

# A Framework for Auditory Data Exploration and Evaluation with Geo-referenced Data Sonification

Haixia Zhao, Catherine Plaisant, Ben Shneiderman

Dept. of Computer Science & Human-Computer Interaction Lab  
Univ. of Maryland, College Park, MD 20742  
{haixia,plaisant,ben}@cs.umd.edu

Jonathan Lazar

Dept. of Computer and Info Science  
Towson Univ., Towson, MD 21252  
jlazar@towson.edu

## ABSTRACT

We first describe an Action-by-Design-Component (ADC) framework to guide auditory interface designs for exploratory data analysis. Applying the framework to the interactive sonification of geo-referenced data, we systematically explored and evaluated its design space. A data exploration tool, iSonic, was implemented for users with vision impairment. In depth case studies with 7 blind users showed that iSonic enabled them to find facts and discover data trends of geo-referenced data, even in unfamiliar geographical contexts. Analysis of user task behaviors and usage patterns confirmed that the framework has captured auditory information seeking actions and components that were naturally adopted by subjects to accomplish geo-referenced data exploration tasks. The results provide evidence for us to extend the framework, and guidance for designers of unified auditory workspaces for general exploratory data analysis.

## Author Keywords

Vision impairment, Sonification, Auditory user interfaces, Information seeking, Universal usability, Maps

## ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI)

## INTRODUCTION

Information visualization has produced many innovative techniques/interfaces for people with normal vision to use their tremendous visual ability to explore and discover data facts/trends. When information is presented with visual properties, such as color and spatial location, it is not easily viewable by users with vision impairment. In addition, visual data interaction is typically done by using pointing devices, such as computer mice, to directly manipulate the visual objects displayed on the screen. Such interaction is hard without sustained visual feedback. Although a few visualization tools (e.g., [26]) allow keyboard-only navigation

inside some visual graphs, most visualizations are not usable for users with vision impairment.

One example is the current access to government statistical data. Such data is often geography-related, such as population distribution by geographical regions, and often presented as choropleth maps that typically use colors to show the value for each map region. Required by Section 508 (www.section508.gov), all USA federal agencies need to make such data accessible.

A widely used accommodation for users with vision impairment to access digital information is to rely on screen readers to speak the textual content. To make non-textual elements accessible to screen readers, textual equivalences are needed. For static graphs, it is a standard practice to provide textual labels during the system development [22]. For dynamic graphs, tabular data presentations are used instead (e.g., [26]), or textual summaries can be automatically generated from the data set.

Several problems exist in the current approaches. First, while a concise textual description is helpful, the data interaction that is a critical part of data exploration process is lost. Automatic textual summarization techniques require pre-defined summary templates and do not have enough flexibility to support all user needs in exploratory data analysis. Second, a tabular presentation may be good for basic data browsing but is hard for in-depth data comprehension and analysis. Third, speech can accurately describe information but tends to be long in duration and hard to realize complex information.

Data interaction has been extensively investigated in visualization systems. But little was done regarding whether techniques in visualizations can be translated for use in auditory data exploration without visual aids, and what design implications are involved. Some research used musical sounds to present sonified “overviews” of simple graphs (e.g., [6]) but support for other task-oriented data interactions is typically missing.

We believe it is important to investigate whether an analogue to standard techniques in visualizations can be established for the auditory mode. In this paper, we first describe an Action-by-Design-Component (ADC) framework for designing auditory interfaces for analytical data exploration. We use a set of Auditory Information Seeking Actions (AISA) to characterize task-oriented data interaction without visual aids, identify Design Components for supporting AISAs, and discuss their general design considerations. This framework

has been used to investigate the design space of geo-referenced data sonification. In our earlier work, we reported on some initial sonification designs [27, 28], but the focus was limited to conveying data distribution patterns on maps and the studies were conducted with blind-folded sighted users.

Guided by the ADC framework, we now developed a general exploratory data analysis tool for users with vision impairment, called iSonic. iSonic contains a highly coordinated map view and table view, and supports AISAs within and across the two views (Fig 1). We will describe iSonic features and discuss the design rationale to illustrate the framework.

Afterwards, we report an empirical evaluation of the keyboard-only version of iSonic with 7 users with complete vision impairment (42 hours of in-depth observation and interview data.) which enabled us to examine the effectiveness of iSonic design choices. After extracting common iSonic usage patterns and analyzing user feedback, we discuss the benefits and limitations of the ADC framework.

**PREVIOUS WORK**

Sonification, the use of non-speech sound, has been used in various interface designs (e.g., non-visual GUI presentations [3, 13]), as well as data presentations [11]. Using the highly structured nature of musical sounds to convey information works even when no everyday auditory equivalence exists, is less tiring and generally more appropriate than everyday sounds [21]. Research has shown that musical sounds enhance numeric data comprehension (e.g. [16]) and humans can interpret a quick sonified overview of simple data graphs (e.g., [2, 6, 7]). Some guidelines were extracted (e.g., [2, 23]) and toolkits were developed to help researchers try different data-to-sound attribute mappings (e.g. [15, 24]). While some allow basic user movements in the graph (e.g., [2]), previous data sonification typically lacks supports for task-oriented data interactions.

In visual data exploration, the information seeking mantra “overview first, zoom, filter, then details-on-demand” [19] characterizes the general visual information seeking process and was an effective visualization design guideline. Several visualization interfaces (e.g., Sage[17], Snap-together[14]) were designed that allow users to construct multiple graphical data views and perform data exploration through unified interaction methods within and across the views.

However, such a framework or interface is absent for data exploration in the auditory mode without visual aids. Some recent models (e.g. [9, 18]) tried to describe interactive data sonification, but they emphasize spatial immersion effects in a physical world modeling of the data set hence may not be suitable for abstract data. More importantly, none has characterized task-oriented information seeking needs in the auditory mode, or addressed design considerations for interaction without visual aids, such as “can users with complete vision impairment operate multiple coordinated auditory views”.

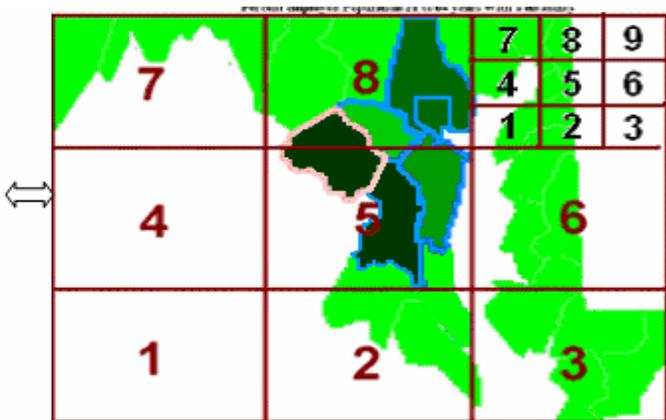
**AN ACTION-BY-DESIGN-COMPONENT FRAMEWORK**

In this section, we first describe the Auditory Information Seeking Actions (AISA), contrasting them with visual actions. Then we briefly mention some general design considerations for the Design Components to support AISAs. Those considerations will be reviewed in more details when we discuss the design of iSonic.

**Auditory Information Seeking Actions (AISA)**

We believe that an exploratory data analysis task in the auditory mode can be accomplished by a series of Auditory Information Seeking Actions (AISA). Many of the actions resemble those in visual information seeking but involve different cognitive processes and present special design challenges due to the highly transient nature of sound.

region	Populati...	Population Wi...	percent Pop...
Prince Geord	479163	69958	14.6
Montgomery	527217	64848	12.3
Baltimore	433820	52058	12.0
Baltimore city	363171	43581	12.0
Anne Arundel	284800	41866	14.7
Howard	151751	19576	12.9
Harford	127011	16511	13.0
Frederick	114675	13761	12.0
Charles	69664	12191	17.5
Washington	70943	10925	15.4
Carroll	86493	10379	12.0
St. Mary's	47934	9539	19.9
Alcomion	47416	9483	20.0



**Figure 1: highly coordinated table and map views of the counties of the state of Maryland. Superimposed on the color coded (choropleth) map is a representation of the recursive 3x3 keyboard exploration grid.**

Obtaining a gist is to experience the overall data trend 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 (see next section) because human auditory perception is much less synoptic than visual perception.

**Navigation** is “moving around” to examine portions of the data set by listening to a sub-gist of that portion. It needs to follow paths that are natural to the data relations. A visual interface provides a sustained display for users to directly manipulate. In auditory interfaces, users need to construct a mental representation of the display space and virtual navigation structures in order to efficiently move in the data set. Without a persistent display, they can easily get lost. To regain the orientation, users need to *situate* themselves by requesting their status. While navigation is an exploratory action, *searching* is a more fixed-goal action that directly lands on the data items by specifying search criteria. 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 helps to trim a large data set to a manipulable size, and allows users to quickly focus on items of interest. In visualization, dynamic query coupled with rapid (less than 100 milliseconds) display update is the goal [20]. In the auditory mode, different goals need to be established because such a short time is usually not enough to present a gist of changes. Results may need to be given after filtering is done instead of continuous display updates during the filtering process.

By **selecting**, users specify special interest in particular data items. Those data items are marked and can be revisited later or examined in other contexts. When the number of items is small, users can listen to the details. While speech is often too lengthy for obtaining an overall gist, it can be an effective presentation at the *details-on-demand* level.

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 feedbacks from multiple views need 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.

### Design Components

Each AISA consists of one or multiple interaction loops in which the user uses an *input device* to 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 and correctly interpret the auditory feedback.

A *data view* is a form of presenting the data items and their relations, such as a table, map, or line graph. Research has

shown that users with vision impairment were able to learn, interpret, and benefit from non-tabular data presentations. There is also evidence [1, 27] that choosing the right data view for a given task dramatically influences performance.

*Navigation structures* should reflect the data relations in the data view. In some previous work, users used a mouse or other input devices to move in the 2-D or 3-D data space to activate sounds of the data items within a certain distance from the cursor position. Such a “torch metaphor” [5] navigation could be useful for some data views, e.g., a scatterplot, but may be inefficient for others, e.g. a node-link diagram.

The choice of *input device* needs to consider 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. [4] categorized physical input devices by their physical manipulation properties and defined several choice factors such as the cost. We can maximize users’ situation awareness by matching an input device’ properties with those of the navigation structures. However, it is important to keep the system device-independent by providing good alternatives in the absence of the desired device. For success by users with vision impairment, a system should provide interactions optimized for keyboard-only operations.

As a general principle, the *auditory feedback* should have a low latency. It should be generally short to fit the short-term memory (STM) or allow pauses for midpoint STM processing. Short and responsive feedback increases user engagement and allows users to 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 [8]. The sounds of multiple items often need to be sequenced along the time dimension instead of being played all at once. This imposes special design challenges when no natural mapping exists from the data relation to the time dimension. When the number of data items is large, data aggregation may be necessary to design short feedback.

### ISONIC AND THE DESIGN SPACE FOR GEO-REFERENCED DATA

Guided by the ADC framework, we have systematically explored the design space for geo-referenced statistical data and designed iSonic (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 the evaluation of some design choices with blind-folded sighted subjects. Auditory interfaces are difficult to describe on paper, so we also submitted a supplementary video.

## Data Views

iSonic provides two highly coordinated data views – a region-by-variable table and a choropleth map (Fig 1). The table shows multiple statistical variables simultaneously. Each row corresponds to a geographical region and columns to variables. Table rows can be sorted by pressing ‘o’ while at the desired column, allowing quick location of low or high values. While geographical coordinates and adjacencies could be added as table columns, such information is better displayed on a map. Subjects in our previous study [27] strongly preferred a map over a table for discovering geographical value trends and performed better on pattern recognition tasks with a map than with a geographical knowledge enhanced table. Other views, such as line graphs or scatterplots, can be helpful for some analytical tasks, but were not used at the current work stage. We wanted to first examine how users could operate multiple coordinated auditory views. Auditory and visual displays are synchronized to allow communication between sighted and blind users.

## Input Devices and Navigation Structures

When choosing input devices, we considered both device availability and how effectively their physical properties match the navigational properties of the two data views.

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 blind users are very comfortable using it. We use the arrow keys as natural means for relative movements in the left, right, up, and down directions. The numerical keypad potentially allows relative movements in 8 directions. The keyboard can also be transformed into a low resolution 2-D absolute pointing device, e.g., by mapping the whole keyboard layout to 2-D screen positions. In iSonic, we map the 3x3 layout of the numeric keypad.

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 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 [12], but we chose not to rely on them because 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.

## 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. The same mapping is used in the table view. Various musical instruments are used to

indicate when users are outside the map or crossing a region border in the touchpad interface, or crossing a water body to reach a neighboring region in the keyboard interface. Stereo panning effects are used to indicate a region’s azimuth position on the virtual auditory map. It is also used in the table to indicate the column order. 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.

There are many alternatives. Sound duration can present the value but would significantly prolong the feedback and is not appropriate when values of many regions need to be presented. Region locations could be mapped to sound locations using virtual spatial sound synthesized with Head Related Transfer Functions (HRTF) [25]. Spatial sound 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 we have connected iSonic to a virtual spatial sound server and plan to investigate the use of individualized HRTF spatial sound, we currently focus on MIDI stereo sound. We also tried to play a piano pitch after each violin value pitch to indicate the region’s elevation position. Unfortunately, such extra sound was not found to increase performance [28].

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.

## Gist

In the table view, a gist is produced by automatically 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, 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 sub-gists of parts of the map. 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 lasts for 9 seconds.

## Navigate

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

sub-gist 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 brings special design challenges. Relative movements between neighboring regions reveal region adjacency but do not convey region shapes, sizes, or absolute locations. Subjects in our previous studies reported that they only had weak location awareness by using this navigation method. 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 the problems, we tested cell-by-cell movements on a mosaic version of the map [28]. However, it did not improve users' location awareness, and was much less preferred because it required more keystrokes to move around.

We expect that navigations based on absolute pointing may help. Kamel and Landay first used a 3x3 grid recursion method via the keypad in a drawing tool [10]. In iSonic, the map is divided into 3x3 ranges (Fig 1) and users use 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 use arrow keys to move around. Pressing '0' sweeps the current zoomed map range or the whole map.

With the touchpad, 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 lifts 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 to compare them in details.

#### **Details-on-demand**

Pressing 'space' plays the details of the current region. Another way to get the 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.

#### **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.

#### **Select**

In both views, users can press 'L' (as for 'Lock') to select/unselect the current region and press 'A' to switch between "all regions" and "selected regions only". In 'selected regions only', AISAs only activate sounds of the selected regions.

#### **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 will be played. By sweeping each of the 9 map ranges, users can roughly but rapidly locate the region on the map.

#### **Filter and Search**

Filtering was done by slider-based queries. It is complex even for sighted novice users and was not evaluated in the current study. Searching is obviously helpful but was not implemented at the time of the study.

#### **EVALUATION**

During early design iteration 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 [27, 28]. However, an exploratory data analysis task is a complex process that involves many interface components. During the process, many inherent human subject variations can come into play, such as experience and cognitive styles. In order to obtain insights into users' auditory information seeking behaviors, we chose to conduct case studies. 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, we conducted intensive case studies with 7 local blind users, producing 42 hours of observation and interview data, with an average of 6 hours per user. Using cross-case analysis, we were able to extract common user behaviors and feedback that allowed us to (1) evaluate the effectiveness of iSonic design choices; (2) identify features helpful to each data exploration task category and examine the utility of the ADC framework; (3) identify task road blocks in order to target training and modifications to the interface and the framework.

#### **Subject Profiles**

All seven subjects possessed basic computer skill and relied on screen readers to access computer information. They were all comfortable with maps and tables, had experience with numerical data sets, and used government statistical data at work. All subjects 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 15 (P1, P5, P6, P7). None of them had residual vision. Among the born blind, 2 were males, one 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 subjects who became blind after 15, one 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 two females, one had a college degree (P5) and the other had a master

degree (P6), both in social science. All subjects volunteered to participate, and were compensated for their time.

### **Material and Tasks**

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

Three data sets were used, one for training, one for testing, and one for post-test free exploration. The data 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 data set contained 8 variables and was about the 50 US states plus the District of Columbia. The test data set contained 12 variables and was about the 24 counties in the state of Maryland. The post-test data set was about the 44 counties in the state of Idaho (but subjects were not told what it was). Subjects' 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 data set. 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 subjects. The testing tasks are summarized below.

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. Subjects also performed a similar set of testing tasks in Microsoft Excel

2002 with their usual screen readers (all happened to use JAWS), and compared the task experience. It was not our intention to compete with Excel. Rather, we considered Excel as the standard tabular data viewer, and used the comparison as a method to solicit user comments on what interface features were helpful to each task. All subjects had some previous experience with Excel, while some were expert users. We did not provide tactile maps when subjects used Excel, because many blind users do not have access to tactile maps (only P5 owns one for Maryland).

### **Procedure**

Each case study was carried out in two sessions on consecutive days, at the subject's home or office. In the first session, the subject listened to a self-paced auditory step-by-step tutorial, tried out all iSonic features and practiced seven sample tasks with the training data set. For each training task, a sample solution and the correct answer were given. Subjects 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 subjects with limited Excel experience were given time to practice. After adjusting the speech rates to the subjects' satisfaction, they performed seven tasks similar to the training tasks in both Excel and iSonic. For each pair of tasks, the subject 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 data set but involved different variables, so data learning between tasks can be ignored. We asked the subject to do the Excel task first because we wanted 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 task might also have benefited strategically from the iSonic training task. An interview was conducted after subjects performed all the testing tasks in both interfaces. Finally, subjects were asked to freely explore an unknown map and data (the post-test data set) for 5 minutes and report things they found interesting. This was to observe what users would do when they encountered a new map and data.

### **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, subjects successfully completed 67% of the training tasks without referring to the sample solution or any other help. After the training, subjects were able to retain their newly acquired knowledge and successfully completed 90% tasks on the next day in a different context without any help. For 74% of the tasks that subjects used different strategies than the given solution in the training, they adopted the sample strategies in the test session.

For tasks that did not require geographical knowledge, the average testing success rates were similar for iSonic and Excel, both at 86%, although subjects ranked iSonic easier than Excel, at 7.9 vs. 7.0 based on a 10-point scale (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 subjects with excellent knowledge about Maryland geography (P3 and P5) achieved a success rate of 67%. Other subjects 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, subjects gave iSonic an average of 8 on easiness for all the 3 geography-related tasks, and gave 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 subjects 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.

Overall it was easy for the subjects 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 subjects 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, swept other columns with “pitch only” to check which one has more consistently increasing pitch pattern. Other subjects 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 subject (P6) said she would have the data plotted in a scatterplot or multi-line graph and had her human reader look for the highest correlation. All subjects, 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 subjects except for P7 who guessed the answer correctly but was very uncomfortable. 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 [28].

Setting aside the above strategic difficulties, incorrect answers in iSonic were caused by two common errors: (1) subjects sorted the wrong variable (a third of all errors). This might be due to the high similarity of variable names, and that the interface did not confirm the variable being sorted when the sorting key was pressed. (2) Some subjects 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 subjects to the 2nd row instead of the 1st.

### **Task Strategies and iSonic Usage Patterns**

**Map vs. table:** All subjects 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 acquire/confirm region locations. 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 subjects to find regions.

**Brush:** All subjects 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 subjects. “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 subjects 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 subjects skipped the use of the map and used their pre-test geographical knowledge to judge if the selected items satisfied the geographical restriction. A few subjects 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 subjects’ memory of the intermediate results and caused some errors. Subjects 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 subjects 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 meet the value restriction, or switched to table and used sorting to compare their values.

**Use of Pitch:** Using pitches to present numeric values was considered intuitive, entertaining and very helpful to data comprehension. It took some subjects 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. “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 subjects 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 waited for a number to be spoken for confirmation purpose (in value comparison). Some subjects were able to tell the absolute value category using only one

pitch while some needed to use other pitches as references. All subjects, except P4, were comfortable with the simultaneous pitch and speech presentations. P4 reported that pitches and speech interfered with each other, and requested to tone down the pitch volume. However, she declined the suggestion to completely remove pitches, because she used pitches exclusively in trend discovery.

**Information Level and Details-on-demand:** All subjects frequently adjusted the information level during a task. Subjects mostly used name plus pitch or name plus pitch together with the value in speech. When the information level with value in speech is used, many subjects 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, e.g., for geographical patterns or correlations, most subjects used the pitch-only level because it let them skim through the data the fastest. A few subjects 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 “name only” level. Details-on-demand was mostly done by increasing the information level to the maximum level instead of pressing the ‘space’ key.

**Gist:** Table sweep was very intuitive. To check value patterns, e.g., for the correlation tasks, some subjects 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 post-test 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 subjects said it was not difficult to understand how the sweep was done, but they need 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 subjects, except P7, were able to easily tell if a variable has a given geographical distribution pattern, by sweeping the nine ranges in pitch only. Unexpectedly, map sweep was also frequently used by all subjects 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 have been selected.

**Navigate:** Navigating the table was easy. All subjects mostly used cell mode because “it allows finer control of what to play”. The row and column mode was used by some subjects to sweep a column for the correlation and close-up

comparison tasks.

All subjects reported that overall it was very easy to navigate the map. The 3x3 exploration was frequently used by all subjects except P2 who mainly used arrow keys to navigate and used sound panning to judge region locations. All subjects 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 are in each map part. While subjects 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 contains that region, then used arrow keys to move to that region. The 3x3 exploration also allowed some subjects 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.

Subjects seemed to be able to zoom into/out of the 9 map ranges and stay aware of their zooming positions. Many subjects played with zooming extensively in training but did not use it in the test. Their explanation was there was no need from 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 subjects expressed the concern that zooming may become mentally intensive as the map scale grows. One observed problem with 3x3 zooming is 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 subjects in adjacency tasks. It was also used often 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 subjects were inclined to use the 3x3 exploration because it gave the absolute region locations.

“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 subjects 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”.

**Situate:** Subjects 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 subjects reset the interface before each task and did not use situating much

since they remembered what they had done. However, all subjects considered this function essential so they do not need to redo the work “after a bathroom visit”.

**Select, Search and Filter:** All subjects were able to use selection and switch their focus between “all regions” and “selected regions only”, even across the two data views. Some subjects requested the ability to select variables besides selecting regions. Subjects also requested first-letter searching of regions. Filtering was not requested since the data sets are small.

## DISCUSSION

It is clear that iSonic enabled subjects to find facts and discover data trends within geo-referenced data, even in unfamiliar geographical contexts. The design choices in iSonic were overall easy to use and allowed subjects to effectively explore data in the map and table views without special interaction devices.

The studies do have limitations. The subjects 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 provide further understanding. We only tested users without any residual vision. Further studies with partially sighted users may reveal different usage patterns and visual- auditory interactions that may modify our results and framework.

However, the studies provided clear evidence that the Action-by-Design-Component framework captured the actions that were naturally employed by blind users during data exploration. The framework, which is analogous to many interactions in visualizations, works for auditory interfaces when applied properly. The key conclusions and design implications were:

- (1) All subjects were capable of choosing and switching between highly coordinated table and map auditory views, in order to complete the tasks. We believe 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 subjects.

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

## CONCLUSION AND FUTURE WORK

We described an Action-by-Design-Component framework for designing auditory interfaces for analytical data exploration. We applied the framework to geo-referenced data and evaluated the resulting interface with 7 blind users. The effectiveness of the interface suggests that the framework may also be useful for other designers. In the future, we hope to extend this framework by applying it to the interactive sonification of graphs. By defining a programming paradigm and incorporating more data views into iSonic, we intend to develop a unified auditory workspace for general analytical data exploration.

## ACKNOWLEDGMENTS

We thank our subjects for participating and thank Alex Silver for helping to run the studies. This work is supported by the National Science Foundation under Grant No. EIA 0129978 and ITR/AITS 0205271.

## REFERENCES

1. Bennett, D.J., Edwards, A.D.N., Exploration of non- seen diagrams, Proc. Intl. Conf. on Auditory Display (ICAD), (1998).
2. Brown, L., Brewster, S.A., Ramloll, R., Burton, M., Riedel, B., Design guidelines for audio presentation of graphs and tables, Proc. ICAD (2003).
3. Brewster, S.A., Using nonspeech sounds to provide navigation cues, ACM Trans. on CHI, 5, 3 (1998), 224-259.
4. Card, S., Mackinlay, J.D., Robertson, G., The design space of input devices, Proc. ACM CHI (1990).
5. Donker, H., Klante, P., Gorny, P., The design of auditory user interfaces for blind users, Proc. NordiCHI (2002).
6. Flowers, J.H., Buhman, D.C., Turnage, K.D. Cross- modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. Human Factors, 39, 3 (1997), 340-350.
7. Franklin, K.M., Roberts, J.C., Pie chart sonification, Proc. IEEE Information Visualization (2003), 4-9.
8. Handel, S. Listening: An Introduction to the Perception of Auditory Events. MIT Press (1989).
9. Hunt, A., Hermann, T., Pauletto, S., Interacting with sonification systems: closing the loop, Proc. IEEE Information Visualization (2004).

10. Kamel, H.M., Landay, J.A., The integrated communication 2 draw (IC2D): a drawing program for the visually impaired. Proc. ACM CHI (1999).
11. Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J., Sonification Report: Status of the Field and Research Agenda (1997) <http://www.icad.org/websiteV2.0/References/nsf.html>
12. Landau, S., and Gourgey, K., Development of a talking tactile tablet, Information Technology and Disabilities, VII(2), 2001.
13. Mynatt, E.D., Weber, G., Nonvisual presentation of graphical user interfaces: Contrasting two approaches, Proc. ACM CHI (1994).
14. North, C., Shneiderman, B., Snap-Together Visualization: A user interface for coordinating visualizations via relational schemata, Proc. Advanced Visual Interfaces (2000).
15. Pauletto, S., Hunt, A., A toolkit for interactive sonification, Proc. ICAD (2004).
16. Ramloll, R., Yu, W., Riedel, B., Brewster, S.A., Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users. Proc. BCS IHM-HCI (2001).
17. Roth, S. F., Chuah, M. C., Kerpedjiev, S., Kolojejchick, J. A., Lucas, P., Towards an information visualization workspace: combining multiple means of expression, Human-Computer Interaction, 12, 1-2 (1997), 131-185.
18. Saue, S., A model for interaction in exploratory sonification displays, Proc. ICAD (2000).
19. Shneiderman, B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, Proc. IEEE Symposium on Visual Language (1996).
20. Shneiderman, B., Plaisant, C., Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th Edition, Addison Wesley (2005).
21. Sikora, C.A., Roberts, L.A., Murray, L., Musical vs. Real world feedback signals, Proc. ACM CHI (1995).
22. W3C, Web Content Accessibility Guidelines, <http://www.w3.org/WAI/intro/wcag.php>
23. Walker, B.N., Lane, D.M., Psychophysical scaling of sonification mappings: a comparison of visually impaired and sighted listeners, Proc. ICAD (2001).
24. Walker, B.N., Cothran, J.T., Sonification sandbox: a graphical toolkit for auditory graphs, Proc. ICAD (2003).
25. Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L., Localization using nonindividualized head-related transfer functions, Journal of the Acoustical Society of America, 94, 1 (1993),111-123.
26. Willuhn, D., Schulz, C., Knoth-Weber, L., Feger, S., Sallet, Y., Developing accessible software for data visualization, IBM Systems Journal, 42, 4 (2003). <http://www.research.ibm.com/journal/sj/424/willuhn.html>
27. Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R., Sonification of geo-referenced data for auditory information seeking: design principle and pilot study, Proc. ICAD 2004.
28. Zhao, H., Smith, B.K., Norman, K., Plaisant, C., and Shneiderman, B., Listening to maps: user evaluation of multiple designs of interactive geo-referenced data sonification, IEEE Multimedia Special Issue on Interactive Sonification , Apr-Jun 2005.