

Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data

HAIXIA ZHAO, CATHERINE PLAISANT, and BEN SHNEIDERMAN

University of Maryland

and

JONATHAN LAZAR

Towson University

We describe the development and evaluation of a tool, iSonic, to assist users with visual impairment in exploring georeferenced data using coordinated maps and tables, augmented with nontextual sounds and speech output. Our in-depth case studies with 7 blind users during 42 hours of data collection, showed that iSonic enabled them to find facts and discover trends in georeferenced data, even in unfamiliar geographical contexts, without special devices. Our design was guided by an Action-by-Design-Component (ADC) framework, which was also applied to scatterplots to demonstrate its generalizability. Video and download is available at www.cs.umd.edu/hcil/iSonic/.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (non-speech) feedback, Evaluation / methodology, User-centered design*

General Terms: Human Factors

Additional Key Words and Phrases: Interactive sonification, auditory user interfaces, information seeking, users with visual impairment, universal usability

ACM Reference Format:

Zhao, H., Plaisant, C., Shneiderman, B., and Lazar, J. 2008. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Trans. Comput.-Hum. Interact.* 15, 1, Article 4 (May 2008), 28 pages. DOI = 10.1145/1352782.1352786. <http://doi.acm.org/10.1145/1352782.1352786>.

This work is supported by the National Science Foundation under Grant No. EIA 0129978 and ITR/AITS 0205271.

Authors' addresses: H. Zhao, Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043; email: haixia@google.com; C. Plaisant, B. Shneiderman, Department of Computer Science, University of Maryland, College Park, Maryland 20740; J. Lazar, Department of Computer and Information Sciences, Towson University, Towson, Maryland 21252.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from the Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permission@acm.org.

© 2008 ACM 1073-0616/2008/05-ART4 \$5.00 DOI: 10.1145/1352782.1352786. <http://doi.acm.org/10.1145/1352782.1352786>.

ACM Transactions on Computer-Human Interaction, Vol. 15, No. 1, Article 4, Pub. date: May 2008.

1. INTRODUCTION

Information visualization is a technique that enables people with normal vision to use their tremendous visual ability to explore data and discover trends. However, when information is visualized using techniques such as color and spatial location, users with visual impairments do not easily comprehend the information. In addition, interaction with information visualizations is typically accomplished by using pointing devices, such as computer mice, to directly manipulate the visual objects displayed on the screen. Such interaction is challenging for users with low or no vision who do not receive sustained visual feedback. Although a few visualization tools (e.g., Willuhn et al. [2003]) allow keyboard-only navigation inside some visual graphs, most are not usable by users with visual impairment.

One example of this challenge is Web-based access to government statistical data. Such data is often geospatial, such as population distribution by geographical regions. This data is often presented as choropleth maps, which use colors to show the value for each map region. U.S. Federal law (www.section508.gov) requires that all government information on the Web be available to users with disabilities, and other governments around the world have similar laws¹. The paradox is that government agencies are presenting more data on the Web through choropleth maps, but at the same time, these government agencies are increasingly required to make that data accessible to people with disabilities. Therefore, this research not only has technical implications, it also has policy implications.

One of the most popular assistive technologies for users with visual impairment is the screen reader. A screen reader is a software program that provides synthesized speech, representing what appears on the screen and alternative text provided by the application [Slatin and Rush 2003]. JAWS and Window-Eyes are the two most popular screen readers among blind users [Lazar et al. 2007]. In addition, new approaches to presenting text using synthesized speech also appear in the research literature. For instance, the BrookesTalk screen reader provides additional navigational information to users [Zajicek et al. 1998] and math markup languages improve the comprehension of math formulas through screen readers [Karshmer and Gillian 2005]. However, the drawback of the new research approaches is that they all focus on other ways of presenting text, not methods for presenting graphical data.

Textual equivalents of graphical data could provide useful output for screen readers. For static graphs, it is standard practice to provide textual labels during the system development [W3C]. For dynamic graphs, tabular data presentations are used instead (e.g., Willuhn et al. [2003]), or textual summaries can be automatically generated from the dataset. Several problems exist in these current approaches. First, while a concise textual description is helpful, the interaction that is a critical part of the data exploration process in information

¹See www.w3.org/wai for government policies around the world.

visualization is lost. Automatic textual summarization methods require predefined summary templates and are not flexible enough to support the multiple goals of different users. Second, a tabular presentation may be good for basic data browsing but is hard for in-depth data comprehension and analysis. Third, synthesized speech can accurately describe data but the duration of listening to large datasets places a large cognitive load on the user.

Sonification (nonspeech auditory presentation of data) shows promise for building data exploration interfaces for users with visual impairment. The success of Braille-based interfaces have been limited in the past both due to the low levels of Braille literacy (some estimates are 20%) and the high cost of refreshable Braille displays [National Federation of the Blind]. Tactile maps would seem to be the most likely choice for understanding data on maps but the high costs of tactile displays mean that they are rarely used [Wall and Brewster 2006]. Even special printing devices, such as printers that can print raised paper maps, are prohibitively expensive and rarely available outside of schools [Wall and Brewster 2006]. The most commonly used assistive technology for users with any level of visual impairment is the screen reader which is inexpensive, requires only limited training, works with the standard computer hardware (the speakers), and only requires a software application.

There has been extensive research on data interaction in visualization systems. The research has produced design guidelines (e.g., the visual information-seeking mantra [Shneiderman 1996] and workspaces (e.g., Visage and Sage [Roth et al. 1997], Snap-together Visualization [North and Shneiderman 2000]) for users to perform data exploration through unified interaction methods. However, little has been investigated regarding whether techniques in visualizations can be translated for use in auditory data exploration without visual aids and what design implications are involved. It is important to investigate how the standard techniques used in visualizations could be utilized in the auditory mode and what adaptations are needed to make them applicable. Such investigation is not only valuable for designing effective auditory interfaces for blind users but may also benefit traditional visualizations. If visualization techniques and guidelines can be adapted for use in the audio mode, visualization systems could naturally incorporate audio counterparts. This would allow users with residual vision to consistently make use of both their visual and auditory abilities. The integration of audio could also be a promising solution to overcome some common visualization challenges such as visual occlusion.

This article presents research results that attempt to answer a set of questions such as: Can blind users handle and benefit from multiple highly coordinated data views using audio? What task-oriented user interaction actions will be supported, and can the actions from the visual information-seeking mantra be applied in the auditory mode? Can blind users handle and benefit from a rich set of such actions? How should the actions be designed? How would the choice of input device affect the interaction effectiveness? How should information be encoded in the auditory feedback during the interaction?

This article first describes a research framework for designing auditory interfaces for analytical data exploration. Our Action-by-Design-Component

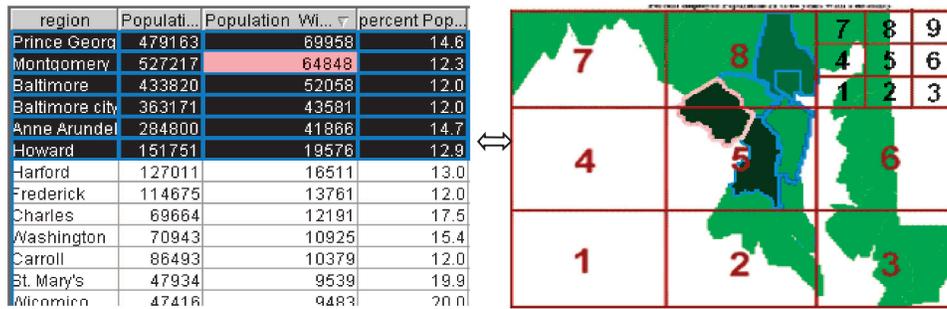


Fig. 1. Highly coordinated table and map views of the counties of the state of Maryland. Superimposed on the choropleth map is a representation of the recursive 3x3 keyboard exploration grid. The numbering is mapped to the layout of a numeric keypad.

(ADC) framework was inspired by information visualization guidelines but includes adaptations to accommodate the transient nature of auditory perception. The ADC framework uses a set of Auditory Information-Seeking Actions (AISA) to characterize task-oriented data interaction without visual aids. A partial preliminary version of the framework, and some initial sonification designs for georeferenced data, appear in earlier publications [Zhao et al. 2004; 2005]. In the earlier publications, the focus was limited to actions for conveying data distribution patterns on maps. Little evaluation has been done regarding the support for general data exploration tasks or the validity of the partial framework. In addition, the initial studies in previous publications were conducted with blindfolded sighted users not blind users.

Second, guided by the ADC framework, we developed a general exploratory data analysis tool for users with visual impairment, called iSonic. It presents three highly coordinated views and supports AISAs within and across the views. Figure 1 shows the table view and map view. We describe iSonic features and discuss the design rationale to illustrate the framework.

Third, we conducted an empirical evaluation of iSonic. A total of 7 blind users took part in the evaluation, with 42 hours of in-depth observation and interview data. This data collection enabled us to evaluate the effectiveness of iSonic design choices and draw general design implications.

Fourth, we applied the ADC framework to scatterplots which helps to demonstrate the generalizability of the framework.

Note that in this article, the term “user with visual impairment” refers to the larger group of users with some level of vision impairment, and the term “blind users” refers to users with no residual vision.

2. PREVIOUS WORK

Sonification, the use of nonspeech sound, has been used in various interface designs such as nonvisual GUI presentations [Brewster 1998; Mynatt and Weber 1994], application interaction control [Igarashi and Hughes 2001], and data presentations [Kramer et al. 1997]. Using the highly structured nature of

musical sounds (sometimes referred to as earcons) to convey information works even when no everyday auditory equivalent exists, is less tiring, and generally more appropriate than everyday sounds [Sikora et al. 1995]. Research has shown that musical sounds enhance numeric data comprehension (e.g., Ramloll et al. [2001]) and humans can interpret a quick sonified overview of simple data graphs (e.g., Flowers et al. [1997]; Brown et al. [2003]; Franklin and Roberts [2003]). From previous research, some guidelines were extracted (e.g., Brown et al. [2003]; Walker and Lane [2001]) and toolkits were developed to help researchers attempt different data-to-sound attribute mappings (e.g., Pauletto and Hunt [2004]; Walker and Cothran [2003]).

Recognizing the importance of interaction, some sonifications allow basic user movements in the auditory scene. Examples include up, down, left, and right movements among adjacent 2-D table cells using arrow keys [Brown et al. 2003] and continuous mouse movement to sonic probe a 2-D scatterplot in which the sound of each icon is activated when swept by the mouse movement [Smith et al. 1990]. To address the fact that blind users have difficulty using mice, Kamel and Landay [1999] developed a novel absolute pointing method with 3x3 grid recursions via the keypad in a drawing tool.

In exploratory data analysis, users often need to examine the data from various aspects. It is important that users are at the center of the interaction and initiate every interaction loop. Information visualization researchers have produced design guidelines (e.g., Shneiderman [1996]) and workspaces (e.g., Roth et al. [1997]; North and Shneiderman [2000]) for users to perform data exploration through unified interaction methods. For example, the visual information-seeking mantra “overview first, zoom, filter, then details-on-demand” Shneiderman [1996] characterizes the general visual information-seeking process and has been widely adopted in visualization designs. Visage and Sage [Roth et al. 1997] provide a workspace that supports multiple visualization views. Some basic operations are supported across all the visualization views such as filter, copy, roll-up/drill down, and scale. Users can also manipulate the data in one visualization to extract what is relevant to the task and reorganize the relevant data by dragging and dropping it to another view to automatically create a new visualization. North and Shneiderman [2000] recognized the value of multiple coordinated visualizations in the Snap-Together visualization interface that enables users to mix and match visualizations and coordinations without programming.

However, a guiding framework is absent for data exploration in the auditory mode without visual aids. Perez-Quinones et al. [2003] described a task by data structure taxonomy for targeted information searching in voice access to Web pages but not for exploratory data analysis. Some recent frameworks or models tried to provide guidance for designing interactions in sonifications, but they did not address task-oriented information-seeking needs in the auditory mode without visual aids. Since they assumed the dataset could naturally be modeled as a physical environment, their framework may not be suitable for abstract datasets. Their sonification and interaction methods emphasized the use of special devices and spatial sound not analytical task needs or the suitability in the absence of visual feedback.

The Model-Based Sonification (MBS) framework [Hermann and Ritter 1999] has virtual sounding objects with specific surface materials and allows users to define the layout of the scenery and its acoustic properties. Model-specific user interactions are defined as a part of the sonification model. Observing how humans interact with real-world physical acoustic objects (musical instrument, tuning in radio channel, etc.), Hunt et al. [2004] emphasized that interaction with sonification should promote user engagement with the system and achieve an interaction style that flows. To achieve these goals, they proposed using high sound complexity, low-latency sound generation, and advanced interaction modes by using various special devices such as an articulated hand, an audio-haptic ball [Hermann and Ritter 1999], or other tangible interfaces, rather than just using standard computer input devices such as mice or keyboards.

SoundWalker [Saue 2000] is a model for sonification of large spatial datasets based on ideas from ecological acoustics. Datasets are modeled as spatially distributed objects in a scene that mimics a real-world environment. Objects are sonified as parameterized auditory icons carrying information about the object such as the size and attribute values. Users walk through the scene by moving an active listener (controlled by mice or similar devices) and listen to the sound rhythmically produced by the nearby data items. SoundWalker is designed to help users to make associations between the dataset and its physical origin, for example, to use a subterranean cave model to present a seismic dataset or a human body model to present a medical dataset. The authors mentioned analysis tasks but did not show how tasks can be performed by walking in the model.

3. AN ACTION-BY-DESIGN-COMPONENT FRAMEWORK

To guide designers in producing effective auditory interactions, the authors developed an Action-by-Design-Component framework. The Action refers to a set of Auditory Information-Seeking Actions (AISA) that designers should consider supporting in an auditory interface for exploratory data analysis. AISA actions include *gist*, *navigation*, *situating*, *searching*, *filtering*, *details-on-demand*, *selecting*, and *brushing*. The Design Component refers to four design aspects to implement supports for the actions. Design Components include *data view*, *navigation structure*, *input device*, and *auditory feedback*.

Many of the AISA actions resemble those commonly used in visualization but involve different cognitive processes and present special design challenges due to the highly transient nature of sounds. Other researchers can further extend the list with more actions. The AISA actions are described in the following.

A *gist* is the overall data trend presented via a short auditory message. It guides further exploration and may allow the detection of anomalies and outliers. A *gist* is an auditory overview but has special design and cognition challenges because human auditory perception is less synoptic than visual perception. Visual overview has tremendous value in visual data exploration. Previous work (e.g., Brown et al. [2003]; Flowers et al. [1997]) showed that

sighted users can interpret data trends from a gist of simple graphs such as scatterplots and line graphs, yet there has been little study of auditory gists and how blind users would use gists in exploratory data analysis.

Navigation is moving around to examine portions of the dataset by listening to a subgist of that portion. In addition to playing, stopping, resuming or rewinding along the time dimension of the dataset's gist, navigation also needs to follow other paths that are natural to the data relations. When navigation is done in a fast tempo along the sequencing path of the gist, the effect of the soundscape produced by navigation is similar to the gist.

A visual interface provides a sustained display that users can directly manipulate. In auditory interfaces, users need to construct a mental representation of the display space and virtual navigation structures so that they can efficiently move in the dataset. Without a persistent display, users can easily get lost. To regain orientation, they need to situate themselves by requesting the status of where they are.

While navigation is an exploratory action, searching is a fixed-goal action that lands on the data items. Searching breaks the process of mental representation construction so situating may be needed to regain orientation after the search is completed.

Filtering out unwanted data items according to some query criteria trims a large dataset to make it easier to manipulate and allows users to quickly focus on interesting items. In visualization, dynamic query coupled with rapid (less than 100 milliseconds) display update is the goal [Shneiderman and Plaisant 2005]. In the auditory mode, different goals are needed because longer time periods are needed to present a gist of changes. Auditory feedback about the filtering results needs to be given after filtering is done instead of continuous display updates during the filtering process.

When the number of items is small, users can listen to the details. As the number of items grows, speech is often too lengthy for presenting a gist, but it can be effective at the details-on-demand level. Since too much detail slows down the sequential presentation and can be overwhelming, multiple information detail levels are needed.

By selecting, users specify interest in particular data items. Those data items are marked and can be revisited later or examined in other contexts.

In visualization, linked brushing allows users to manipulate the data in one view while seeing the results in other views. It requires users to construct and maintain multiple mental representations of the data views simultaneously, which can be mentally intensive in the auditory mode. Additionally, auditory feedback from multiple views needs to be clearly distinguished to avoid confusion and overloading. In the auditory mode, brushing can be done in a sequential style by selecting data items in one view, then explicitly switching to another view to examine them in a different data relation. For example, users may want to select a few geographical regions from a list alphabetically ordered by region names and have them plotted on a map to view their geographical locations.

As illustrated in Figure 2, an AISA is an interaction loop involving four Design Components. In each interaction loop, the user uses an input device to

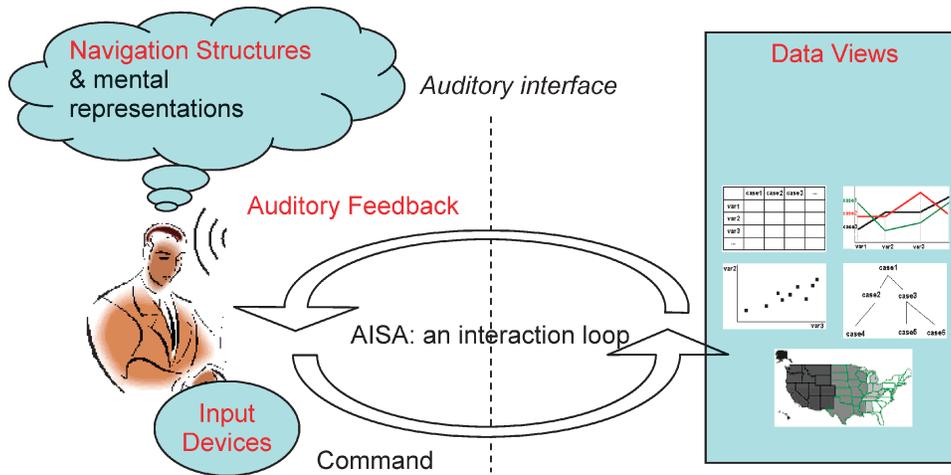


Fig. 2. An Action-by-Design-Component framework. An Auditory Information Seeking Action (AISA) is an interaction loop involving four Design Components (Data Views, Input Devices, Navigation Structures, and Auditory Feedback).

issue a command and listen to the auditory feedback. The center of the loop is the data view that governs the navigation structure, allowing the user to build a mental representation of the data space so that he/she can interpret the auditory feedback.

A data view is a form of presenting the data items and their relations. Examples include a table, map, scatterplot, or line graph. Previous research has shown that users with visual impairment were able to learn, interpret, and benefit from nontabular data presentations. There is also evidence [Bennett and Edwards 1998] that choosing the right data view for a given task dramatically influences performance in auditory interfaces.

Navigation structures are the paths that users follow to move around in the dataset. The paths should reflect the relations of the data items in the data view. In previous research, blind users utilized a mouse or other input device to move a cursor in a 2-D or 3-D view of the dataset to activate sounds of the data items near the cursor position. Donker et al. [2002] named this style of navigation the *torch metaphor*. A torch in a nighttime forest only illustrates the nearest trees. The cursor implements an auditory torch that activates only the nearby data items. The torch metaphor could be useful for some data views, for example, a scatterplot, but may be inefficient for others, for example, a node-link diagram.

The choice of input device depends on both effectiveness and universal availability. Speech as input can be tempting but lacks the kinesthetic feedback users can get from operating physical input devices. Sensory feedback can help with users' orientation and mental representation in the interaction. Card et al. [1990] categorized physical input devices by their physical manipulation properties and defined several important factors such as the cost. For users with visual impairment or full blindness, a system should provide interactions

optimized for keyboard-only operations since a keyboard is standard equipment for nearly all computers and blind users have typically memorized the keyboard layout.

Auditory feedback should have a low latency. Short and responsive feedback increases user engagement so that users can quickly refine their control activities in the exploration process. It should synchronize with other display modalities to allow perceptual combinations. While humans are good at selective listening, attending to multiple simultaneous sounds is difficult and the amount of accurate information that can be extracted from simultaneous sound streams is limited [Handel 1989]. The sounds of multiple items often need to be sequenced instead of being played all at once. When the number of data items is large, data aggregation may be necessary to design short feedback.

4. ISONIC AND THE DESIGN SPACE FOR GEOREFERENCED DATA

Guided by the ADC framework, we developed iSonic, a tool for georeferenced statistical data (Figure 1). Two users without residual vision were involved in the iterative design process. Many iSonic design decisions were based on their suggestions as well as results from earlier evaluations of some potential design choices.

4.1 iSonic Design Components

4.1.1 Data Views. iSonic provides three highly coordinated data views – a region-by-variable table, a choropleth map, and a bivariate scatterplot. The scatterplot view was integrated into iSonic as an extension after user evaluation and more detail is provided in Section 7 of this article. The next paragraphs provide information on the region-by-variable table and the choropleth map.

The table shows multiple statistical variables simultaneously. Each row corresponds to a geographical region and each column to a variable. Table rows can be sorted, allowing quick location of low or high values. An example of the data table in iSonic appears in Figure 1.

While geographical coordinates and adjacencies could be added as table columns, such information is better displayed on a map. Participants in our previous study [Zhao et al. 2004] strongly preferred a map to a table for discovering geographical value trends and performed better on pattern-recognition tasks using a map as opposed to a geographical knowledge-enhanced table. Other views, such as line graphs or scatterplots, can be helpful for some analytical tasks; however, they were not included in the evaluation. Auditory and visual displays are synchronized to allow communication between sighted users and users with any level of visual impairment. An example of the map view in iSonic appears in Figure 1.

4.1.2 Input Devices and Navigation Structures. When choosing input devices, both device availability and how effectively their physical properties match the navigational properties of the two data views should be considered.

In iSonic, the table navigation follows the row and column table structure. It is discrete and relative because what matters is the relative row/column order, not the exact spatial location or size of each table cell. On the other hand, the map navigation follows the regions' positions and adjacencies. Both the relative region layout and the absolute region locations and sizes are useful.

iSonic works with a keyboard alone. A keyboard is available on most computers and users with any level of visual impairment are very comfortable using the keyboard. The arrow keys are used as a natural means for relative movements in the left, right, up, and down directions. The numerical keypad potentially allows relative movements in 8 directions. Furthermore, blind users tend to have limited or no experience with using mice [Kamel and Landay 1999]. The keyboard can also be transformed into a low-resolution 2-D absolute pointing device, for example, by mapping the whole keyboard layout to 2-D screen positions. In iSonic, the 3x3 layout of the numeric keypad is utilized for positioning.

iSonic also works with a touchpad. Touchpads are relatively common. A 14" touchpad costs less than \$150. A touchpad provides high resolution 2-D absolute pointing and allows continuous movements by the fingers. The kinesthetic feedback associated with arm and finger movements, combined with the touchpad frame as the position reference, may help with users' position awareness on maps. Tactile maps placed on the touchpad can be helpful [Landau and Gourgey 2001] but have limited usefulness since they need to be changed when the map changes, and tactile printers are expensive and rarely available. When resources are available, a generic grid with subtle tactile dots may be used instead as a position and direction aid.

4.1.3 Auditory Feedback. iSonic integrates the use of speech and musical sounds. Values are categorized into 5 ranges, as in many choropleth maps, and mapped to five violin pitches. An auditory legend is available by pressing the "L" key. The same mapping is used in the table view. Violin pitches present the region values, while other musical instruments indicate when users are outside the map or crossing a region border in the touchpad interface or crossing a body of water to reach a neighboring region in the keyboard interface. Region border crossing is not sonically indicated in the keyboard interface since each key press implies navigating to a neighboring region. Stereo panning effects (range 0-127) are used to indicate a region's azimuth position on the virtual auditory map, with 0 for the leftmost region(s) and 127 for the rightmost region(s). For any region in between, the stereo panning position is an interpolated number between 0 and 127, according to the region's position on the visual map. Stereo panning is also used in the table to indicate the column order, with 0 indicating the first column and 127 the last column. Using the plus and minus keys, users can switch among four information levels for each region: region name only, musical sound only, name and sound, name and sound plus reading of the numerical value.

Sound duration can present values but would significantly prolong the feedback and is not appropriate when values of many regions need to be presented in a short time. Region locations could be mapped to sound locations using

virtual spatial sound synthesized with Head Related Transfer Functions (HRTF) [Wenzel et al. 1993]. HRTF describes how a given sound wave input is filtered by the diffraction and reflection properties of the head, pinna, and torso before the sound reaches the eardrum and inner ear. Utilizing HRTF, computers can synthesize sound that the listener will perceive as if the sound comes from the 3-D real-world space, although the sound is actually heard through a stereo headphone. Spatial sound from the HRTF method provides high perceptual resolution in the azimuth plane but is not satisfactory in the elevation plane, especially when a generic HRTF is used. Using individualized HRTF could improve the elevation perception but its measurement is a long process requiring special equipment and careful calibration. Additionally, HRTF spatial sound is computing-intensive. While iSonic was connected to a virtual spatial sound server for potential evaluation, the final iSonic tool focused on MIDI stereo sound. Another enhancement that was evaluated, a piano pitch after each violin value pitch to indicate the region's elevation position, did not improve performance [Zhao et al. 2005].

4.2 iSonic Auditory Information-Seeking Actions

iSonic supports AISAs in both the table and the map views, including sequential brushing between the two views. Each interface function can be activated from a menu system that also gives the hotkey and a brief explanatory message for the function. The menu system is presented auditorily and visually at the same time. It can be brought up anytime by pressing F10, which is also the default key to activate menus in GUI.

4.2.1 Gist. In the table view, a gist is produced by playing all values in a column or a row. The sequencing follows the values' order in the table, from top-to-bottom or left-to-right. In the map view, there is no natural mapping from the geographical relation to the time relation. Research has shown that sequencing that preserves spatial relations helps users to construct a mental image of the 2-D representation. Sequencing is done by spatially sweeping the map horizontally from left-to-right, then vertically. All regions are sorted into a multirow linear order, according to the regions' spatial locations, in relation to each other. Each left-to-right horizontal sweep involves a subset of regions in one sweep row. The values of regions in one sweep row are played sequentially. When one sweep row finishes, the sweep row below it follows. During a gist, only value sounds are played. The value sounds' stereo panning positions value provide the regions' geographic location cues. Like in a typewriter, when the end of sweep row is reached, a tick mark sound is played and the stereo effect reinforces the change. A bell indicates the end of the sweep of the whole map. The same sweep order holds for subgists of parts of the map. A subgist is a gist presenting only a portion of the map, specified during the navigation. For both views, the current information level controls the amount of detail in the gist, thus controlling its duration. For example, when the information level is set to "musical sound only", a sweep of the entire US state map containing 51 regions includes 51 violin pitches and lasts for 9 seconds.

4.2.2 Navigation. Table navigation is done by using arrow keys to move up, down, left, right, and to top, bottom, left and right edges. Users can press ‘u’ to switch between two modes. In the cell mode, the current cell is played. In the row/column mode, a subgrid of the whole row or column is played. While it is easy to navigate the table, using a keyboard to navigate maps with irregularly shaped and sized regions presents special design challenges. Relative movements between neighboring regions reveal region adjacency but do not convey region shapes, sizes, or absolute locations. Participants in the previous studies reported that they only had weak location awareness by using this navigation method even with the sound stereo panning effect providing some position cues. Furthermore, it is a challenge to define a good adjacency navigation path for a map that is not a perfect grid. A movement may deviate from the direction users expect. Reversibility of movements can also be a problem in which a reversed keystroke may fail to take the user back to the original region. To tackle some of these problems, cell-by-cell movements were tested on a mosaic version of the map [Zhao et al. 2005]. However, this did not improve users’ location awareness, and was preferred less because it required more keystrokes to move around.

It is expected that navigations, based on absolute pointing, may help. Kamel and Landay [1999] first used a 3x3 grid recursion method via the keypad in a drawing tool. In iSonic, the map is divided into 3x3 ranges (Figure 1), and users utilize a 3x3 numerical keypad to activate a spatial sweep of the regions in each of the nine map ranges. For example, hitting ‘1’ plays all regions in the lower left of the map, using the same sweep scheme as the overall gist. Users can use Ctrl+[number] to zoom into any of the ranges, within which they can recursively explore using the 3x3 pattern or arrow keys to move around. Pressing ‘0’ sweeps the current zoomed map range or the whole map.

With the touchpad approach, users drag their fingers or press spots on the smooth surface touchpad to activate the sound of the region at the finger position. Stereo sounds provide some complementary direction cues. The sound feedback stops when the finger is lifted off. The touchpad is calibrated so that the current map range is mapped to its entire surface. Preliminary observations suggest that both the keyboard and touchpad navigations allow users to gain geographical knowledge. A controlled experiment is planned in the future, to compare them in detail.

4.2.3 Details-On-Demand and Adjust Information Level. Pressing the spacebar key plays the details of the current region. Another way to get the region details is to increase the information level to the maximum level in which all details of a region are given by default when users navigate to that region.

4.2.4 Situate. When users press ‘I’ (as for Information), iSonic speaks the current interface operational status. In the table, it includes the row/column counts, headings of the current table position, navigation mode, sorting status, regions selected, and so on. In the map, it includes the name of the variable displayed, navigation position, regions selected, and so on.

4.2.5 *Select.* In both views, users can press ‘L’ (as for ‘Lock’) to select/unselect the current region. Each region is marked either as selected or not selected. Pressing ‘L’ on the current region toggles the region’s selection status. Users can also press ‘A’ to switch between “all regions” and “selected regions only” operation modes. In the “selected regions only” mode, AISAs only activate sounds of the regions that are currently marked as selected. For example, a gist in the “selected regions only” mode plays sounds only for selected regions. Navigating in the “selected regions only” mode will only move among selected regions.

4.2.6 *Brush.* Brushing is done by users switching back and forth between the two views. The views are tightly coupled so that action results in one view are always reflected in the other. For example, users can select a region in the table view and show selected regions only. When users switch to explore the map view, only the selected region is played. By sweeping each of the 9 map ranges, users can rapidly (but roughly) locate the region on the map.

4.2.7 *Filter and Search.* Filtering in iSonic is done by slider-based queries. It is complex even for sighted novice users and was not examined in the current evaluation. Searching is obviously helpful but was not implemented at the time of the evaluation.

5. ISONIC IMPLEMENTATION

iSonic is implemented as a Java application that also runs from the Web through Java Web Start Technology [Sun Microsystems]. The GUI part is written in Java JFC/Swing and the musical sounds are produced through the Java MIDI Sound technique. Speech is produced by sending commands to an accompanying networked speech server built on Microsoft Speech SDK 5.1. Although the speech operates in a manner similar to a screen reader, including the speech directly in the application instead of utilizing a screen reader application provides more flexibility and also more stability of the speech output. Screen readers frequently conflict with applications and cause crashes which are one of the most frequent causes of frustration for blind users [Lazar et al. 2007]. Incorporating the speech output directly into the iSonic application greatly decreases the likelihood of a screen reader-caused crash and is more likely to lead to accurate results in evaluating the iSonic tool. iSonic can also produce virtual spatial sounds by sending a command network datagram to a spatial sound server that simulates real-world sounds using Head-Related Transfer Functions (HRTF) [Wenzel et al. 1993]. The spatial sound server was developed at the Perceptive Interfaces and Reality Laboratory [Zotkin et al. 2002]. Additionally, iSonic can be configured to load and play pre-recorded speech or musical sound files. iSonic is a real-time application, giving real-time highly synchronized visual and auditory responses to every user action with delays of usually less than 100 milliseconds.

The iSonic implementation is based on the Model-View-Controller (MVC) paradigm [Reenskaug] commonly used as a design framework for modern GUI interfaces. It extends the MVC paradigm by treating the auditory display and

GUI display as two parallel subviews (a visual view and an auditory view) residing inside each data relation view (e.g., map, table, and scatterplot). For each data relation view, the auditory display is produced by the creation and execution of various sound objects (e.g., MIDI events, speech objects). The two displays are synchronized by monitoring the execution of the sound objects. Each data relation view has related controller classes that define the mapping from input devices to interaction commands. Model classes that contain the input dataset and each data item's interaction status (e.g., whether a data item is selected) are shared among all data relation views. Each data relation view may also have its own model classes (e.g., map shape data for the map view).

iSonic has a rich set of configuration parameters that can be used to customize its visual, auditory, and interaction behaviors. This allows iSonic to be adapted for different users and to be used as a research tool to compare some design options.

To generate map gists, iSonic accepts user-defined map spatial sweep orders as part of the input data. It can also automatically produce sweep orders in the absence of predefined orders. Automatically producing map spatial sweep orders, which are similar to what people may define manually based on their visual impressions, can be very challenging. iSonic uses a preliminary algorithm that first recursively splits the regions on the map into multiple subsets with as little inter-set intrusion as possible, then connects the regions inside each subset to produce the sweep path. The algorithm uses a greedy approach and does not guarantee globally optimized solutions but produces good results for several U.S. maps tested. More details on the problem of map spatial sweep order generation are available in Zhao [2006].

6. EVALUATION

During early design iterations for iSonic, controlled experiments were conducted to compare the effectiveness of design alternatives including the choice of data views, map navigation methods, and sound encoding schemes [Zhao et al. 2004, 2005]. However, an exploratory data analysis task is a complex process that involves many interface components. During the process, many inherent human variations can come into play such as experience and cognitive styles. In order to obtain insights into users' auditory information-seeking behaviors, case studies were conducted. Through a combination of direct observation, thinking aloud protocol, and in-depth interview, case studies can reveal the underlying design strengths and weaknesses and capture common user behaviors as well as individual differences.

During the summer of 2005, intensive case studies were conducted with 7 local users who are blind, producing 42 hours of observation and interview data with an average of 6 hours per user. Using cross-case analysis, the common user behaviors and feedback were extracted. This feedback allowed for (1) evaluating the effectiveness of iSonic design choices, (2) identifying features helpful to each data exploration task category, and (3) identifying task road blocks in order to target training and modifications to the interface and the framework.

6.1 Participant Profiles

All 7 participants possessed basic computer skills and relied on screen readers to access computer information. They all reported to be comfortable with maps and data tables, had experience with numerical datasets, and used government statistical data at work. All participants had some previous experience with Excel, while some were expert users. They also had various previous experience with maps through tactile maps or map reading before the vision loss. All participants were in the age range of 23 to 55. Three of them were born blind (P2, P3, P4), and the others became legally blind after age 15 (P1, P5, P6, P7). None of them had residual vision, and none of them were newly-blind. Among the born blind, 2 were males, 1 with a college degree (P2) and the other with a doctorate degree in law (P3). The remaining female (P4) had a masters degree in English. Among the participants who became blind after 15, 1 was a male (P7) with a college degree in business and commerce. The other male (P1) was about to finish college in science and technology. For the 2 females, 1 had a college degree (P5) and the other had a masters degree (P6), both in social science. All participants volunteered to take part and were compensated for their time.

6.2 Material and Tasks

The case studies used the basic iSonic configuration with equipment that is available to most computer users: stereo auditory feedback through a pair of speakers and a standard computer keyboard as the input device.

Three datasets were used, one for training, one for testing, and one for posttest free exploration. The data used was 2003 census data on general population information, employment of population with a disability, housing value and vacancy, education levels, and household income. The training dataset contained 8 variables and was focused on the 50 US states plus the District of Columbia. The test dataset contained 12 variables and was focused on the 24 counties in the state of Maryland. The posttest dataset was about the 44 counties in the state of Idaho (but participants were not told what it was). Participants' geographical knowledge of US states and Maryland counties ranged from excellent to very poor. This allowed us to observe the influence of geographical knowledge on task behaviors and interface usability.

Seven tasks were designed for each dataset. The tasks were based on those used in previous research on finding statistical data on government Web sites [Ceaparu and Shneiderman 2004]. Three tasks required value comparison in the geographical context (T5, T6, T7) and four did not need any geographical knowledge (T1, T2, T3, T4). Task orders were different between the training and testing sessions but were consistent for all participants. The testing tasks are summarized here.

T1: (Find min/max) Name the bottom 5 counties with the lowest housing unit value.

T2: (Find the value for a specific item given the name) What is the population of Dorchester county?

- T3: (Correlation) Which of the two factors is more correlated to “Median household income”: “percent population with bachelor’s degree and above”, or “Percent employed population”?
- T4: (Close-up item comparison) For what factor(s) does Montgomery county do better than Frederick county: (1) employment rate for population with a disability, (2) percent population with at least college education, (3) household income, and (4) average housing unit value.
- T5: (Find items restricted first by value relations then by geographical locations) How many of the bottom 5 counties with the lowest housing unit value are in the western part of the state? Name them?
- T6: (Find items restricted first by geographical locations then by value relations) For all three counties that border Frederick, plus Frederick, which one has the highest percent housing unit vacancy?
- T7: (Value pattern in geographical context) Comparing “population with a disability” and “percent population with a disability”, which variable generally increases when you go from east to the west and from the north to the south.

Participants also performed a similar set of testing tasks in Microsoft Excel 2002 with their usual screen readers (all happened to have experience with the JAWS screen reader) and compared the task experience. It was not our intention to compete with Excel. Rather, Excel is a standard tabular data viewer and the comparison is useful as a method to solicit user comments on which interface features were helpful for each task. All participants had some previous experience with Excel, while some were expert users. Tactile maps were not provided when participants used Excel because many blind users do not have access to tactile maps (only P5 uses one, and it is owned by the state government of Maryland).

6.3 Procedure

Each case study was carried out in two sessions on consecutive days at the participant’s home or office. In the first session, the participant listened to a self-paced auditory step-by-step tutorial, tried out all iSonic features, and practiced seven sample tasks with the training dataset. For each training task, a sample solution and the correct answer were given. Participants could either first try to solve the task on their own or directly follow the sample solution.

In the beginning of the second session, those participants with limited Excel experience were given time to practice. After adjusting the speech rates to the participants’ satisfaction, they performed seven tasks similar to the training tasks in both Excel and iSonic. For each pair of tasks, the participant first did the Excel task, then the iSonic task, and finally compared the interface experience for that task. The iSonic task was similar to the Excel task but modified. They used the same testing dataset but different variables so data learning between tasks should be minimal. Participants were asked to perform the Excel task first to minimize the effect on the Excel task resulting from the geography learning in the corresponding iSonic task. While there was a chance of strategy transfer from the Excel task to the iSonic task, the Excel

Table I. Per-Participant Task Success Rates in the Case Studies

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | Avg |
|--|-----|-----|-----|-----|-----|-----|-----|-----|
| iSonic training task success rate w/o help | 4/7 | 6/7 | 6/7 | 5/7 | 5/7 | 3/7 | 4/7 | 67% |
| iSonic testing task success rate w/o help | 7/7 | 7/7 | 7/7 | 4/7 | 6/7 | 7/7 | 6/7 | 90% |
| New strategy carry-over rate | 4/4 | 2/2 | 2/2 | 1/2 | 2/3 | 4/4 | 0/1 | 74% |
| iSonic testing task success rate (non-geo) | 4/4 | 4/4 | 4/4 | 2/4 | 3/4 | 4/4 | 3/4 | 86% |
| iSonic testing task success rate (geo) | 3/3 | 3/3 | 3/3 | 2/3 | 3/3 | 3/3 | 3/3 | 95% |
| Excel testing task success rate (non-geo) | 3/4 | 4/4 | 4/4 | 3/4 | 3/4 | 3/4 | 4/4 | 86% |
| Excel testing task success rate (geo) | 1/3 | 1/3 | 2/3 | 1/3 | 2/3 | 0/3 | 0/3 | 33% |

task might also have benefited strategically from the iSonic training task. An interview was conducted after participants performed all the testing tasks in both interfaces. Finally, participants explored an unknown map and dataset (the posttest dataset) for 5 minutes and reported things that they found interesting. This was to observe what users would do when they encountered a new map and data.

6.4 Cross-Case Task Completion and Error Analysis

After spending an average of 1 hour 49 minutes going through all the interface features by following the tutorial, participants successfully completed 67% of the training tasks without referring to the sample solution or soliciting any other help. After the training, participants were able to retain their newly acquired knowledge and successfully completed 90% of the tasks on the next day in a different context without any help. For 74% of the tasks for which participants used different strategies than the given solution in the training session, they adopted the sample strategies from the training session to use in the test session. Table I lists the per participant task success rates in the case studies. More details about each participant are available in Zhao [2006].

For tasks that did not require geographical knowledge, the average testing success rates were similar for iSonic and Excel, both at 86%, although participants ranked iSonic easier than Excel (7.9 vs. 7.0 based on a 10-point scale with a higher number being easier). The explicitly reported reasons, in decreasing order of frequency, included: (1) the pitch was helpful in getting the value pattern and comparing values; (2) it was easier to sort in iSonic because sorting was done by pressing one key in the desired column to toggle the sorting status instead of handling multiple widgets in the dialog window as in Excel; (3) it was helpful to isolate a few regions from other interfering information by selecting; (4) it was flexible to adjust the information level during the task; (5) there was more than one way to get the same information.

For geography-related tasks, the average testing success rate was 95% in iSonic. In Excel, the two participants with excellent knowledge of Maryland geography (P3 and P5) achieved a success rate of 67%. Other participants either skipped some tasks due to the lack of geographical knowledge or tried to make an educated guess but gave incorrect answers, resulting in an average success rate of 20%. On a 10-point scale, participants gave iSonic an average of 8 on easiness for all 3 geography-related tasks and Excel an average of 5.8 for the tasks they performed. The explicitly reported reasons included: (1) the map was easy to use and very helpful (mentioned by all participants in all 3

tasks); (2) it was great to be able to switch between the map and table, select things in one view then look at them in another; (3) the pitch was helpful in getting the value pattern and comparing values; (4) there was more than one way to get the same information.

In general, it was easy for the participants to choose an efficient combination of interface features to do the tasks (average 7.4 on a 10-point scale with 10 being easy). Correlation tasks, however, turned out to be challenging. Most participants understood the concept but did not know how to do it efficiently in iSonic until they viewed the sample solution. Only P7 easily came up with the sample solution. He sorted the main variable ascending in the table view, then in the row/column navigation mode, and swept other columns with “pitch only” to check which one had more consistently increasing pitch pattern. Other participants mostly went across all requested columns to check if the pitches or numbers were consistently small or large for each region. Some also sorted one or all columns. One participant (P6) said she would have the data plotted in a scatterplot or multiline graph and then would have her human reader look for the highest correlation. All participants, except P4, were able to learn from the sample solution and successfully applied it in the test session. The geographical value pattern tasks were easy for most participants except for P7 who guessed the answer correctly but was very uncomfortable with the task. Instead of visualizing the map, he emphasized accuracy by trying to calculate and compare the average value for each of the 9 map ranges. This was consistent with our earlier finding that task strategies affect geographical value pattern recognition [Zhao et al. 2005].

Setting aside the strategic difficulties, incorrect answers in iSonic were caused by two common errors. First, participants sorted the wrong variable (a third of all errors). This might be due to the high similarity of variable names and to the fact that the interface did not confirm the variable sorted when the sorting key was pressed. Second, some participants skipped the 1st region in the table (a third of all errors) because pressing the down arrow key after hearing “already top edge” took the participants to the 2nd row instead of the 1st.

6.5 Task Strategies and iSonic Usage Patterns

6.5.1 Map vs. Table. All participants used the table for most value comparisons and used the map when they needed to compare items in the geographical context (e.g., T7) or to find out/confirm the location of a region. The table was often used to change the variable to display on the map, but, more importantly, the sorting feature was used to find minimum or maximum values, named regions, and values of specific regions. The table was also used to compare the values of multiple regions and to check correlations. The map was used sometimes by a few participants to find regions.

6.5.2 Brush. All participants became proficient in switching between the table and the map views according to the changing needs for data relations during the task. The tight coordination between the map and the table views

was considered the most significant strength of iSonic by all participants. One participant noted that “It is cool to select things in one view and look at them in the other. The biggest advantage of this tool is the ability to quickly change between the table view and the map view”. To find items restricted first by value relations, then by geographical locations (e.g., T5), most participants first used the table to find items meeting the value restriction, selected to isolate them, then switched to the map to check their geographical locations. Some participants skipped the use of the map and used their pretest geographical knowledge to judge if the selected items satisfied the geographical restriction. A few participants first used the map to find all items that met the geographical restriction, remembered them, then sorted the table to find items satisfying the value restriction and reported the intersection of the two sets. The latter two strategies relied on participants’ memory of the intermediate results and caused some errors. Participants said they would have used selecting to mark items during view switching if the number of items were larger. To find items restricted first by geographical locations, then by value relations (e.g., T6), most participants first found and selected items meeting the geographical restriction on the map, then either used the pitch and value in speech to check if they met the value restriction or switched to table view and used sorting to compare their values.

6.5.3 Use of Pitch. Using pitches to present numeric values was considered intuitive, entertaining, and very helpful for data comprehension. It took some participants a few tasks to get used to this idea but they became increasingly inclined to using pitches for both trend analysis and close-up value comparison. Comments from participants included: “Pitch makes it a lot easier and quicker to compare values.” “Tones are very helpful to find patterns in a series of values. In some extent it helps me to do things I used to do with (visual) graphs.” “All the other applications are boring. iSonic has its personality. It has the map that I really enjoyed. The tones are entertaining and fun.” To use pitches, most participants either changed to the pitch-only information level (especially for trend analysis), or used the level with both pitches and numbers in speech but quickly navigated through items, only waiting for a number to be spoken for confirmation purpose (in value comparison). Some participants were able to tell the absolute value category using only one pitch, while some needed to use other pitches as references. All participants, except P4, were comfortable with the simultaneous pitch and speech presentations. P4 reported that pitches and speech interfered with each other and requested that the pitch volume be toned down. However, the participant declined the suggestion to completely remove pitches because she used pitches exclusively in trend discovery.

6.5.4 Adjust Information Level and Details-On-Demand. All participants frequently adjusted the information level during a task. Participants mostly used region name plus value pitch, or region name plus value pitch, together with the value in speech. When the information level with value in speech was used, many participants cut it short by navigating to another item before the value speech finished and only waited for it to finish when they wanted to confirm the value. In automatic map-sweep searching for a region, spoken values

were typically removed. To sweep the map or a table column for value patterns, for instance, for geographical patterns or correlations, most participants used the pitch-only level because it let them skim through the data the fastest. A few participants chose to keep the names on to keep track of the meaning of each sound while still being able to go through the data at a decent pace. To find a named region on the map, P7 often used the “region name only” level. Details-on-demand was mostly done by increasing the information level to the maximum level instead of pressing the space key.

6.5.5 Gist. Table sweep was very intuitive. To check value patterns, for example, for the correlation tasks, some participants used an automatic pitch-only sweep of each column by navigating the table in the row/column mode.

Automatic sweep of the whole map was typically done with pitch only or with the region name spoken along with the pitch. P3 said “automatic sweep will be my first step to get acquainted with a new map to get the big picture” During the posttest free exploration of an unknown map, P3 swept the map several times in pitch-only to obtain a rough idea of where the highly populated regions were before starting to explore. P2 swept the unknown map once and accurately reported that most highly populated regions were in the west by judging from the pitches and the sound panning positions. P2 was the only one that consciously used stereo panning cues in tasks. Most participants said it was not difficult to understand how the sweep was done but they needed to know what the map looked like to make sense of it. Once they broke the whole map into nine smaller ranges and swept each range using the keypad, it made more sense. All participants, except P7, were able to easily tell if a variable had a given geographical distribution pattern by sweeping the nine ranges in pitch only. Unexpectedly, map sweep was also frequently used by all participants to locate a region on the map. This was typically done with the region names spoken and often combined with the arrow key navigation and 9-range sweep. It was also used to check what regions had been selected.

6.5.6 Navigate. Navigating the table was considered easy by users. Participants mostly used cell mode because “it allows finer control of what to play”. The row and column mode was used by some participants to sweep a column for the correlation and close-up comparison tasks.

All participants reported that overall, it was very easy to navigate the map. The 3x3 exploration was frequently used by all participants except P2 who mainly used arrow keys to navigate and used sound panning to judge region locations. All participants understood the mapping between map locations and the 3x3 layout of the keypad. They were able to use the 3x3 exploration to find the map location of a specific region and to find what regions were in each map part. While participants mostly looked for a region by navigating the table (typically by first ordering it alphabetically), sometimes they used the map. They often first used the 9 numeric keys to find out which range contained that region, and then used arrow keys to move to that region. The 3x3 exploration also allowed some participants to acquire knowledge about the overall map shape and the region layouts. During the study, P3, P5, P6, and P7 reported the overall map shape and region density distribution. P7 also used two-level

recursive 3x3 exploration to find the county layout in the central and eastern parts of northern Maryland.

Participants seemed to be able to zoom into/out of the 9 map ranges and stay aware of their zooming positions. Many participants played with zooming extensively in training but did not use it in the evaluation. Their explanation was there was no need to use zooming in the tasks and the Maryland map only has 24 counties. If there was a need to focus on one area on a much bigger map, zooming could be helpful. Many participants expressed the concern that zooming might become mentally intensive as the map scale grows. One observed problem with 3x3 zooming was that some adjacent regions are assigned to different ranges and thus not reachable from each other after zooming. The problem can be remedied by allowing zooming centered on a region of interest.

Arrow key navigation was essential to find a region's geographical neighbors and was used by all participants in adjacency tasks. It was also often used to explore regions in a small map range, typically identified earlier with the 3x3 exploration. While P2 mostly used arrow keys to navigate the map, most participants were inclined to use the 3x3 exploration because it gave the absolute region locations.

Participants noted: "Arrow key navigation takes me everywhere on the map. It is not efficient especially when I am not familiar with the map. The nine keys tell me what are in the northwest and so on. It narrows me down to a specific range".

To address the irreversibility problem in arrow-key navigation, iSonic supports previous/next navigation to let users go through every region once and only once, following their order in the map sweep. Although a few participants mentioned the irreversibility problem, they thought it was a natural fact about maps and no one seemed to be bothered. No one used previous/next navigation after the training because "there is no need for it" or "it does not make sense on maps".

6.5.7 Situate. Participants used situating to get the table sorting status, the current table position, the current map and map position, and the number of selected regions. Many participants reset the interface before each task and did not use situating much since they remembered what they had done. However, all participants considered this function essential so they do not need to redo the work after a break.

6.5.8 Select, Search and Filter. All participants were able to use selection and switch their focus between "all regions" and "selected regions only", even across the two data views. Some participants requested the ability to select variables besides selecting regions. Participants also requested first-letter searching of regions. Filtering was not requested since the datasets are small.

6.6 Discussion

It is clear that iSonic enabled participants to find facts and discover data trends within georeferenced data even in unfamiliar geographical contexts.

The design choices in iSonic were overall easy to use and allowed participants to effectively explore data in the map and table views without special interaction devices.

The studies do have limitations. The participants might have made favorable comments because they wanted to please the experimenters. An average of 6 hours' use was not enough to go beyond the novice usage stage. Investigation of the tool's long-term use in real work circumstances will hopefully provide further understanding. Only users without any residual vision took part in the evaluation of iSonic. Further studies with partially sighted users may reveal different usage patterns and visual-auditory interactions that may modify the results and design framework for these users.

However, the studies provided clear evidence that the blind users dramatically benefited from the set of task-oriented user actions (AISA) and the use of multiple highly coordinated data views offered by the Action-by-Design-Component framework. Some widely used visualization techniques, such as the visual information-seeking mantra and the use of multiple highly coordinated visualizations, with appropriate adaptation, work in the auditory mode. The key conclusions and design implications were as follows.

- (1) All participants were capable of choosing and switching between highly coordinated table and map auditory views, in order to complete the tasks. It is possible that users could also deal with more and different views such as graphs.
- (2) Using musical pitches to present numerical data makes it easier to perceive data trends in data series and enhances close-up value pair comparison. The integrated use of musical sounds and speech allows users to listen to overall trends and to get details.
- (3) A single auditory feedback detail level is not sufficient. Our 4 levels were all used productively. While it is hard to understand a data element without the appropriate context, too much detail slows down the sequential presentation and can be overwhelming for gaining the big picture. Designers need to carefully select multiple information levels and let users adjust it to fit their tasks.
- (4) A rapid auditory gist is valuable in conveying overall data trends and guiding exploration. For maps, perceiving spatial relation from a sequence of sounds can be difficult but sweeping the map as separate smaller ranges in a consistent order was effective.
- (5) Navigation structures should reflect the data relation presented by the data view. In the map, designers would do well to provide 3x3 exploration using the numeric keypad and adjacency navigation using arrow keys. Users benefited from absolute localization and relative movements. Even a coarse map partitioning mapped to the physical spatial layout of a numeric keypad, can provide valuable geographical knowledge. Stereo sound panning can be helpful but seems to be secondary in giving location cues for most participants.

- (6) Selecting was valuable for all participants in focused data examination. They were able to operate selection within and across data views and accomplish brushing.

iSonic is also being tested in a separate study underway at the University of Rome la Sapienza. This study is comparing the performance of sighted users, users with acquired blindness, and users with congenital blindness, using either a tablet or solely the keyboard as input device for various map recognition tasks (see the iSonic Web page for links to the upcoming results).

7. EXTENSION TO GRAPHS

We believe that the ADC framework can be applied to other graphical data views such as line graphs and scatterplots. In the visual mode, a line graph clearly shows the value change for one or more variables over the period defined by another variable (typically time stamps). The table view in iSonic, with one variable sorted and a second variable swept automatically, produces the effect of a single variable line-graph sonification in which X presents the sorted variable and Y presents the second variable. Details on interactive line-graph sonification are available in Zhao [2006]. In this article, the focus is on the sonification of a bivariate scatterplot and its integration in iSonic.

Flowers et al. [1997] found that a sonified overview of a 50-data sample bivariate scatterplot was quite as efficient as the visual equivalence in conveying sign and magnitude of correlation. The sonification was produced by a dot-scanning method in which a vertical scan line moved along the X axis at a constant speed, and the Y value of each data sample (dot) was presented by a pitch when encountered by the scan line.

While information on individual dots can be interesting, sonifying every single dot could become overwhelming for scatterplots containing a large number of data items or many overlapping data items. To sonify a scatterplot, a dot aggregation method is used in which dots are aggregated into cells in a 2-D heat map. In Figure 3, the 2-D space of the scatterplot is equally divided into 9x9 grid cells. The choice of 9x9 resolution is guided by the 3x3 number pad exploration design. For each cell, the number of dots it contains is the value shown in the corresponding heat map. This spatial clustering and binning method is also fundamental in calculating entropy, a measurement strongly advocated by MacEachren et al. [2003] for analyzing 2-D scatterplots as opposed to the correlation coefficient.

Guided by the ADC framework, designing the scatterplot sonification was straightforward. Many interface designs for choropleth maps can be directly applied to the scatterplot heat maps such as the sound-encoding schemes, spatial sweep, 3x3 recursive 2-D space exploration, and arrow-key navigation, filtering, and so on. In terms of the actual implementation, much code from the map view can be reused in the scatterplot view with little modification due to the high similarity of their sonification and interaction methods. One main effort lies in the design and implementation of data structures to generate, store, and maintain dot coordinates and semantically zoomable heat map

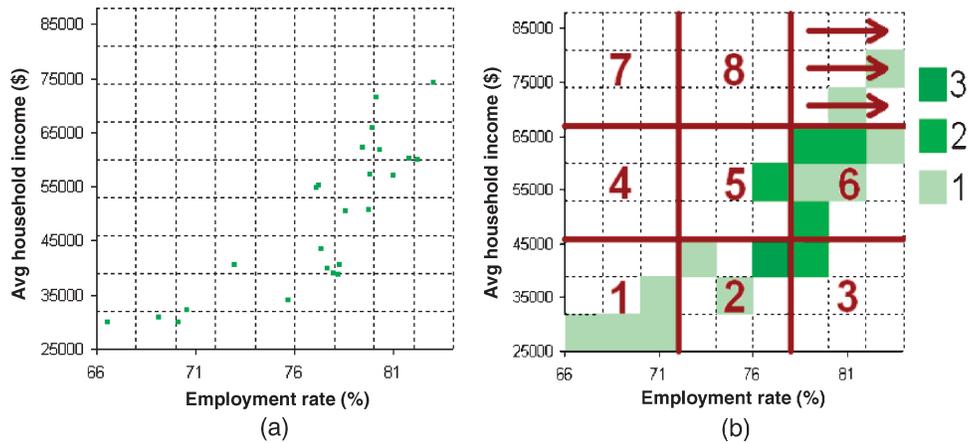


Fig. 3. (a) A bivariate scatterplot and (b) its heat map at the resolution of 9x9 grid.

representations of the scatterplot. The other part involves visually rendering the scatterplot. A brief description of the scatterplot view in iSonic and its integration with the table and map views follows.

7.1 Gist

A gist is produced by spatially sweeping the grid cells. Empty grid cells can be included or omitted. When empty cells are included, users can expect the same gist length, sweep path, and speed for all scatterplots using the same grid resolution. Each empty cell can add either a pause or a background sound to the gist.

7.2 Navigate

A heat map can be navigated in a relative or an absolute style like in a choropleth map. Figure 3(b) illustrates a 3x3 partition and the absolute navigation by activating a subgist of each of the 9 ranges. Users can zoom into a range or a grid cell for in-depth examination. Upon zooming, the heat map disaggregates. Figure 4(a) shows zooming into range 6 of the scatterplot in Figure 3, and Figure 4(b) shows the disaggregated heat map. The relative and absolute navigation methods can be recursively applied to the zoomed range.

7.3 Information Level and Details-On-Demand

Four information levels are provided. At level 0, the grid cell position is spoken. At level 1, the heat pitch is played. At level 2, the heat value is also spoken. At level 3, the IDs (the region names) of the dots in the cell are spoken as well. When users press the spacebar to request details, the X and Y value ranges of the current cell are given in addition to the information given at level 3.

7.4 Select

Selection is applied to individual grid cells in the heat map and affects all dots inside the operated cell. Users can switch among viewing “all regions”



Fig. 4. 3x3 zooming: (a) zoom into range 6 of the scatterplot in Fig 3; (b) the 9x9 heat map for the partial scatterplot in (a).

or “selected regions only”. In the later status, unselected dots fade into gray, and a new heat map is calculated to count only selected dots. This is useful especially when users select some regions in the table or map views and want to examine only those regions in the scatterplot view (see Section 7.6). It is also usually possible to select a single dot in a multidot cell by zooming into that cell to get a more scattered view. The dot-aggregation sonification method could be combined with the dot-scanning method as two sonification modes.

7.5 Situate and Filter

When users press ‘I’, iSonic speaks the name of the X and Y variables, the X and Y range of the current view, the number of dots in the view, the current navigation position, and so on. Filter is done in a similar way as for the map and table views.

7.6 Brush and Coordination with Other Data Views

The scatterplot view is highly coordinated with the table and map views during all AISAs such as selection, navigation, filtering, and adjusting information level. Users can change the X and Y variables of the scatterplot view by selecting the desired variables in the table view. Regions selected in the table or map view are automatically selected in the scatterplot view, and vice versa.

8. CONCLUSION AND FUTURE WORK

This article presents an Action-by-Design-Component framework for designing auditory interfaces for analytical data exploration. The framework was applied to georeferenced data, and a data exploration tool called iSonic was built. Evaluation of iSonic with 7 blind users showed that the rich set of task-oriented actions (AISAs) and the use of multiple coordinated data views offered by the ADC framework were effective and beneficial to blind users to accomplish

complex tasks. The application to scatterplots demonstrates the generalizability of the framework which can lead to an auditory workspace with multiple highly coordinated graphical data views and a set of unified user actions for general analytical data exploration. We hope that other designers and researchers will apply the ADC framework in their auditory graph development. The similarity among the sonification and interaction designs of choropleth maps and scatterplots indicate that some design methods may be effective across various graphs. Identifying what methods are commonly effective and what methods only work for specific graphs will help to extract rules for interactive sonification of many datasets.

ACKNOWLEDGMENT

We greatly appreciate the efforts of our participants and the work of Alex Silver in helping to run the studies. Thanks also to the reviewers and Associate Editor Marti Hearst who guided us through constructive revisions.

REFERENCES

- BENNETT, D. J. AND EDWARDS, A. D. N. 1998. Exploration of nonseen diagrams. In *Proceedings of the International Conference on Auditory Display*.
- BROWN, L., BREWSTER, S. A., RAMLOLL, R., BURTON, M., AND RIEDEL, B. 2003. Design guidelines for audio presentation of graphs and tables. In *Proceedings of the International Conference on Auditory Display*.
- BREWSTER, S. A. 1998. Using nonspeech sounds to provide navigation cues. *ACM Trans. Comput. Hum. Interact.* 5, 3, 224–259.
- CARD, S., MACKINLAY, J. D., AND ROBERTSON, G. 1990. The design space of input devices. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- CEAPARU I. AND SHNEIDERMAN, B. 2004. Finding governmental statistical data on the Web: Three empirical studies of the FedStats topics page. *J. Amer. Soc. Inform. Sci. Tech.* 55, 11, 1008–1015.
- DONKER, H., KLANTE, P., AND GORNY, P. 2002. The design of auditory user interfaces for blind users. In *Proceedings of the Nordi Conference on Human-Computer Interaction*.
- FLOWERS, J. H., BUHMAN, D. C., AND TURNAGE, K. D. 1997. Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors* 39, 3, 340–350.
- FRANKLIN, K. M. AND ROBERTS, J. C. 2003. Pie chart sonification. In *Proceedings of the Annual IEEE Symposium on Information Visualization*. 4–9.
- HANDEL, S. 1989. *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, Cambridge, MA.
- HERMANN, T. AND RITTER, H. 1999. Listen to your data: model-based sonification for data analysis. In M.R. Syed, Ed. *Advances in Intelligent Computing and Multimedia Systems*, International Institute for Advanced Studies in System Research and Cybernetics, 189–194.
- HUNT, A., HERMANN, T., AND PAULETTO, S. 2004. Interacting with sonification systems: closing the loop. In *Proceedings of the Annual IEEE Symposium on Information Visualization*.
- IGARASHI, T. AND HUGHES, J. F. 2001. Voice as sound: Using non-verbal voice input for interactive control. In *Proceedings of the 14th ACM Annual Symposium on User Interface Software and Technology*. 155–156.
- KAMEL, H. M. AND LANDAY, J. A. 1999. The integrated communication 2 draw (IC2D): A drawing program for the visually impaired. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- KARSHMER, A. AND GILLIAN, D. 2005. Math readers for blind students: Errors, frustrations, and the need for a better technique. In *Proceedings of the 2005 International Conference on Human-Computer Interaction (HCI)*.

- KRAMER, G., WALKER, B., BONEBRIGHT, T., COOK, P., FLOWERS, J., MINER, N., AND NEUHOFF, J. 1997. Sonification report: Status of the field and research agenda, <http://www.icad.org/websiteV2.0/References/nsf.html>.
- LANDAU, S. AND GOURGEY, K. 2001. Development of a talking tactile tablet. *Inform. Techn. Disabil. VII2*.
- LAZAR, J., ALLEN, A., KLEINMAN, J., AND MALARKEY, C. 2007. What frustrates screen reader users on the Web: A study of 100 blind users. *Int. J. Hum.-Comput. Interact.* 22, 3, 247–269.
- MAC EACHREN, A. M., DAI, X., HARDISTY, F., GUO, D., AND LENGERICH, G. 2003. Exploring high-D spaces with multiform matrices and small multiples. In *Proceedings of the IEEE Symposium on Information Visualization*. 31–38.
- MYNATT, E. D. AND WEBER, G. 1994. Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- NATIONAL FEDERATION OF THE BLIND. Braille usage: Perspectives of legally blind adults and policy implications for school administrators. <http://www.nfb.org/brusage.htm>.
- NORTH, C. AND SHNEIDERMAN, B. 2000. Snap-together visualization: A user interface for coordinating visualizations via relational schemata. In *Proceedings of the ACM International Working Conference on Advanced Visual Interfaces*.
- PAULETTO, S. AND HUNT, A. 2004. A toolkit for interactive sonification. In *Proceedings of the International Conference on Auditory Display*.
- PEREZ-QUINONES, M. A., CAPRA, R. G., AND SHAO, Z. 2003. The ears have it: A task by information structure taxonomy for voice access to Web pages. In *Proceedings of Interact*, IOS Press, The Netherlands.
- RAMLOLL, R., YU, W., RIEDEL, B., AND BREWSTER, S. A. 2001. Using non-speech sounds to improve access to 2-D tabular numerical information for visually impaired users. In *Proceedings of the BCS IHM-HCI Conference*.
- REENSKAUG, T. The Model-View-Controller (MVC): Its past and present, http://heim.ifi.uio.no/~trygver/2003/javazone-jaoo/MVC_pattern.pdf.
- ROTH, S. F., CHUAH, M. C., KERPEDIJEV, S., KOLOJEJCHICK, J. A., AND LUCAS, P. 1997. Towards an information visualization workspace: Combining multiple means of expression. *Hum.-Comput. Interact.* 12, 1–2, 131–185.
- SAUE, S. 2000. A model for interaction in exploratory sonification displays. In *Proceedings of the International Conference on Auditory Display*.
- SHNEIDERMAN, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Language*.
- SHNEIDERMAN, B. AND PLAISANT, C. 2005. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* 4th ed., Addison Wesley.
- SIKORA, C. A., ROBERTS, L. A., MURRAY, L. 1995. Musical vs. Real world feedback signals. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- SLATIN, J. AND RUSH, S. 2003. *Maximum Accessibility*. Addison-Wesley.
- SMITH, S., GRINSTEIN, G. G., AND BERGERON, R. D. 1990. Stereophonic and surface sound generation for exploratory data analysis. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 125–132.
- SUN MICROSYSTEMS. Java Web Start Technology, <http://java.sun.com/products/javaWebstart/>.
- W3C. Web content accessibility guidelines. <http://www.w3.org/WAI/intro/wcag.php>.
- WALKER, B. N. AND LANE, D. M. 2001. Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. In *Proceedings of the International Conference on Auditory Display*.
- WALKER, B. N. AND COTHAN, J. T. 2003. Sonification sandbox: A graphical toolkit for auditory graphs. In *Proceedings of the International Conference on Auditory Display*.
- WALL, S. AND BREWSTER, S. 2006. Feeling what you hear: Tactile feedback for navigation of audio graphs. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 1123–1132.

- WENZEL, E. M., ARRUDA, M., KISTLER, D. J., AND WIGHTMAN, F. L. 1993. Localization using nonindividualized head-related transfer functions. *J. Acoustical Soc. Amer.* 94, 1, 111–123.
- WILLUHN, D., SCHULZ, C., KNOTH-WEBER, L., FEGER, S., AND SAILLET, Y. 2003. Developing accessible software for data visualization. *IBM Sys. J.* 42, 4. (<http://www.research.ibm.com/journal/sj/424/willuhn.html>).
- ZAJICEK, M., POWELL, C., AND REEVES, C. 1998. A Web navigation tool for the blind. In *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility*. 204–206.
- ZHAO, H., PLAISANT, C., SHNEIDERMAN, B., AND DURAIWAMI, R. 2004. Sonification of georeferenced data for auditory information seeking: design principle and pilot study. In *Proceedings of the International Conference on Auditory Display*. Sydney, Australia. International Community for Auditory Display.
- ZHAO, H., SMITH, B. K., NORMAN, K., PLAISANT, C., AND SHNEIDERMAN, B. 2005. Listening to maps: User evaluation of multiple designs of interactive georeferenced data sonification. *IEEE Multimed.* (Special Issue on Interactive Sonification) 12, 2, 26–35.
- ZHAO, H. 2006. Interactive sonification of abstract data – framework, design space, evaluation, and user tool. Ph.D. dissertation, University of Maryland, College Park. (Also available as Human-Computer Interaction Lab. Tech. rep. HCIL-2006-06).
- ZOTKIN, D. N., DURAIWAMI, R., AND DAVIS, L. S. 2002. Customizable auditory displays. In *Proceedings of the International Conference on Auditory Display*. International Community for Auditory Display. 167–176.

Received July 2006; revised June 2007; accepted August 2007 by Marti Hearst