Sonically-Enhanced Tabular Screen-Reading

Jonathan Cofino, Armando Barreto, Fatemeh Abyarjoo, Francisco R. Ortega

Florida International University

jcofi001@fiu.edu, barretoa@fiu.edu, fabya001@fiu.edu, forte007@fiu.edu

Abstract

The World Wide Web has made more information readily available than at any time in human history. This information is often presented visually, which can be an inaccessible medium for people with blindness or low-vision. Presently, screen readers are able to verbalize on-screen text using text-to-speech (TTS) synthesis. However, much of this vocalization is inadequate for browsing the Internet, as it cannot properly convey the structure and relationships that exist in a visual presentation. We have created and tested an auditory interface that incorporates auditory-spatial orientation within a tabular structure. When information is structured as a two-dimensional table, links can be semantically grouped as cells in a row within the auditory table, which provides a consistent structure for auditory navigation. Our auditory display prototype was tested with sixteen legally blind participants, who each navigated four sonified tables enhanced with prepended tones, which were varied with stereo spatialization and tonal variation. The sonified tables were presented in a randomized order to avoid ordering/learning effects. Results from the experiment showed that stereo panning was an effective technique for audio-spatially orienting non-visual navigation in a five-row, six-column HTML table as compared to a centered, stationary synthesized voice.

Keywords

Screen-reading; sonification; assistive technology; tabular navigation; accessibility

Introduction

Screen-reading software has made it possible for many individuals who have lost their vision to use computers. Using dynamically created synthesized speech, this invaluable tool provides a portal to the world of e-mail, news, banking, entertainment, and countless other activities that have migrated to the digital world. Internet access is now ubiquitous and often necessary for gainful employment, and the World Wide Web (WWW) has connected communities across the world. The preeminence of the WWW in the daily lives of many individuals cannot be overstated. However, the WWW is primarily a visual medium, and the nature of this visual content makes it difficult to render in the inherently sequential speech modality of the contemporary screen-reader. When a screen-reader renders visually designed content, spatial orientation and relationships are often lost or disregarded, leaving the blind user to perform complex searches and organizational tasks. To make matters worse, visual content is often arrayed in inconsistent and arbitrary configurations, which can be confusing to the blind user. Much of the Internet is now an exercise in navigation, and this navigation is often intended to be visual. To fully include all potential users, the WWW must be accessible to auditory navigators using screen-readers.

Related Work

Tabular Sonification

Researchers have investigated various modalities of presenting information non-visually. The auditory icon technique involves sonic metaphors posing for visual icons (Gaver). For example, a “cow” icon would sound like “moo.” The “earcon” technique involves composing melodies/rhythms to indicate information (Blattner, Sumikawa, and Greenberg). These two sonification techniques, while often useful, have inherent limitations. Auditory icons are not appropriate for purely visual icons like a book, or functions that have no obvious sound, like "file...save...as." Composed melodies and rhythms are arbitrary and do not necessarily reflect a meaningful representation for the user.

Researchers have proposed using pitched tones to represent numerical data in a spreadsheet (Ramloll et al.). Considering that each row or column of a spreadsheet is a sequence, the data in that sequence could be proportionally scaled and shifted to integer values. These integer sequences can then be sonified such that when played as successive tones, the sequence is perceived as a melody. This is an ingenious method for non-visually portraying trends in numeric data; however, its application is not easy to generalize, as textual data has no obvious analog in pitch.

A method for audio browsing that involves both 3-D auditory spatialization and metaphorical sonifications was proposed (Goose and Möller). In their implementation, users hear space-ship lift-off and landing sonic metaphors to connote intra-document linking while stereo spatialization conveys the browser's relative focal position within the document. This serves to maintain orientation while giving a sense of overall document length.

Blind individuals, on average, have a significantly better ability to localize sound than do sighted people (Ohuchi et al.). This ability was found by arranging a circular array of loudspeakers around both blind and sighted individuals who were then asked to indicate the speaker from which a test sound emanated. This capability inspired the stereo spatialization technique implemented in this research.

Screen-Reader Optimization and Webpage Reformatting

Researchers have explored automated semantic organization techniques for making the Web more accessible to blind users (Nagao, Shirai, and Squire). These techniques look for similar content to group together. Gupta *et al.* have proposed parsing HTML documents to strip out extraneous content based on the Document Object Model (DOM) tree structure inherent in HTML documents (Gupta *et al.*). Advertisements, navigational bars, and empty tables can be stripped out to provide a concise presentation for the non-visual navigator. Our spatialized tabular browser presupposes that content can be semantically arranged into a tabular format for a consistent, predictable layout.

Methods

Subjects

We recruited sixteen research study participants from a pool of legally blind screen-reader users. Our implementation assumed no previous knowledge of existing screen-reader software. The participants ranged in age from 20 to 57 years old, with a mean age of 43.17 (SD = 12.73) years. They were recruited from various Lighthouses in the greater South Florida area.

Spatialization Techniques

An inherent limitation in screen-reader presentation is verbosity. We seek to eliminate some verbosity by replacing verbally explicit descriptions of cellular focus with dynamic relative spatialization. We aimed at leveraging two fundamental properties of human psychoacoustics, namely stereo localization and pitch perception, to accomplish sonic spatialization.

Stereo Panning

The psychoacoustic phenomenon of sonic localization can be exploited to artificially localize sound sources. Blumlein discovered that by simultaneously playing a sound from two loudspeakers at a differential in amplitude, a person will perceive the sound as originating from somewhere along the continuum of space between the two loudspeakers. The perceived location of origination is relative to the ratio of amplitudes, where the speaker with greater amplitude will drag the virtual sound source closer to it from the middle. The technique of varying the perceived sound location is called stereo panning. The synthesized speech output from the screen reader can be panned linearly to indicate the location of a text item along the horizontal dimension of the table.

Tonal Variation

Humans can perceive changes in sonic frequency as distinct pitches on a one-dimensional continuum. Musical scales, or series of notes progressively increasing in frequency, can be exploited to correspond to sonic locations. A tone, prepended to the TTS representation of each item in a table can be manipulated to achieve graduated relative auditory orientation. This prepended tone also serves to alert the user-navigator that she is focused on a link.

Other Sonic Indicators

We utilized sound effects to indicate navigational status, such as an “electrified fence” sonic metaphor for reaching boundaries and audible buzzes to indicate erroneous key presses. These non-verbal sound effects are critical to indicate navigational status, error, and events.

Procedures

Table 1. Grocery Navigational Table

Vegetables	Carrot	Potato	Cucumber	Onion	Lettuce
Fruits	Banana	Apple	Lemon	Orange	Cherry
Bakery	Bread	Cake	Pie	Cookie	Muffin
Meat	Beef	Chicken	Pork	Turkey	Duck
Drinks	Beer	Juice	Milk	Soda	Tea

An online purchasing task was implemented to test the feasibility of sound-spatialized tabular browsing. Each sound spatialization technique was implemented in a separate navigational table (A, B, C, and D), each consisting of five categorical rows and six total columns, serving as a consistent, predictable structure (see Table 1). The relatively small number of columns and rows was influenced by research suggesting that an average person can store “seven +/- two” items as a working memory (Miller). Each user was asked to navigate each navigational table, using the variations in stereo panning and/or changes in the prepended tone as navigational references. We tested the techniques according to a double-binary design (see Table 2).

Table 2. Sonic Augmentation and Enhancement Methods

Stereo/Tonal	Constant	Varying
Stationary	**A**	**B**
Mobile	**C**	**D**

We trained our participants to familiarize them with tabular navigation. Participants were able to use shortcut keys to immediately access either end of a row. Keystrokes were provided to recall the prompted information as well as to recall the order of the category-rows (see Table 3).

Table 3. Keyboard Commands

Key	**Mnemonic**	**Function**
c	categories	repeats all of the category headers
e	end	focuses to the rightmost cell in the current row
r	repeat/recall/remind	prompts user to find a particular link
s	shortcut	focuses to the leftmost cell in the current row
spacebar	N/A	selects a link
left-arrow	left	shifts focus to the immediate left
right-arrow	right	shifts focus to the immediate right
down-arrow	down	shifts focus down one cell
up-arrow	up	shifts focus up one cell

Data Collection

In each of the navigational tables (A, B, C, and D), the participant was verbally prompted to find ten items, in sequence. Primary data such as the time-to-target (TTT) and number of moves were recorded, as well as ancillary data such as number of boundaries reached, wrong links selected, and wrong keys pressed. This ancillary data serves to indicate confusion and disorientation, which may or may not be a result of the system itself.

Results

Quantitative Analysis

As each participant browsed through the navigational tables, his/her navigational data were recorded, such as the TTT and the number of moves needed to reach the stated target. While these two aspects of navigation are readily recorded and understood, it is less clear how to compare navigational data from one user to another as well as from one navigation task to another. The user's mental model heavily influences non-visual screen reader usage (Kurniawan, Sutcliffe, and Blenkhorn). Each of the navigations was unique in that the user must use her mental model of the table to map out a course toward the intended target. For blind navigators, this path is not straightforward. Many blind navigators will choose to return to the category-column (left-most column of each table), move vertically to the appropriate category-row (e.g. "Fruits" of the "Grocery" table), and then navigate across to the intended cell (e.g., "orange"). There is no ideal path. The existence of a shortcut to the category-column complicates the notion of a shortest or ideal path.

To meet the normality and equality of variances assumptions of ANOVA, we transformed our experimental TTT data using the natural logarithm. The results show that the logarithmically transformed TTT was significantly affected by the tonal variation method, $F(1,15)=6.194$, $p=0.025$. The results show that the logarithmically transformed TTT was marginally affected by the stereo spatialization method, $F(1,15)=4.240$, $p=0.057$. The results show that the logarithmically transformed TTT was not significantly affected by the interaction of both methods, $F(1,15) = 1.381$, $p=0.258$. These results suggest that some confusion may be caused in the subject when employing both of these methods simultaneously.

Qualitative Feedback from Subjects

Based on oral interviews conducted with the participants, many users remarked that they would be more likely to engage in spreadsheet editing and on-line purchasing if tabular navigation were streamlined and presented in an audio-spatial modality.

Several subjects reported that they reflexively moved their heads to follow the perceived origin of the moving voice. This clearly shows that the stereo spatialization is readily perceived and intuitive. There was a mixed reception to the pitched tones; some users noticed them and made use of them, while others simply ignored them. Some users noted that their sense of hearing was weak in one ear and that the tones were more effective for auditory guidance.

Discussion

Noting the significant effect of stereo spatialization on tabular navigation, we would strongly encourage web designers to consider a tabular layout as a blind-accessible alternative to conventional visually oriented web pages. While there has been progress in automating the conversion of arbitrarily arranged content into semantically structured web-pages, we believe that having a tabular format in mind from the onset would benefit the non-visual browser.

Our design philosophy was to keep everything as simple and intuitive as possible. We opted to use generic, off-the-shelf stereo speakers. Headphones, which vary in style and shape, were not considered, as they tend to create the psychoacoustic perception of hearing sound move within one's own head and restrict a blind person from remaining aware of her ambient surroundings. While a 3-D spatialized sound environment may have been novel and allowed for far more possibilities of sonic localization, the availability of such systems is prohibitive to a user who must be able to access her computer on the go. 3-D sonic spatialization would require a highly customized implementation of hardware and software not available to the typical user.

We used what we believed to be intuitive features of auditory perception: sonic localization and relative pitch. These intuitive features should then not need extensive training and explanation. Through the course of testing, a few of the most experienced screen-reader users remarked that they ignored the tonal guidance to focus better on the navigational task. While this was an unexpected reaction, it does indicate that those who wish to ignore it can tolerate the tones without great annoyance. Some other users stated that their hearing had deteriorated in one ear relative to the other and, consequently, they would prefer the tonal variation. For more effective implementation, users must be given a choice to enable one or both of the sonic enhancements to adjust for limitations in pitch perception and sonic localization.

Conclusions

The marginally significant effect of stereo spatialization supports our expectation that the presence of auditory enhancements could have an impact on the efficiency and comfort experienced by the experimental users in navigating the tables. This highlights the potential benefits from solely implementing a simple form of stereo panning that do not require any hardware changes. Fortunately, the minimal software modifications needed would be transparent to an end-user of this approach.

The results obtained in the experiments and associated statistical analyses have revealed that the tonal variations we applied to indicate spatial relations between cells were not as intuitive and easy to assimilate, as we would have hoped. We speculate that this may be due to the mental effort necessary to map pitch change to spatial displacement during the browsing of our navigational tables. Musical training may be necessary to appreciate tonal variation fully.

Noting the lack of help provided by the tonal variation method, the interaction of the combined enhancements did not mutually complement each other, as we would have hoped. Since the tonal variation method itself does not seem to be assimilated well, it does not enhance or complement the stereo spatialization method. Rather than making navigation more efficient, the combination of the two methods seems to add to the listener-navigator's cognitive burden.

Acknowledgments

The authors of this paper wish to thank Dr. Elly du Pré and Eric Barrette of the Lighthouse of Broward County, the Lighthouse for the Blind of the Palm Beaches, and the

Disability Resource Center at Florida International University. SoundManager2, the JavaScript API coded by Scott Schiller, was instrumental in the implementation of this research. National Science Foundation grants HRD-0833093 and CNS-0959985 sponsored this work.

Works Cited

Blattner, Meera M, Denise A Sumikawa, and Robert M Greenberg. “Earcons and Icons: Their Structure and Common Design Principles.” *Human–Computer Interaction* 4.1 (1989): 11–44. Print.

Gaver, William W. “The SonicFinder: An Interface That Uses Auditory Icons.” *Human–Computer Interaction* 4.1 (1989): 67–94. Print.

Goose, Stuart, and Carsten Möller. “A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure.” (1999): 363–371. Print.

Gupta, Suhit et al. “DOM-Based Content Extraction of HTML Documents.” *WWW ’03: Proceedings of the 12th International Conference on World Wide Web*. ACM, 2003. Print.

Kurniawan, Sri Hastuti, Alistair G Sutcliffe, and Paul Blenkhorn. “How Blind Users’ Mental Models Affect Their Perceived Usability of an Unfamiliar Screen Reader.” 3 (2003): 631–638. Print.

Miller, George A. “The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *Psychological review* 63.2 (1956): 81. Print.

Nagao, K, Y Shirai, and K Squire. “Semantic Annotation and Transcoding: Making Web Content More Accessible.” *MultiMedia, IEEE* 8.2 (2001): 69–81. Print.

Ohuchi, Makoto et al. “A Comparative Study of Sound Localization Acuity of Congenital Blind and Sighted People.” *Acoustical Science and Technology* 27.5 (2006): 290–293. Print.

Ramloll, Rameshsharma et al. “Using Non-Speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users.” (2001): 515–529. Print.