An Argument for Using Techniques Found in Audio Games as the Next Generation of User Interfaces for Blind People
Chris Hofstader
December 2005

Introduction: Reality-based Interfaces and the Blind Computer User

The majority of blind people who use computers (including the author of this position paper) access the information on their device using a program typically called a screen reader. Some users, mostly on the GNU/Linux platform, use software that augments the text-based console interface or a text editor to access their information[1]. For the purposes of this paper, though, I will treat them identically as the interface to both classes of access technology are very similar and neither is based in anything resembling what a human might regard as reality.

The interface to screen readers (whether *JAWS*, the most popular in the world, or *VoiceOver*,, the newest entry to the market) which will be described in greater detail later in this paper, is a one-dimensional stream of words spoken by a text to speech engine[2]. So, unless one's reality is in a single dimension, the interface for blind people has no basis in human experience.

There are, however, pioneers of three-dimensional user interfaces making and selling audio-only programs[3] to blind users. These people come from the world of audio gaming and have advanced to a point in which blind people can, in real time, navigate through complex environments similar to those enjoyed in video games by millions of sighted people[4].

The position described in this paper is that the techniques used in the more advanced audio games should define many aspects of the next generation of interfaces for blind people in Computing, orientation and mobility technologies and smart spaces/environments.

Inefficiency: The Achilles' Heel of Screen Readers

Screen reader vendors have made improvements to the usability of their products over the last decade[5]. *JAWS*, the most popular screen reader, has added hundreds of features that a blind person can employ to use the more advanced functions of the *Microsoft Office* and

---

[1] The Blinux Project, http://www.leb.net/blinux/.

[2] Many screen readers also offer the user the ability to have text output to a refreshable Braille display. For the purposes of this paper, though, all tactile I/O devices will be ignored.

[3] Wikipedia, the Free Encyclopedia, "Audio Game History", *Audio Games* (November 8, 2005) http://en.wikipedia.org/wiki/Audio_Games#Audio_game_history.

[4] Cohn, David, "The Blind Fragging the Blind", *Wired News* (14 March 2005) http://www.wired.com/news/games/0,2101,66879,00.html.

[5] Virtually every review of a new release of a screen reader that one can find in the assistive technology press makes mention of new features and enhancements to the product they are writing about.

Internet Explorer suite of products. *JAWS* has also added a number of keystrokes that make document navigation simpler in some of its more recent releases.

Apple Computer's *VoiceOver,* which has been poorly received by critics[6], does get applause from some users for being simpler to use than the more complex Windows products. *Voiceover's* simplicity, unfortunately, comes at the expense of the many features that blind users have grown accustomed to having available to them. Its ease of use does not, therefore, lead to improved efficiency as many of the features in a product like *JAWS* designed to increase efficiency are simply not present.

Fundamentally*, JAWS*, *Window-Eyes* and *HAL*, the three most popular screen reading tools, no matter what features they add to increase a user's efficiency, can only deliver one little piece--either a syllable or pause--of information at a time. Thus, the interface presented to a blind user is linear and the information received at any moment is merely a single point on a long string of auditory tokens.

To access a specific piece of information, a blind user must wait for the point on the line that contains the bit of information he is interested in hearing to be spoken by the screen reader. Certainly, blind users can employ navigation commands to jump from one point to another on this line of information, but doing so without jumping past the point one is seeking can be difficult for even the most experienced users.

For lack of another term, I call the amount of time it takes to deliver one of the aforementioned auditory tokens a gTick (grammatical tick). One gTick is equal to the average amount of time at a given speech rate (each user will have her own level of comfort in words per minute) it takes a speech synthesizer to deliver a single syllable plus the amount of time a speech synthesizer pauses for white space and different types of punctuation (following the rules of the human language the synthesizer is speaking). Thus, the most productive a blind person can be using one of these screen reading tools is measured in the total number of gTicks that pass in the time allotted for their task as, using a purely speech interface, they are not presented with more than a single gTick of information at a time.

Audio Games and the 3D Interface

Many blind people enjoy the game *Shades of Doom* (SOD) from GMA Games. SOD presents a complex environment where the user (just like the video game with a similar name) must navigate in real time a sequence of mazes, fight off mutants, find useful items and, ultimately, find his way to the center of the secret laboratory and kill the mad scientist before his experiment escapes and turns all of humanity into mutants[7].

---

[6]Leventhal, Jay, "Not What the Doctor Ordered: A Review of Apple's VoiceOver Screen Reader", AccessWorld, September 2005.
(http://www.afb.org/afbpress/pub.asp?DocID=aw060505)
[7] David Greenwood, "The Shades of Doom Status Report", *Audyssey Magazine*, Issue 21 (January-February 2000)

The people who play this game must navigate by the sound of the wind (always out of the east), the echo of their own footsteps against walls, the sound of their own breath and other ambient sounds which, after a while, the player learns to locate in relation to their current position. The author of the game claims that the player is presented with up to 32 simultaneous sounds[8] which they must decipher to perform the various tasks, navigate the maze and fight off the mutants.

Shades of Doom presents multiple simultaneous audio data streams. *JAWS* and its counterparts present a single stream of information. Hundreds, maybe thousands, of game players enjoy SOD and the other entries from GMA. Approximately 150,000 blind people worldwide use *JAWS* to access information on Windows-based computers. How is it that the gamers can process so much information at once, but the screen reader users are bound by the tiny bit of information expressed in a single syllable or pause?

The amount of semantic information delivered by a single sound in SOD in a single gTick is actually greater than that delivered by a speech interface. A player who hears a particular "bloop" which takes about as long as a syllable to present, knows (after a little practice with the game) that there is an object of use to them nearby. For *JAWS* to speak, "there is an object of use nearby" requires 15 gTicks (9 syllables and six spaces).

There is a fair amount of research by cognitive psychologists about the number of audio data points a person can comprehend simultaneously[9]. There has also been work studying how humans perceive the location of an object presented by a three-dimensional interface[10]. There is no research, to my knowledge, that suggests how many data streams a blind person can use effectively to perform a given task on their computer, to find an object in their home, or to more effectively move independently through real world spaces.

Applying 3D Interfaces for Use By Blind People

The game developers have certain advantages over people who design interfaces for screen readers. Specifically, they know the narrative that they are presenting and so can predict, within a small set of actions, what the user will do next. A screen reader cannot predict what word will appear next in an email, what value will be present in a random cell in a spreadsheet or which link on a web page the user will select next.

---

http://www.audiogames.net/page.php?pagefile=_Audyssey_Magazine_issue_21_-_January_-_February_2000_.

[8] Referenced in the games tutorial information and on the GMA Games web site, http://www.gmagames.com.

[9] Bregman, Albert S., Auditory Scene Analysis: The Perceptual Organization of sound. Cambridge, Massachusetts: The MIT Press, 1990.

[10] Begault, Durand R., 3-D Sound for Virtual Reality and Multimedia. NASA technical report, Moffett Field, California: 2000.

A sighted computer user can learn a number of bits of information about the state of their computer by merely diverting their glance from one portion of the screen to another. Using three-dimensional audio an interface can be built that, by placing audio objects in different locations in a simulated spatial environment, will deliver far more simultaneous information to a blind person. These techniques, if applied correctly, can greatly improve the efficiency of computer use for those who rely on an audio interface.

3D Interfaces for Non-Computer Applications

Although I have not tested these concepts yet, I can envision a blind person using 3D audio to help them locate objects in a smart space. If, for instance, one keeps their keys attached to a fob that contains an RF tag and the room has an array of RF readers the blind person can ask their user agent, "Where are my keys?" And, through directing the audio in the space be presented with a tone that sounds as though it is emulating from the spot where the RF system has located the object of interest.

Orientation and mobility technologies can likely be greatly improved if expanded to include a third dimension. There are a wide variety of technologies ranging from sonar to infrared to laser range-finding equipment that can report a tremendous amount of information about the solids in an outdoor environment. Delivering this through a TTS in a single dimension would take an overwhelming amount of time. In sound, though, like the sounds of walls and other objects in *Shades of Doom*, a portable 3D tool can shape the surroundings for the blind person as they walk. This type of interface will provide a terrific amount of information about obstacles that one might miss with a cane.

Conclusion

I could provide many examples of how 3D audio interfaces can be used to increase the productivity of blind computer users and to improve safety and quality of life at home. For the most part the audio portion of the software engineering is done science, as Microsoft and the other game platform developers have figured it out for the very lucrative video game market. What the HCI people, including me, now need to do is to discover how to do the specific implementations for blind people so they can be increasingly productive, safer and have greater conveniences in their homes in the future.