objects (the pitch of a file icon was related to its size). Mercator, which provides access to graphical interfaces for people who are blind, uses auditory icons to replace standard graphical icons [3]. For example, a toggle button sounds like an chain-pull light switch, conveying the notion of a two-state control. Editable text fields sound like old-fashioned typewriters, while noneditable text fields sound like printers. As in the SonicFinder, the auditory icons used in Mercator can also convey attributes of the graphical objects such as highlighting. Many other interfaces for simulating physics environments [4], monitoring surgery patients [5], drawing programs, supervising satellite ground control [6], monitoring background computer activity [7], and others have been built using auditory icons.

In short, a number of interesting and useful interfaces employing auditory icons have been constructed. But, excepting usability testing, these interfaces have been constructed on an ad hoc basis. Currently, there are few guidelines for the use of auditory icons. Likewise, a methodology for designing auditory icons does not exist. This paper describes work in progress to demonstrate a design methodology for selection, use, and evaluation of auditory icons. Factors that affect the usability of auditory icons are identified, and a design methodology corresponding to these factors is described. Two steps in this methodology are determining how well people identify typical auditory icons, as well as how people map auditory icons to concepts in user interfaces. The paper details a series of experiments evaluating the identifiability of many everyday sounds and the possible conceptual mappings between auditory cues and common concepts in graphical interfaces. The results from these experiments are used to derive a number of preliminary guidelines for using auditory icons.

2 Using Auditory Icons

Evaluations of auditory interfaces reveal a number of factors that affect the usability of auditory icons. These factors are:

1. *Identifiability:* The user must be able to recognize the sound. The ecological frequency (how common the sound is) versus the relative uniqueness of the sound help determine its overall identifiability [8]. If the user is unable to correctly identify the sound, then the usefulness of the sound is likely to be reduced. For example, in Cohen's ShareMon interface, the sound of a key turning in a lock represented a registered user logging into a computer system. Although the mapping between the sound and the interface concept is intuitive, the sound was not identifiable and therefore not utilized [7].

2. *Conceptual Mapping:* How well does the sound map to the aspect of the user interface it is representing? Sound seems to lend itself to storytelling. Evaluators of auditory interfaces are often amazed at the stories users will build to explain the sounds that they hear. My favorite example is again taken from the ShareMon system [7] where a knocking sound immediately followed by the sound of footsteps indicated that a user had logged onto a computer system

and was accessing the shared file space. One test subject explained the sounds by describing a father expecting a baby who was knocking at a door (no one responds) and then pacing back and forth. The metaphorical power of auditory icons is immense, but controlling these mappings is a difficult design task.

3. *Physical Parameters:* The physical parameters of the sound such as length, intensity, sound quality (for example, sampling rate) and bandwidth or frequency range may affect its usability. These physical parameters not only affect the user's reaction to the auditory cue, but they also affect the usability of the cue from the designer's perspective. The overall quality of the sound is largely determined by the sampling process. Users may negatively react to cheap (phone quality) or noisy sounds. The frequency composition of the cue may affect how it can be manipulated with standard filters [9].

4. *User Preference:* How the user responds emotionally to the auditory icon is also important. Is the sound harsh or too cute? Almost all evaluations of auditory interfaces have uncovered sounds that users did not like.

The auditory icons used in an interface must also be evaluated as a cohesive set, again addressing the above issues. Concerning identifiability, the cues must be relatively distinct from each other so that they can be individually identified. The conceptual mappings in the interface must not be counterintuitive. The physical parameters of the auditory icons such as length, intensity and sound quality must also be carefully controlled as the user might attempt to attribute some meaning to any perceivable differences. These issues are important even if the sounds are not simultaneously presented. In addition to the above issues, concurrent presentation of auditory cues must also address basic psychoacoustic issues such as masking and streaming.

3 Proposed Design Methodology

A methodology for designing with auditory icons can be derived from the above factors. Some basic steps in this methodology are:

1. Choose short sounds that have a wide bandwidth, and where length, intensity, and sound quality are roughly equal.

   Short sounds will conserve disk space and will ideally be recognizable in a fraction of a second by experienced users. Assuming an interactive system where the user will be able to quickly stop and start sounds (for example, by moving from one auditory object to another), encoding information in the length of a sound is not advisable. Likewise, equalizing the length of the auditory cues will help ensure that novice users do not incorrectly interpret the length of a cue in an unintended manner. Since the designer may not have complete control over the presentation of the interface, for example, volume control and speaker quality, equalizing the intensity and sound quality of the cues may help avoid errors in interpreting the cues.

2. Evaluate the identifiability of the auditory cues using free-form answers.

   Although free-form answers are difficult to interpret quantitatively, anticipating possible responses is difficult. Working with free-form answers will assist the designer in realizing the range of possible identifications.

3. Evaluate the learnability of the auditory cues that are not readily identified.

   Unfortunately most previous work (including the work reported here) has focused on the usability of auditory cues by novice users or test subjects. A general principle from interface design is to focus the design on the habits of the anticipated users of the system and not to design only for first-time, novice users. It follows that the learnability of auditory cues, both inside and outside of the context of a specific interface, needs to be addressed.

4. Test possible conceptual mappings for the auditory cues using a repeated measures design where the independent variable is the concept that the cue will represent.

   After deciding on a collection of sounds that are easily identified or learned, evaluate possible uses of the auditory cues in the interface being designed. By performing the evaluation outside of the actual interface, confounding factors from other aspects of the interface may be avoided.

5. Evaluate possible sets of auditory icons for potential problems with masking, discriminability and conflicting mappings.

   The auditory icons must also be evaluated as a set taking into consideration how they will be used in the interface. Optimizing for the set may require that the ideal sounds for some concepts (as indicated by high scores in the individual evaluations) may have to be discarded for sounds that form a more cohesive set. As stated earlier, sounds that will be presented simultaneously need to be evaluated for masking, streaming, and other possible psychoacoustic phenomena.

6. Conduct usability experiments with interfaces using the derived auditory icons.

   Finally the complete interface should undergo usability evaluations with subjects or users conducting specified tasks, as well as long-term evaluations that monitor performance over a long period of use.

4 Mercator–A Design Context

Currently this methodology is merely a proposal based on factors that appear to affect the usability of auditory icons and general principles in computer interface design. My current work has focused on employing this methodology while designing an auditory interface for the Mercator environment. Mercator is a system that provides access to X Windows applications for computer users who are blind by transforming the graphical interface into an auditory interface [3]. This interface must convey objects and actions common in graphical interfaces in an auditory-only presentation. The following sections detail two experiments that evaluated the identifiability of potential auditory cues, and then the possible conceptual mappings between graphical interface concepts and those cues that were most readily identified.

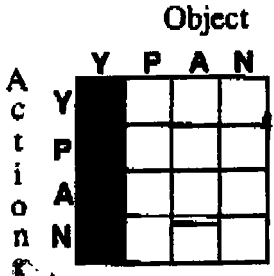
5 Identifying Auditory Icons

5.1 Method

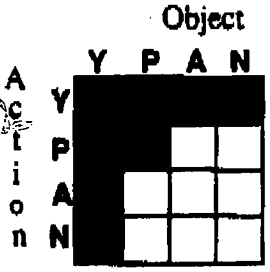
To assess the identifiability of common auditory cues, I collected 64 sounds (short "everyday" samples from sound effects CDs with minimal editing). The sounds were digitally recorded on a Silicon Graphics Indigo and stored on a digital audio tape (DAT). During the experiment, the

Y	Correct identification
P	(Partial) Either the subject identified a base material, an object with the same affordance, or a similar action.
A	(Alternate) The subjects commonly selected a consistent (alternative), yet incorrect answer.
N	Incorrect or no answer

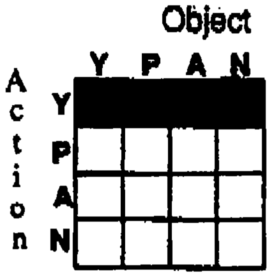
Table 1: Possible ratings for identifying the object and action associated with a sound.



Object Score



Overall Score



Action Score

sounds were played over two BOSE speakers to a classroom of 83 students. The acoustics of the room were sufficient to ensure that the sounds were not distorted from increased volume levels. Under tight time constraints, subjects were asked to describe (free-form) the sounds as best they could. The sounds were presented in the following manner:

"Please identify Sound #1" ⟨play Sound #1⟩ [2 seconds] ⟨play Sound #1⟩ [7 seconds]...

Since many of the sounds are difficult to identify, I suggested that the subjects attempt to identify an object and an action that could be causing the sound.

The subjects' responses were analyzed by transforming their answer into a point in a 4 × 4 matrix. One axis on the matrix indicates how well the subject identified the object associated with the sound. The other axis indicates how well the subject identified the action associated with the sound. Per object and action, four scores were possible.

In my analysis, I have focused on the three scores illustrated

The overall score provides a general measurement for how well the sound was identified. The object and action scores indicate respectively how well the object and action associated with the sound were identified.

5.2 Observations

5.2.1 Identifiability Scores are Low

The first result to note is that, in general, the identifiability scores were low with few sounds (13) showing an overall score over 80 (eighty) percent. This result is not surprising given the difficulty of the task and the short length of the sounds (under two seconds). In fact, the second experiment, discussed next, indicates that repeated exposure to the sounds increases their overall identifiability. On the other hand, it was easy to discern when sounds were relatively unidentifiable. Common answers in these cases were drumming or dragging.

5.2.2 Distinction between Actions and Objects

The distinction between identifying a sound as an object or an action is clearly evidenced. Many sounds are consistently identified as objects (camera, typewriter, door, zipper) while other sounds are consistently identified as actions (closing, tearing, winding, locking). For example, for the sound of a camera taking a picture, specifically the shutter operating, subjects easily identified the object as a camera, but were unable to identify the correct action (taking picture, flashing, winding film). Another sound of winding a parking meter was easily identified as winding but the object (meter, typewriter, music box) was not identified. This separation supports the guideline for choosing some sounds to represent objects while using other sounds to represent actions. In general, the designer should decide if an action or an object (or both) need to be conveyed, and then choose the appropriate auditory cue.

5.2.3 Partial and Alternative Answers Merit Further Research

Many of the sounds resulted in high partial (P) or alternative (A) scores. In other words, the subjects are hearing some of the information in the sound. A partial answer for the object was recorded if the subject identified the base material of the object, or identified an object with the same affordance. For example, for the sound of closing a metal mailbox, recognizing the metallic material or answering "closing a file cabinet" resulted in a partial object score. For actions, identifying similar actions resulted in partial scores such as dropping or throwing, swinging or striking, or closing, latching or locking.

An alternative answer was scored if the pool of subjects tended to identify the object incorrectly but consistently. Cultural biases may partially explain alternative answers. For example, in my pool of undergraduate Georgia Tech students, whip cracks were identified as gun shots and popping a cork out of a champagne bottle was identified as making a popping noise with one's finger and cheek.

The large number of sounds with partial and alternative scores motivates future work in the learnability of auditory cues. One hypothesis is that sounds with high partial answers are easy to learn whereas sounds with high alternative answers are difficult to learn since it would require the subject to break the existing mapping between the sound and the alternative concept.

5.2.4 Sounds Naturally Conflict

Many sounds naturally conflict. For examples, machines that afford typing such as typewriters, keyboards, cash registers are natually confused. Likewise printing and copying sounds are often mistaken for each other. These conflicts indicate when it may be difficult to present two similar concepts in an interface using auditory cues. For example, although the sounds of a copier and printer may be quite distinct, it may be difficult to correctly identify them when they are both used in the same interface.

6 Mapping Auditory Icons To Graphical Interfaces

This experiment assessed how users expected auditory cues to be used to represent concepts common in graphical interfaces.

6.1 Method

Twenty-eight sounds were selected from the initial collection of sounds used in the first experiment. The sounds selected had high identifiability scores from the first experiment. Fifteen concepts (objects and actions) from graphical interfaces were chosen as the independent variable. During the experiment, each subject was presented with the graphical interface concept and five sounds. Their task was to choose which sound best matched the interface concept. The presentation order for the concepts and sounds was randomized such that each sound was presented with each concept. The subject's current favorite sound for each object would then be presented alongside four new sounds the next time the concept was tested, such that, by the end of the test, a best sound for each object had been chosen. Nothing in the test prevented the subject from choosing the same sound for multiple concepts. In each trial the graphical concept, the five sounds presented, the chosen sound, and the time required to make selection was recorded.

Each sound was a short (under 2 seconds) sample recorded at 16-bit, 44.1 kHz. Each concept was represented with a short, textual description and a picture of what the concept could look like in a graphical interface. A description of the sounds used are in Table 2. The screenshots and descriptions used to represent the graphical concepts are indexed in Table 3.

Twenty-six subjects completed the test. Basic demographic data such as name, age, gender, and major was collected as well as the number of years that the subject had lived in the U.S. and an indication of their experience with music and audio production. I debriefed each subject at the end of the computerized test. During the debriefing, subjects were asked to write short descriptions of each sound that was used in the experiment. Each sound was presented once during this test. Next, using a paper-copy of the interface concepts presented in the experiment, I asked the subjects what sounds they had chosen and what had motivated their selection. These discussions were videotaped with the subject's consent.

6.2 Observations

The observations reported here are based on debriefing the subjects and an quantitative inspection of the numerical data gathered during the experiment.

6.2.1 Four Types of Selection

During the debriefing, the subjects described why they had chosen a particular sound for an interface concept. The reasonings used to guide their selections seem to fall into four major categories. First, the subjects (as predicted) chose sounds that they felt represented the semantics of the interface concept. For example, the sound of typing on a typewriter (multiple keystrokes) was a common choice to represent an editable text field. Likewise, for closing a window, the sound for closing a car door was a common choice, followed by the sound of closing a zipper. This observation is not surprising, but it does indicate that sounds can be used to represent semantic information in the interface.

Second, several subjects indicated that they choose sounds that mapped well to the perceived importance of the interface concept. For example, the sound of a short tap on a typewriter was the most common sound for the check boxes that implement nonexclusive choices. Subjects reasoned that selecting a check box is easily undone, and therefore less important than pushing a button that was represented by a longer, more complex sound such as cracking a whip.

Third, subjects also indicated that they chose sounds that mapped to the perceived length of the interface concept. For example, selecting a radio button (common sounds: short and long

Drop of water falling into puddle	Popping the cork out of a bottle-long	Airplane flying by
A camera shutter	Printer printing out a line	Rain
Opening a creaky, wooden door	Playing a slide whistle—pitch ascending	Motorcycle driving by
Car driving by	Typing on a typewriter	Cars driving by
Revving the engine on a motor scooter	Cracking a whip	Flipping papers
Zipping (up) a zipper	Someone knocking at a door	Airplane taking off
Slide whistle playing—descending pitch	Pulling a light switch	Closing a car door
Popping the cork out of a bottle—short	Pinball dropping into hole	Winding a music box
Zipping (down) a zipper	Dropping ice into a glass	Ripping paper
Someone walking (footsteps)	One keystroke on a typewriter	

Table 2: Sounds Used in Mapping Experiment.

pops) was described as taking less time than opening a folder (common sounds: opening door and slide whistle-pitch ascending).

Fourth, the subject choose sounds that mapped to the physical appearance of the interface concept. For example, a slide whistle (pitch descending) was the most common sound for a pull-down menu. Subjects explained that pop-up menus would have a different sound although the menu concept was shared by both objects. Several subjects explained that the sound of a camera shutter mapped well to the concept of a window appearing on the screen since the sound reminded them of square shapes being presented in quick succession.

6.2.2 Identifiability Ratings Consistent

After subjects completed the computerized mapping experiment, they were asked to write short descriptions of the sounds that were used in the experiment. Each sound was played once before the subject responded. Although this identification task is more difficult than the task in the first experiment, the identification rates remained consistent or increased. Two theories may explain this increase. First, the initial mapping task forced the subjects to think about the contents of the sounds more than the simple identification task. Second, the repeated exposure to the sounds over a period of approximately 40 minutes may have helped the subjects learn the sounds even though the sounds were never identified during the course of the tests.

6.2.3 Feedback is the Expected Use of Sound

Subjects' comments during the debriefing indicated that they expected the auditory cues to be used as feedback in the interface, usually as the result of a direct manipulation task. This use is in contrast to using sound to simply convey the identity of interface objects independent of a particular action. For example, in Mercator, navigating to an object causes a sound to be played that conveys the type of object. This distinction between conveying the identity of an object separate from feedback resulting in manipulating the object is important in auditory-only interfaces where the contents of the interface must be conveyed aurally [3].

6.3 Switch Stimulus and Response

I repeated the mapping experiment by using the auditory cues as the stimulus and asking the subjects to pick the best interface concept to represent the auditory cue. The structure of the experiment, including the debriefing, remained the same. Only eight subjects participated in this experiment.

1.	An editable text field.	9.	Dragging an icon.
2.	A slider or scrollbar. Control used to increment or decrement a value or location.	10.	A pull-down menu—a collection of menu buttons organized in a group.
3.	Selecting an item from a list.	11.	Radio buttons. Buttons in which only one button can beselected. Buttons which operate as an exclusive choice.
4.	A scrollable list. A collection of list items organized in a list that can be scrolled.	12.	Copying an item.
5.	Selecting a menu item from a menu.	13.	A push button. A control that the user selects to cause an action to occur.
6.	Closing an item.	14.	An noneditable text area—where the application displays error messages and other information.
7.	Check boxes. Buttons in which more than one can be selected.	15.	Opening an item. Buttons which implement nonexclusive selection.
8.	A window appearing on the screen.		

Table 3: Textual descriptions of interface concepts used in mapping experiment.

6.4 Combining Results

One difficulty in utilizing the data gathered in each mapping experiment was that the subjects rarely converged on one sound for each interface concept. Generally, a number of favorite sounds were identified. Also, nothing prevented the subjects from picking the same sound for multiple interface concepts. By combining the data from both experiments, it is possible to highlight the best sounds.

In Table 4, the best sounds for each interface concept are listed. The chi-square value in second column describes the distribution of selected sounds. A low chi-square value indicates an equal distribution where none of the sounds were favored. A high chi-square value indicates that a subset of the sounds were favored to represent the given individual concept. The third column indicates the probability of the distribution of choices occurring by chance. The sounds that the subjects thought best matched the interface concept are listed in the fourth column. The value in the fourth column is that sound's individual contribution of the overall chi-square value. The last column indicates the number of subjects who matched the sound and concept given the sound as the stimulus.

Conclusions

Cohen concludes his summary of ShareMon by discussing the pitfalls of sound design [7]. Despite excellent example systems that use auditory cues, these pitfalls are all too common. Determining the ideal sounds to use in a human-computer interface is a difficult and painstaking process. At this time this process is more of an art than a science, dependent on skilled and gifted designers. But for auditory interfaces to become commonplace in the world of computer-based interaction, this process must mature into a science. This work aims to be one step in this maturation effort. Insights from systematically designing auditory interfaces should help speed the creation

Concept	X^2	α	Sounds	$X_c^{2_s}$	$\#_c$ [0-8]
Copying	136	<.001	camera shutter	96.86	2
Closing	113	<.001	closing car door	59.10	4
			zipping down	30.62	2
A text field	101	<.001	typing	76.82	6
A slider	92	<.001	whistle up	59.10	2
Check boxes	89	<.001	keystroke	59.10	2
Dragging	78	<.001	whistle up	43.70	0
			cars driving by	11.42	2
Opening	75	<.001	open door	19.86	3
			whistle up	19.86	3
			motorcycle drive	11.42	1
A push button	73	<.001	crack whip	19.86	0
			short pop	11.42	2
			keystroke	11.42	1
			water drop	11.42	1
A menu	68	<.001	whistle down	31.62	0
			flipping papers	11.42	2
			open door	5.30	1
			pinball drop	5.30	0
Radio buttons	61	<.001	short pop	19.86	2
			long pop	19.86	2
A scrolling list	59	<.001	winding	19.86	0
			whistle up	5.30	1
			whistle down	5.30	0
			motorcycle drive	5.30	0
			flipping papers	5.30	0
Select from menu	52	.01	keystroke	30.62	1
Window appear	43	.05	winding	11.42	0
			whistle up	5.30	1
			cars drive by	5.30	2
Select from list	38	.25			
A message	34	.25			

Table 4: Mapping interface concepts to auditory icons.

of a shared design knowledge that can be used to create not only useful, but usable auditory interfaces.

Acknowledgments

This work is sponsored by Sun Microsystems, the NASA Marshall Space Flight Center and the Georgia Institute of Technology.

References

[1] Mayhew, D. J. *Principles and Guidelines in Software User Interface Design.* Prentice Hall, 1992.

[2] Gaver, W. "The SonicFinder: An Interface that Uses Auditory Icons." *Human-Comp. Interaction* 4 (1989): 67–94.

[3] Mynatt, E. "Mapping GUIs to Auditory Interfaces." In *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer, 533–556. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[4] Gaver, W. and Smith, R.B. "Auditory Icons in Large-Scale Collaborative Environments." *Interact '90: Proceedings of the IFIP TC 13 International Conference* (1990): 735–740.

[5] Tecumseh, W. T. Fitch, and G. Kramer. "Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System." In *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer, 307–326. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[6] Albers, M. Workshop on the Future of Speech and Audio, ACM Conference on Human-Computer Interaction. 1994.

[7] Cohen, J. "Monitoring Background Activities." In *Auditory Display: Sonification, Audification and Auditory Interfaces*, edited by G. Kramer, 499–532. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[8] Ballas, J. A. "Delivery of Information Through Sound." In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, 79–94. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[9] Ludwig, L. F., and M. Cohen. "Multidimensional Audio Window Management." *International J. Man-Machine Studies* **34(3)** (1991): 319–336.