

Audiocentric Interface Design: A Building Blocks Approach

*Chad Thornton, Anthony Kolb,
Francine Gemperle*

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 12603
{dthornto, akolb, fg24}
@andrew.cmu.edu

Tad Hirsch

Massachusetts Institute of Technology.
77 Massachusetts Ave
Cambridge, MA 02139
tad@media.mit.edu

ABSTRACT

The advent of mobile, wearable, and ubiquitous computing presents opportunities for audiocentric interfaces that use sound as the primary or only means of displaying information to users whose eyes are otherwise engaged. While interface designers have a wealth of technological capabilities at their disposal for capturing, storing, transmitting, and displaying sound, there is a lack of appropriate resources to inform and inspire the design of compelling new audiocentric interfaces. This paper presents work towards developing guidelines for audio interface designers by developing a suite of interface “building blocks:” common interface elements that can be incorporated into the design of complex interfaces. Several audio progress meters and experiments in directing user focus in a spatialized audio environment are discussed.

1. INTRODUCTION

Interface designers have long recognized the value of sound in mediating human-computer interaction (HCI). Audio often augments graphical user interfaces (GUIs) to provide feedback to users, and to reduce cognitive load in complex operating environments. Audiocentric interfaces, in which sound is the primary or only information display, are less well established in HCI. However, their use is growing, fueled by the advent of mobile, wearable, and ubiquitous computing. As computing moves “off the desktop,” audiocentric interfaces are gaining importance for users who need eyes- (and hands-) free access to information.

While designers currently have a wealth of technological capabilities at their disposal for capturing, storing, transmitting, and displaying sound, there is a lack of appropriate resources to inform and inspire the design of audio-centric interfaces that are both usable and engaging. Much prior work in audiocentric interface design is application-specific, making it difficult to draw conclusions that inform the design of new interfaces. There is a need for resources to educate interface designers – many of whom are only trained as visual thinkers – of the possibilities and limitations of audio as an interface component.

To address this need, a team of interaction designers and engineers at Carnegie Mellon University is developing an audio interface designer’s guide. This guide draws upon prior work in psychoacoustics and audio interface design, and also incorporates original work in the design and development of audiocentric interfaces. There is a particularly emphasis placed on spatialized audio environments, as we believe that spatialization will play a

significant role in the design of next-generation audiocentric interfaces. The goal of this effort is to present the interface design community with useful information in a manner that is useful and appropriate to common design discourse.

2. RELATED WORK

Substantial prior research exists in audio interface design. The scope of this work has included sounds to complement graphical user interfaces (GUIs) [6], audio documents [9], factory simulations [7], and games [4]. [12] identifies many challenges facing audio interface designers, and along with [10, 2, and 5] suggests the value of appropriate metaphors for guiding interaction design. [9] discusses the role of aesthetics in audio interface design. While the majority of prior research has focused on auditory display, several novel proposals have been made to provide means of interacting with sound, including an “audio cursor” [11], and “query by humming” techniques [8].

Because much prior work is application-specific, it is difficult to make general recommendations for the design of new interfaces. Our approach is to develop widely applicable guidelines through the design of general-purpose interface components.

3. METHODOLOGY

We have taken a “building blocks” approach to design research. Our method has been to identify small, common interaction design problems, and design and evaluate several solutions for each. These solutions – and the lessons learned from producing them – can be then incorporated into the design of more complex interfaces. To date, we have addressed two “building blocks”: progress displays, and user focus.

For each solution produced, five users – all graduate students – were recruited for informal user testing. Think-aloud protocols and subject interviews were used to record their responses. Attention was paid both to usability issues and to users’ aesthetic experience.

A rapid prototyping environment was developed using Cycling74’s MAX/MSP package. Spatialization was achieved with the spat~ plug-in developed by the Institut de Recherche et Coordination Acoustique/Musique (IRCAM).

4. DESIGN

4.1 Audio Progress Indicator

Displaying progress of ongoing processes – such as transferring files, printing documents, or rendering images – is a common interface design challenge. Because users often require only peripheral awareness of progress, progress displays often exist in the background of the users’ attention. Such displays present a challenge for audio-only interfaces, where users’ ability to monitor multiple information displays is significantly less than in visual interfaces. It has been suggested that spatialization can enhance the effectiveness of an audio progress meter [13].

To simulate a realistic multitasking scenario, a National Public Radio news broadcast and audio feedback of an AOL Instant Messenger chat were placed 60° apart, directly in front of the user. Three audio progress meters simulating the printing of a large document were then designed and placed in the environment (Figure 1). The meters also used motion as a design element – sound moved in 180° arc in front of the user.

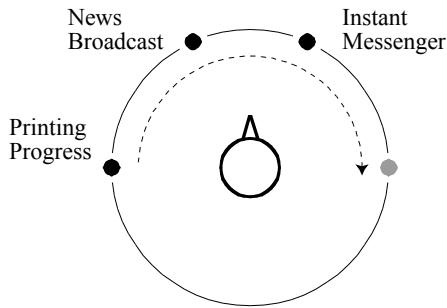


Figure 1: An audio-centric multitasking scenario

Progress Meter 1: Water

In our first prototype, the sound of water pouring into a container was chosen to indicate progress, because it is harmonically rich, elegantly simple, intuitively understood, and emotionally pleasing. Using a Sennheiser 421 microphone connected to an M-Audio Duo USB Preamp and recording directly to hard disk, we recorded a number of containers being filled with water. Microphone placement, pouring speed, the container’s material and shape, and various pouring techniques were explored to create several dozen examples of water filling a container.

Progress Meter II: Diatonic scale

For our second design, we chose an ascending diatonic scale (do-re-mi-fa-so-la-ti-do). Like the sound of water filling, we felt that an ascending diatonic scale was simple, intuitive, and pleasing. Using a scale also allows for longer progressions, as notes can be repeated (do, do re, do re mi...). Using two software-based synthesizers – Absynth by Native Instruments and Reason by Propellerheads – we played a diatonic scale using a number of instruments, with and without major chord flourishes to signal completion.

Progress Meter III: Bouncing ball

Sound also can be an indication of physical phenomena, such as the size and weight of a metal hammer striking a bell [2]. For our third prototype, we simulated the effects of gravity on a bouncing ball, shortening the time between bounces as it tended towards settling on the ground. Again,

our expectation was that this would provide a simple, intuitive, and pleasing audio experience.

4.2 User Focus

Directing user focus is another common challenge facing interface designers. In a complex operating environment, a user multitasks between several applications and/or interacts with multiple information sources simultaneously. “Focus” refers to which of several information objects a user attends at a given time. In visual interface design, there are various standard techniques for directing user focus, including layout (for example, placing important information in the middle of the screen), layering (placing “in focus” objects on top of other objects), and highlighting (including, for example, the use of color, size and/or motion to call a user’s attention to a particular piece of information).

For audiocentric environments, we hypothesized that focus might be indicated through placement, volume, and filters that alter the timbre of a sound. Several solutions were implemented with a Griffin Technologies rotary dial provided a physical control over the audio environment.

Focus I: Position

Though we hear from all directions, we generally orient ourselves toward what currently holds our attention. Sounds that are directly in front may then be presumed to hold our focus. The three sounds were placed 120° apart on a circle around the head (Figure 2a). The rotary dial allowed a user to rotate the circle of sounds, controlling which sound was front most, and thereby in focus.

Focus II: Amplitude and Filter

Inspired by Cohen’s notion of “filtears” [1] – subtle cues applied to sounds to convey additional information – we manipulated amplitude and timbre to indicate focus. The three sounds were again placed 120° apart, but remained static while moving the dial rotated an arc along the circle that increased the amplitude and removed a low-pass filter of any sound it was located upon, thereby indicating focus (Figure 2b).

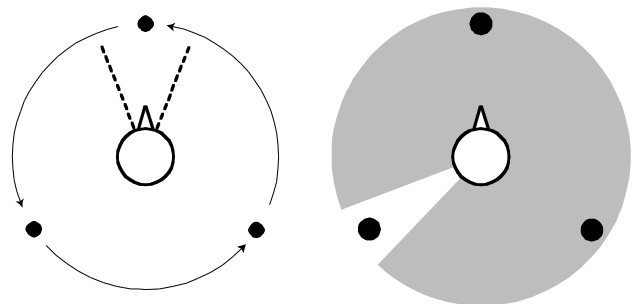


Figure 2: Focus through (a) position or (b) filter

Focus III: Interaction

In comparing the “rotating sounds” and the “rotating filter” approaches, we also considered the dynamics and feedback of the rotation. We prepared one prototype that allowed for fluid, unrestricted rotation when using the dial, and another that discretely moved from one of the three locations around the circle directly to the next (Figure 3). A beep was also tested as feedback to indicate target acquisition in the discrete approach.

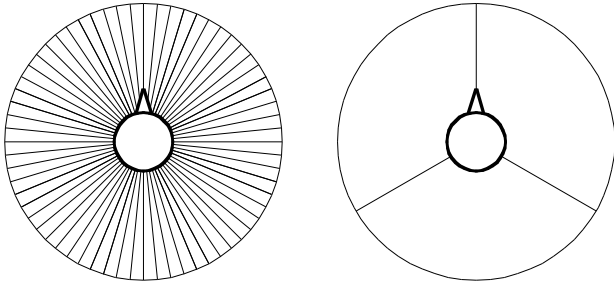


Figure 3: (a) Fluid or (b) discrete rotation

5. RESULTS

Our first concerns were with the effectiveness of the spat~plug-in, and users' ability to perceive motion and spatialization. We found the prototyping environment to be largely effective, although a sound moving in a 360° circle seemed to move much more quickly around the back 180° than it did when crossing the front 180°.

The system suffered from similar limitations as other spatialized audio environments [14]. For example, changes in distance from the listener were difficult to notice - a sound moving away was usually indistinguishable from the same sound not moving but decreasing in volume. Elevation was also largely ineffective. It was difficult to notice the effects of an elevation change with one sound, and impossible when multiple sounds were playing. However, we found that noticing a static pitch changing in elevation was easier when another reference pitch remained parallel to the listener. This effect may be worthy of further investigation.

Our study focused on motion along arcs centered around the listener and along the same horizontal plane. Listeners hearing the sounds - two non-moving and one moving in a 180° arc - could tell that a sound was moving, which sound was moving, and indicate the direction and general speed of movement. Consistent with previous findings, localization was fairly general [3].

All users were able to accurately localize the various sounds. However, several felt that the axis around which the sounds were rotating was tilted slightly - sounds in the front were described as passing "over the top" near the forehead, and in the back as "down below", near the base of the neck.

The duration and basic rhythmic pattern of sounds affected the perception of movement. Listeners tended to have a better sense of movement with persistent sounds such as water running and vibraphones sustaining than they did with discrete, repeated patterns such as the ball hitting the floor. The sustaining sounds were stronger in attracting and maintaining a user's attention as well.

Users also had an acute sense of the mapping between the physical dial and the audio response, and they quickly figured out how fast they could turn the dial and still expect a response.

5.1 Audio Progress Indicator

Water

Users immediately identified the sound as water filling a glass and understood that fullness signaled completion. All users liked the sound and had a good sense in advance of

roughly when the glass would be full. One user noted that it was easy to focus on without needing to be too loud. Several mentioned their ability to relate to making the sound and being able to fill a glass by sound alone. Given the choice, about half of the users felt this was their preferred sound for indicating progress.

Diatonic scale

On first listen, all users understood the progression by the third note, and had a good sense of how long the progression would take to complete. Users also understood both the short version (do re mi fa...) and the longer version (do, do re, do re mi...) of the indicator. All users felt that a scale was a good way to indicate progress. Most users preferred instruments that had longer sustains, such as vibraphones. About half of the users preferred the diatonic scale to the other progress indicators. One user felt that the scale was the best complement to the type of movement used.

Bouncing ball

Users noted that the sound also gave a sense of progress to completion. However, some listeners felt that the speeding up of the bounces as it settled was somewhat anxiety inducing. Users were also less certain about how long the sound would last. It was the least favored of the three.

5.2 Focus

The quality of particular sounds affected user attention. When several types of sounds were heard in concert, speech and music (particularly that containing vocals or drums) gathered the most attention. Sounds that were infrequent, erratic, or unfamiliar also strongly attracted the listeners' attention.

The relative volume levels amongst sources did have a strong effect in indicating persistent focus. Changes in volume to indicate a change in focus were sometimes effective, though there were some conflicts between the amplitude effects used to generate the head-related transfer function (HRTF) and the amplitude effects that attempted to indicate a focus change.

Several users also expressed the desire to have the ability to "put away" a sound, referring to away as either a change in volume or a change in position resulting in a volume change.

Position vs. filter rotation

Changing the positions of sounds, especially coupled with fluid rotation and higher overall volume, was initially confusing for one user and dizzying for another. Others had less trouble, and all quickly began rotating sounds to the front for a better listen without prompting.

All users perceived the filter effect to be a change in volume; most also noted that one sound was "more clear" or "less muffled" than the other two. This seemed to be especially noticeable when speech was present in the audio. Several users mistakenly perceived a change in motion when the filter effect was used. One user thought that the front-most sound remained in place, while the back two were orbiting around an axis no longer centered on her head. This may have been an unintended consequence of using a low-pass filter for the study, and may not be an issue with other effects, such as reverb.

There was no clear preference amongst users for either the position or filter-based rotation. Users who were initially

disoriented by the rotating sounds prototypes preferred the rotating filter prototypes. All understood the change in position or filter to be a means of directing focus without it being explained to them. They were generally impressed by the effects and most ruminated on how they might incorporate it into their current computing platform. The system appeared to be most effective when sounds were more consistent, or when auditory feedback accompanied the discrete rotation.

Discrete vs. fluid rotation

Users appeared equally adept at using both systems. However, users generally reported a decreased sense of spatialization when using discrete motion to change the position of or the filter on the three sounds. All users listening to the discrete rotation prototypes preferred a light beep of auditory feedback to accompany the change, though they felt that the feedback would be more useful if localized.

6. CONCLUSIONS & FUTURE WORK

The approach taken offers interesting considerations to auditory interface researchers and designers.

1. A design challenge generally has a number of solutions. The results of our testing on these early-stage prototypes indicates that there are often several tenable solutions to a given auditory interface problem. A competent designer has the ability to recognize the issues, understand the possibilities, and design an appropriate solution. For audio interface research that is intended to inform the design community, it appears that demonstrating multiple solutions to a design problem may be a more effective means of conveying the possibilities for design than reporting on a single implementation.

2. Aesthetics matter. Users all commented on the types, and sometimes the qualities, of the sounds that were presented. While we placed more emphasis on aesthetics than most auditory interface research, there remains a need for more sustained, ongoing inquiry into the role of aesthetics in auditory interfaces – both the sounds and the means of interaction. Complex interactions involving multiple sound elements, multiple simultaneous sounds, and the means to interact with those sounds create a more lively and compelling experience for users, but it's these complex interactions that are also the least understood. The quest for harmonically rich, elegantly simple, intuitively understood, and emotionally pleasing sounds is likely to show even greater dividends in longer-term studies on use of auditory interfaces.

7. REFERENCES

[1] Cohen M. Throwing, Pitching and Catching Sound: Audio Windowing Models and Modes. *Journal of Man-Machine Studies* 39, 1993.

[2] Beaudouin-Lafon, M, Gaver, W.W. "ENO: Synthesizing Structured Sound Spaces," *UIST'94*.

[3] Begault, D. "Auditory and non-auditory factors that potentially influence virtual acoustic imagery," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999.

[4] Drewes, T.M., Mynatt, E.D., Gandy, M. "Sleuth: An Audio

Experience," *ICAD'00*.

[5] Fernström, M., Brazil, E., "Sonic Browsing: An Auditory Tool for Multimedia Asset Management," *ICAD'01*.

[6] Gaver W. *Auditory icons: Using sound in computer interfaces*. Human-Computer Interaction, 1986; 2: 167-177, Lawrence Erlbaum Associates, Inc.

[7] Gaver, W., W. and Smith, R., B. and O'Shea, T., "Effective Sounds in Complex Systems: The ARKola Simulation", *CHI'91*, April 28-May 2, 1991.

[8] Ghias, A. et al "Query by Humming" *ACM Multimedia'95*.

[9] Goose, S., Möller, C. "A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure," *ACM Multimedia'99*.

[10] Hindus, D., Arons, B., Stifelman, L., Gaver, B., Mynatt, E., Back, M. "Designing Auditory Interactions for PDAs," *UIST'95*.

[11] Kobayashi, M., Schmandt, C. "Dynamic Soundscape: mapping time to space for audio browsing" *CHI'97*.

[12] Mynatt, E.D. and Edwards, W.K. "Metaphors for Nonvisual Computing," in *Extraordinary Human-Computer Interaction*, Edwards, A.D.N (ed) Cambridge University Press, 1995.

[13] Walker, V.A. and Brewster, S.A. "Spatial audio in small screen device displays". *Personal Technologies*, 2000, 4, 2.

[14] Wenzel, E. "Localization in Virtual Acoustic Displays," *Presence*, 1:1, 1992.