Human Factors and the Acoustic Ecology:
Considerations for Multimedia Audio Design.

Bruce N. Walker
Psychology Department,
Rice University
P.O. Box 1892
Houston, TX 77251
(713) 527-8101
Fax: (713) 285-5221
walkerb@rice.edu

Gregory Kramer
Clarity/Santa Fe Institute
310 NW Brynwood Lane
Portland, OR 97229
(503) 292-8550
Fax: 503-292-4982
kramer@listen.com

Running Head: multimedia human factors

ABSTRACT

Acoustic designers face new challenges as audio, video and computers converge into the synergy of multimedia. The sound engineer must consider the human factors and acoustic ecology involved in the new "conversation" between listeners and a multimedia system. An organized and centralized sound design will make the audio element a rich ingredient in the multimedia experience.

Human Factors and the Acoustic Ecology:
Considerations for Multimedia Audio Design.

Bruce N. Walker
Psychology Department, Rice University

Gregory Kramer
Clarity/Santa Fe Institute

Acoustic designers face new challenges as audio, video and computers converge into the synergy of multimedia. The sound engineer must consider the human factors and acoustic ecology involved in the new "conversation" between listeners and a multimedia system. An organized and centralized sound design will make the audio element a rich ingredient in the multimedia experience.

0 INTRODUCTION: INTERACTIVE AUDIO

Audio has typically been part of the recreational and entertainment worlds, while computers have mostly been information retrieval tools. There have been some notable exceptions (Speeth, 1961), but sound was generally used only as a warning, and often redundant with a visual display (Patterson, 1982; Kramer, 1994a). Visual displays have been a huge part of computing for decades, so there is a well-established display style. Audio, on the other hand, is really only beginning to play a significant role in consumer computing. Until very recently, the few guidelines that did exist for auditory displays focused on issues related to warnings, such as audible thresholds and masking (Patterson, 1982; Sorkin, 1987). Now, however, audio designers face new and interesting challenges as audio, video and computers converge into the synergy of multimedia.

The listener now is an active part of the system; the designer must consider many issues that have not been as important in other, more traditional, audio realms. While perhaps new to some audio engineers, these ideas come from the well-established field of human factors. In particular, three issues to consider are: the nature of the interaction; the user audience; and the resulting acoustic ecology.

1 THE INTERACTIVE LISTENER

The sound engineer must accept a "conversational model" of cooperative interaction between the user and the system (Shneiderman, 1992). This is quite different from the traditional "concert model" of sound presentation, which treats the listener as a passive recipient of the audio presentation. The multimedia listener pursues the dual goals of being entertained and of gaining information. While entertainment may or may not involve interaction, information retrieval and efficient learning is usually interactive, so the listener has evolved into an active player in the audiovisual exchange. This means that the display system, and therefore the sound designer, must be sensitive to what the user wants from the system at any given time. Of course, in a multimedia context, this balance of information and entertainment will change dynamically, providing a challenge for the audio design. But just as in any conversation, communication norms will emerge.

For example, when the aim is mostly entertainment and the application is not a game, the user will generally watch and listen to the resulting presentation in a fairly passive way, similar to the way one traditionally watches a movie on video. The computer can take more control over the presentation, and present an active sound environment with many simultaneous audio events that add to the richness of the experience. This is the time for the full orchestra - all

the bells and whistles, as it were. Note that the listener will not appreciate even well-intended interruptions from the system. The caveat to this mode of interaction is that it may be very short-lived in the multimedia environment, as new information sources are chosen, or as the user moves on to a new task.

On the other hand, if the aim is to extract specific information from any particular audio source, be it recorded speech, a musical selection, a data sonification, or an audible system warning, other sound sources should not interfere. In this situation, the user will play a more active role and will expect the system to respond in an active way. This mode of interaction is more dynamic, and the system needs to respond in an appropriate and timely fashion. For the audio designer, this means keeping the sound sources focused on the information task at hand. Simplify the audioscape and give control of it back to the user.

The system needs to seek and provide feedback in this style of conversation. Remember that the listener is still the leader of the conversation and this may result in a non-linear, non-sequential path through the sound and information space. The designer must respect the roles of the participants in the dialog, considering their goals, pace, need for privacy and the willfulness of the user. The user must maintain the feeling of control over the audio environment (and the interface in general); the multimedia system is simply an integrated tool to help achieve the listener's goals for the session.

2 KNOW THE USER!

Accepting the multimedia user as an integral participant, it is imperative that the designer know exactly who the users are. The "Six W's" are useful for thinking about the human and psychological factors involved.

2.1 Who

Designers must study audience demographics (*who*), since different groups interact differently. As the importance of information increases, so does the effect of psychological variables like learning style, learning rate, memory, attention and frustration, not to mention broad differences in perceptual skills and computer experience (Shneiderman, 1992). For example, adults tend to interact in a more consistent and structured way than children. More experienced users tend to be more active participants than novices, who tend to let the system make decisions. Listeners with more musical training may prefer an audio display with more simultaneous sound sources. And if a particular sound is used to represent an action or event, some listeners will take longer to make the association or may not retain the information as well.

2.2 Why

We have already mentioned that different users have different reasons (*why*) for participating in an interactive session, each requiring a unique balance of information and entertainment elements (Friedlander, 1995). The reason a person has for using a system can provide a lot of information to the audio system designer. A business user will more likely be seeking information, so a simpler audio environment may be appropriate. The home user is more likely to want all the richness the audio system is capable of. However, business users must still have the option of a complex audio environment, just as the home user must have a controlled soundscape when it is time for a homework assignment.

2.3 What

The actual "topic of conversation" (*what*) will also vary among users and among applications, even within the same multimedia system. In a school context one student might be browsing through music video clips, looking for a sound bite to include in a report, while the next student might be analyzing the results of a chemistry experiment. Naturally, the design of the audio interface will be quite different for each task that will be performed with the system, and for different types of content.

2.4 When and where

When and *where* the user interacts with a multimedia system will have a major impact on the audio display. Set in a busy office or classroom, a noisier environment and typically lower quality sound equipment means perceptual problems for the listener, possibly resulting in frustration or lack of interest (Friedlander, 1995; Strawn, 1994). Better audio hardware and speakers are becoming more common, but noise and interruptions are common in many settings. Headphones may help in some cases, but they cannot be considered standard equipment. Flexibility must be engineered into the interface to allow for noisy (or quiet) conditions. Subtle differences between sounds will often be difficult to perceive and so must be avoided in noisy settings, especially if decisions are to be made based on the perception of those sounds. Large differences in pitch, volume, rhythm and timbre will help a great deal. For example, if one system message is presented using the

sound of a violin, then a second message should be presented with a snare drum, rather than a cello, for example, in order to maximize comprehension (Brewster, 1994).

2.5 How

The "sixth W" is *how* the designer actually uses sound to communicate with the listener. Choices of mappings and metaphors, affective aspects, and general aesthetics are crucial at all levels of the auditory display (Blattner, Sumikawa, & Greenberg, 1989; Brewster, 1994; Das, DeFanti, & Sandin, 1995; Gaver, 1988; Kramer, 1994a; Patterson, 1982). The audio designer must learn to communicate using the tools of music and cultural stereotypes (Deutsch, 1982; Lakoff & Johnson, 1980). For example, due to cultural learning in Western music major scales may be taken to represent "good" or "positive" or "correct" ideas; minor scales may represent "bad", "sad" or "wrong". A louder sound may represent “more” of a selected data value, as may higher pitch. Higher pitch may also be associated with "up", while lower pitch generally means "down". A rapid tempo tends to convey a sense of urgency; long, slow sounds project tranquillity. And large objects (whether drums, elephants or computer files) ought to sound "big" -- that is, they should have rich, low pitched sounds with gentle onsets and offsets.

However, care is suggested when using these metaphors; they may not be understood by everyone in the intended audience and meanings might not transfer well to different cultural groups (Lakoff & Johnson, 1980). Above all, the whole audio environment must sound "good" and "right" to the *audience*, and not just to the designer. So the designer seeks to create an integrated, sensible audio environment, consistent across situations and across applications. Remember that for a successful audio element, guessing can never substitute for testing! When in doubt about the aesthetics or comprehensibility of any auditory display, ask the users (or find a better alternative).

3 THE ACOUSTIC ECOLOGY

The acoustic ecology of a multimedia / multiapplication environment becomes very difficult to manage (Kramer, 1994b). Speech, music, synthesized and spatialized sounds, notifications and data sonifications may all be part of the multimedia interface. At the basic level, mapping information or mood onto a feature of the auditory output must make sense to the user and must successfully transmit the designer's intent. With variable audiences and only preliminary auditory design guidelines available (Brewster, 1994; Kramer, 1994b), providing a successful audio addition to the multimedia computer system becomes extremely difficult. At the system level, the need for standardization is clear (Glinert & Blattner, 1993). There are benefits to a broadly implemented, as opposed to highly individualized, approach to sound implementation. A user might run several applications concurrently, each trying to add sound to the listener's environment, based, perhaps, on different designers' views of what "sounds good." If the acoustic result is cacophony, the user will quickly shut it off! We need to know how the rest of the system will react when one application produces a given sound. Part of the solution will depend on the emergence of a full-featured sound manager, much like the windows managers that coordinate the graphical aspects of most computer systems. This pair of display managers, audio and graphical, will soon evolve into a true multimodal display manager that makes informed decisions about how to present given information, automatically using the available system resources in the most effective manner possible (Glinert & Blattner, 1993; Papp & Blattner, 1995).

However, the first and most significant step toward true integration is an effective and consistent implementation by audio designers. Whichever style of audio display users like the most (which may be determined by market share, which breeds familiarity) will likely become the de facto standard in common applications. Thus, the industry is in need of reaching two main goals. First, we need to make these audio interfaces good: audio designers, engineers, musicians, programmers, and human factors specialists must cooperate to find out what the users want, what works, and then to test it all out a lot! This is an iterative process, especially in such a new field. Second, we need to get these well-designed products out there soon! The sooner we can present a common audio interface style, the faster we will see audio technology making the substantial positive impact on computing that it has had on more mature fields such as telecommunications and home entertainment.

4 CONCLUSIONS

The use of sophisticated audio in a multimedia computing context is increasing rapidly. The sound engineer must consider the human factors and acoustic ecology involved in the new "conversation" between active listeners and a multimedia system. It is imperative to understanding as much as possible about the user, and to accept the fact that the listener is no

longer a "passive element" in the audio system. The industry as a whole needs to work together to implement effective auditory displays. An organized sound design along with commonly accepted and implemented design principles will make the audio element a rich and vital ingredient in the multimedia experience.

5 REFERENCES

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4(1)*, 11-44.

Brewster, S. A. (1994). Providing a structured method for integrating non-speesh audio into human-computer interfaces. Unpublished doctoral dissertation, Department of Computer Science, University of York, UK.

Das, S., DeFanti, T., & Sandin, D. (1995). An organization for high-level interactive control of sound. In G. Kramer and S. Smith (Eds.), *Proceedings of the Second International Conference on Auditory Display, ICAD '94* (pp. 203-215). Reading, MA: Addison Wesley.

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

Friedlander, L. (1995). Space of experience: On designing multimedia applications. In E. Barrett and M. Redmond (Eds.), *Contextual media: Multimedia and interpretation* (pp. 163-164). Cambridge, MA: MIT Press.

Gaver, W. W. (1988). Everyday listening and auditory icons. Doctoral dissertation. University of Califonia, San Diego.

Glinert, E. P., & Blattner, M. M. (1993). Programming the multimodal interface. In *ACM Multimedia '93 Proceedings* (pp.189-197). New York: ACM.

Kramer, G. (1994a). Some organizing principles for representing data with sound. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces* (pp. 185-221). Reading, MA: Addison-Wesley.

Kramer, G. (1994b). An Introduction to Auditory Display. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces* (pp. 1-77). Reading, MA: Addison-Wesley.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Papp, A. L. III, & Blattner, M. M. (1995). A centralized audio presentation system. In G. Kramer and S. Smith (Eds.), *Proceedings of the Second International Conference on Auditory Display, ICAD '94* (pp. 231-240). Reading, MA: Addison Wesley.

Patterson, R. D. (1982). *Guidelines for auditory warning systems on civil aircraft.* Paper No. 82017, Civil Aviation Authority, London.

Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction.* Reading, MA: Addison-Wesley.

Sorkin, R. D. (1987). Design of auditory and tactile displays. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 549-576). New York: Wiley & Sons.

Speeth, S. D. (1961). Seismometer sounds. *Journal of the Acoustical Society of America, 33*, 909-916.

Strawn, J. (1994).Digital audio presentation and processing. In J. F. K. Buford (Ed.), *Multimedia systems* (pp. 65-107). New York: ACM Press.