Micro-Narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation

Maribeth Back, *Xerox PARC*
D. Des, *Xerox PARC*

Abstract: This paper reviews sound design techniques used in professional audio for music and theater and proposes a conceptual approach to the construction of audio based in narrative structure. The sound designer does not attempt to replicate "real" sounds; the task is rather to create the impression of a real sound in a listener's mind. In this attempt to create a sound in the listener's mind, the sound designer is aided by user expectations based upon cultural experience as well as physical experience. Practical sound manipulation techniques are discussed in view of their usefulness in matching a listener's mental model of a sound. Narrative aspects of audio design in computational environments are also delineated. Some keywords involved in this paper are sound design, auditory display, multimodal interaction, interface design, narrative, sonic narrative, micro-narrative, and audio.

Reading Sonic Media

Composers and sound designers draw from a system of metaphoric constructions that often seem cliched; the sound of slow violins is sad, and the crash of thunder is threatening or ominous. Designers make choices because they know how their audience will react to these sounds. This shared understanding has two main sources: cultural conditioning and natural cognitive mappings based on experience. For example, the use of thunder as a metaphor for trouble may have its origin in common-sense, but encoding this sound both in spoken language ("he looked thunderous") and in its constant use as a metaphor in various media reinforces the cultural understanding of this sound as metaphor.

Music and sound design constantly employ a vast vocabulary of such impressionistic metaphors and sound icons. As our culture changes, so do the contents and the semantic values of this vocabulary. Some sounds are less definite than others, and some have a multiplicity of uses depending on context. The effectiveness of the designer lies in recognizing the varying metaphoric values of different sound elements and in combining these elements to tell the right story.

Narrative in Nonspeech Audio

In the most overt sense of story, narrative is both a powerful cognitive mechanism and one of our primary methods of entertainment (Schank & Abelson, 1995). The design of multimodal artifacts with computational ability is one challenge in itself, but imparting to the user how to think about and use this artifact is harder. An artifact can help by telling the user its own story and by using the right story for any given state or activity. In this sense, the narrative of an artifact or an environment is not an unwavering start-to-finish plotline; the narrative is rather a set of smaller stories, little tales, and events that all add up to a cohesive world. In the familiar example of the modern personal computer, one might throw out the trash, open a file, find a name, or speak to a friend; all of these small stories are affordances of the desktop metaphor.

Grammars for dynamic media such as film and video are described as formalizations of the methods developed over many years of production. Elements such as camera angle, focus, shot width, and camera motion are combined in known ways to produce certain effects. In dynamic media, such grammars are not prescriptive, for they do not exclude new possibilities. These lexical structures instead add to possible creative combinations by allowing the artist to consider scenes with internal complexities as single units in much the same way a writer might use a favorite phrase that is made out of individual words. One possible model for understanding narrative structures in sound design involves using dynamic or

generative grammars such as those for more limited sonic domains proposed by Albert Bregman ([1990](#)) or David Cope ([1992](#)). However, these structures do not adequately account for the highly developed semantic systems already embedded within sound design (speech or music).

We learn to "read" the sound on film as though it were real when, in fact, this sound is nearly all artificial. Foley artists and radio producers share a sense of the unreality of actual recorded sound; we often do not recognize a recorded sound because it does not fit the mental model we have for that sound ([Mynatt,](#) 1994). Thus, thunder must crack, boom, or roll, and seagulls must utter high lonesome cries or harsh squawks; listeners will reject any of the myriad of other sounds made by thunder or seagulls as not authentic. To be understood, a recorded sound must match the listener's mental model of it. The sound itself must tell the right story; only then will that sound be a useful device for communication.

This narrative model of nonspeech audio design is partially based on Gaver's "everyday listening" techniques ([Buxton, Gaver, & Bly](#), 1989; [Gaver](#), 1992). However, the designer profits from exaggerating or caricaturing especially useful aspects of some kinds of sound or some elements of a particular sound. In that sense, we use everyday sound as a stepping stones to narrative design.

Icon to Metaphor to Narrative: Moving to the Next Dimension

Iconic sound systems rely on the use of different sounds to impart various information. Perhaps the most extensive use of auditory icons to date involves Gaver's Sonic Finder that, when added to the Macintosh Finder, appropriates sounds for actions (e.g., opening a file, dragging an object, or emptying the trash) ([1989](#)). Yet many dynamic parameters, such as volume, voice stacking, pitch shifting, etc., can also be effective ([Kramer](#), 1992). Whatever the sounds in the auditory construct are, they must reside within a consistent structure. Elements of such a

structure include but are not limited to

- Context: Where am I?
- Story: What is going on?
- Dynamics: What is happening now?
- Expression: How does it feel?
- Instruction: What happens next?

Narrative takes the constructed world a step further than metaphor does by consciously including time into its construction. If time is part of the structure, then story becomes possible, for behavior is possible. In fact, the desktop "metaphor" does operate as a narrative in some instances, for the animation sequences on the opening or closing of a file folder or an application is one instance of this narrative function. A miniature story is being told with that animation; the open window appears to shrink into the icon representing its "container."

Nonlinear Narrative in Sound

What makes a story work is not so much the choice of particular events but rather the shape and ordering of those events (e.g., elements such as sequence, pace, surprise, revelation, and style of presentation). It is difficult to create interactive narratives because (in most attempts thus far) artistic control of these elements is weakened or absent as control is given to the user instead of to the designer. Nonlinear narrative in sound can either be a designer's dream or his or her nightmare, for such sound creates a set of tools for the user that are difficult to misuse. In composing for interactive instruments (a very similar task), transition points are key. The question is how to get from one small narrative section to another. The content of each segment must be adjusted to link well to other segments, and some overall protocol must be established for manipulating the linkage. In music, this linking may mean attention to key, tempo, voice, and volume; in a series of everyday sounds, linking means arranging a graceful exit from one sound while another begins to play and making

sure that the overlapping sounds do not create any unwanted sonic artifacts. This linking also means establishing a story for the environment that allows the listener to make the transition between sounds logically and without losing context (Ballas, 1992).

Monet and Time: A Visual Analogy

Consider the temporal explorations of the Impressionist painter Claude Monet (1840-1926). Monet did several series of paintings that focus on the same subject (e.g., a haystack painted during the course of a day--Monet would set up several easels, and as the light changed, he would stop work on one painting and move on to the next.) Monet also painted more than forty studies of the facade of Rouen Cathedral in various lights, at different times of day, and during different seasons. The same structure (the facade of the cathedral) is differently narrated by the painter; in one painting, the facade's shadows are blue and brown while they appear gray and red in another. In a third work painted under overcast skies, no sharply defined shadows appear at all, and the sky changes hue in tandem. Monet was exploring how light and color change with the time of day and year. From our own experience, we understand how such changes come about, and we correctly interpret the artist's narration.

How does Monet tell this story? What are his narrative elements? The shadows of the cathedral facade are some of the characters in the painting; the details of the facade are others, and the light itself is another. If we compare two cathedral paintings, we can contrast the shape, color, and length of the shadows, the color of the light at the edges, and the scaling of color across a limited palette of hues as the eye travels over the shape of an arch or a sculpted detail. The story painted about the doorway shadow is a mini-narrative supporting and supported by a similar mini-narrative about the color of the light on top of the rose window and another about the sky in the distance. At an even more minute level, Monet's palette moves from dark blues and browns to greens to grays and then

to reds. His color language changes, and the gestural expression found in his brush strokes also changes.

In a similar fashion, the sound designer controls elements of the sonic palette by adjusting sound styles and the relationships between foreground and background sounds, lengthening or shortening sounds, and layering sounds into an effective pattern. Because sound is inherently a time-based art, the designer has a much finer resolution in the grain of detail over time; matters such as the length of the echo or the sound of a coin dropping can be tailored to fit a particular set of circumstances.

Just as with visual disciplines, the sound designer's ears and hands require training. For example, in the field of multitrack music recording, it is said to take five years in the studio to "learn to hear." Ear training, which is common among musicians, is equally important for sound designers. However, the sound designer listens for both additional sets of sonic structure and musical relationships. The entire sonic event, including variables like background noise or the frequency effect of various speaker systems, is the responsibility and the artistic product of the sound designer.

Microstructures: Resculpting the Waveform

The concept of narrative microstructures powerfully informs the construction of a single sound. Layered detailed manipulation of the waveform is a major part of the sound designer's work; this manipulation is also one of the most creative parts of the sound designer's work. In a sound-manipulation software analogous to a word processor, the sound designer mixes files, shapes loudness levels, adds or subtracts frequencies, and lengthens, shortens, or reverses the sound. All of this work serves one purpose; it matches the mental template the designer has of the sound that must be built.

The Recording Session

When a sound designer record a door slam, that designer must consider many variables that effect the recording: microphone type, microphone orientation and distance, amount and type of background noise, the acoustics of the space, the materials of the door and door frame, the nature of the latch, and the amount of force used to slam the door. The designer also considers the desired use of the particular sound. Perhaps the sound of the latch will be vital in the final context, so an extra mike might be added close to the latch. In the recording session, the designer makes sure that the raw materials are there for later manipulation, and the designer organizes the recording for the greatest flexibility later on in the editing and production process.

Caricature

Often, if a constructed sound cue fails to work despite the craft accorded each element, caricature becomes a useful paradigm. Exaggerating the most salient features of the sound under construction can help define it. For example, if the sound of a door is not being clearly read as a door, designers will often retailor the sound to include details such as a latch click and doorknob release. These sounds may even be laid into the sound file at an exaggerated loudness in order to emphasize their effect. In the same example, a more door-like sound might also have been achieved through the manipulation of the frequency content of the file, which would add more amplitude in the middle ranges (1000-3000 Hz). Such "equalization" will give the door a sharper sound. This effect can then be emphasized by shaping the amplitude envelope of the sound so that the door will strike sharply and with no fade-in at all. If the sound is still not satisfactory, another sound can be brought in and layered on top of the original sound (perhaps a hammer or croquet mallet strike). Adding artificial reverberation or a sharp slap echo both increases the definition of the sound and gives it a sense of place.

The Microstructure of Space

Achieving a sense of place in the auditory realm means establishing the illusion of consistant acoustic behavior among all sonic features. Again, the key word is illusion; many times a simple, low-resolution reverberation program will create the desired impression. As demonstrated above, good storytelling can achieve better believability than an accurate reproduction; the importance of storytelling holds true for auditory illusions of space.

For example, the interior of an ice cave is a place with a lot of acoustic reflection as is the interior of a cathedral. The character of the reflected sound is quite different in these two places due to differing surface qualities (which act as frequency filters), object placements (frequency and amplitude filters and added reflections), and the shape and the distance of reflective surfaces (walls, ceilings, and floors). By treating all the sonic objects in the entire environment with a good approximation of the appropriate kind of effect, the sonic story of a place is told efficiently. Manufacturers of musical equipment name their reverberation programs after the kinds of space they mean to replicate; Large Hall, Small Hall, Garage, Basement, Tunnel, Stadium, and Cathedral are all common names among the ready-made programs found in MIDI-controllable special effects processors. The designer can tailor these programs to a specific use by changing internal parameters such as how much delay a high-frequency sound will get as opposed to a low-frequency one, how many and how closely packed the reflections are, and so on. How much "dry" (unprocessed) sound gets mixed in with the "wet" (processed) sound is also a controllable real time parameter. In an interactive system, this mixing of sounds provides a very effective means of establishing the approach or the recession of sound-emitting objects. As an object making sound recedes, direct sound from that object diminishes and a more reflected sound is heard. This more-reflected sound is usually used in combination with a standard proximity-to-volume algorithm.

The Microstructure of Character: Design Techniques

The design mechanisms for giving the desired characteristics to a sound include identifying the physical components of those characteristics and suppressing things that detract from them. For example, if the sound has a great deal of dynamic range (very loud to very soft), it is likely that the softer parts will get lost either through accompanying sound effects or as a result of destination sound system resolution. An audio process called compression allows the soft parts to be brought up in level while keeping the louder ones at the same level. This compression process is an effect used in nearly all electronic media.

One problem with compression is that a certain amount of expressiveness is lost as dynamic range is lost. Audio compression is measured in ratios; while a radio announcer's voice may receive a 20:1 compression, a well-sung vocal track will usually not get more than a 4:1 compression. Compressed sound is known for its ability to make voices (or other sounds) "stand out" from the sonic background.

Compression is an example of a physical design variable that operates at the low-level frame of the waveform. The designer is literally changing the individual values of "samples" of sound stored in the computer. Each of the many effects available to the designer at this waveform level has its own set of design variables that interact with each other and with other effects. These effects include many kinds of equalization or frequency filtering, doubling or chorusing, pitch shifting, inversion, editing with crossfades, amplitude variation, and envelope shaping. The sound designer must track these relationships carefully; an aural "smudge" may be created from unintended conflict between effects.

Conclusion

As we integrate more sophisticated aural work into our interfaces and auditory displays, we should question how we obtain these sounds and how we should manipulate and combine them for the greatest possible effect. Simply using a sound taken straight from a CD library or recorded live on a portable DAT player fails to take advantage of the full richness of the medium. By thinking of the sound design task as telling a story, the sound designer can make a more detailed analysis of needed sonic details and decide on appropriate technical approaches. The finished design will present a richer and more consistent feel to the user.

References

Ballas, J. A. (1992). Delivery of information through sound. In Auditory display: Sonification, audification, and auditory interfaces Reading, MA: Addison-Wesley.

Bregman, A. (1990). Auditory scene analysis. Cambridge, MA : MIT Press.

Buxton, B., Gaver, W., & Bly, S. (1989). Non-speech audio at the interface. Book in manuscript; partially published as tutorial notes. In Nonspeech Audio, CHI '89 Conference Proceedings New York : ACM Press.

Cope, D. (1992). On algorithmic representations of musical style. In Understanding music with AI. Menlo Park, CA: The AAAI Press.

Gaver, W. (1989). The sonic finder: An interface that uses auditory icons. Human-Computer Interaction 4 (1).

Gaver, W. (1992). Using and creating auditory icons. In Proceedings of the International Conference on Auditory Display Reading, MA: Addison-Wesley.

Kramer, G. (1992). An introduction to auditory display. In Auditory display: Sonification, audification, and auditory interfaces Reading, MA: Addison-Wesley.

Mynatt, E. D. (1994). Designing with auditory icons. In
Proceedings of the 2nd International Conference on Auditory
Display Reading, MA: Addison-Wesley.

Schank, R. C., & Abelson, R. P. (1995). Knowledge and
memory: The real story. In R. S. Wyer, Jr. (Ed.), Advances in
social cognition, volume VIII. Hillsdale, New Jersey:
Lawrence Erlbaum Associates.

Maribeth Back and D. Des.
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304
(415) 812-4409
http://brainop.media.mit.edu/~mbb/