

ABSTRACT

Title of Document: INTERACTIVE SONIFICATION OF
ABSTRACT DATA – FRAMEWORK, DESIGN
SPACE, EVALUATION, AND USER TOOL

Haixia Zhao, Doctor of Philosophy, 2006

Directed By: Professor Ben Shneiderman
Department of Computer Science

For people with visual impairments, sound is an important information channel. The traditional accommodation for visually impaired users to access data is to rely on screen readers to speak the data in tabular forms. While speech can accurately describe information, such data presentation tends to be long and hard to realize complex information. This is particularly true in exploratory data analysis in which users often need to examine the data from different aspects. Sonification, the use of non-speech sound, has shown to help data comprehension. Previous data sonifications focus on data to sound attribute mapping and typically lack support for task-oriented data interaction.

This dissertation makes four contributions. (1) An Action-by-Design-Component (ADC) framework guides auditory interface designs for exploratory data analysis. The framework characterizes data interaction in the auditory mode as a set of Auditory Information Seeking Actions (AISA). It also discusses design considerations for a set of Design Components to support AISAs, contrasted with actions in visualizations. (2)

Applying the framework to geo-referenced statistical data, I explore its design space. Through user evaluations, effective design options were identified and insights were obtained regarding human ability to perceive complex information, especially those with spatial structures, from interactive sounds. (3) A tool, iSonic, was developed, with synchronized visual and auditory displays. Forty-two hours of case studies with seven blind users show that iSonic enables them to effectively explore data in highly coordinated map and table views without special devices, to find facts and discover data trends even in unfamiliar geographical contexts. Preliminary algorithms are also described to automatically generate geographical region spatial sweep orders for arbitrary maps. (4) The application to geo-referenced data demonstrated that the ADC framework provided a rich set of task-oriented actions (AISAs) that were effective for blind users to accomplish complex tasks with multiple highly coordinated data views. It also showed that some widely used techniques in visualization can adapt to the auditory mode. By applying the framework to scatterplots and line graphs, I show that the framework could be generalized and lead to the design of a unified auditory workspace for general exploratory data analysis.

INTERACTIVE SONIFICATION OF ABSTRACT DATA
– FRAMEWORK, DESIGN SPACE, EVALUATION, AND USER TOOL

By

Haixia Zhao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Ben Shneiderman, Chair and Advisor
Associate Research Scientist Catherine Plaisant, Co-chair and Co-advisor
Assistant Professor Ramani Duraiswami
Associate Professor Carol Espy-Wilson
Assistant Professor François Guimbretière

© Copyright by
Haixia Zhao
2006

Acknowledgements

First, I would like to thank my advisor, Dr. Ben Shneiderman, for his inspiration, support and guidance. Ben first helped me build my interest in HCI through his well-renowned HCI class. I was fortunate to later on work with him, first as a teaching assistant, then as a research assistant, and finally as a colleague. His respect for and trust in me have really encouraged me to proceed with confidence. He has also inspired me in so many ways, both in research and in life.

I was very fortunate to have Dr. Catherine Plaisant to be my co-advisor. She was always available for intellectual discussions and constantly provided constructive feedback. My Ph.D. research would have been extremely difficult if not for her trust, support, and close guidance.

There are several faculty members, student colleagues, and external partners whose inputs contributed to the advance of my Ph.D. research. Dr. Ramani Duraiswami introduced me into the realm of virtual spatial sound, and provided the spatial sound server used in the research. Dr. Carol Espy-Wilson and Dr. François Guimbretière provided invaluable fresh perspectives to my dissertation. Benjamin K. Smith, Dr. Kent Norman, and Dr. Jonathan Lazar helped me with user evaluations. Dr. Dmitry Zotkin and Dr. Elena Grassi helped me work with the spatial sound server. Dan Newsome introduced me to his blind students Yasmin Reyazuddin and Ahmad Zaghafal who both gave invaluable feedback.

I would like to thank all of HCIL members. Cheerful and encouraging faculty

members and friendly student colleagues supported my research by making HCIL a great place to work. Bill Kules deserves my special thanks for his continuous constructive feedback on my research and papers. Anne Rose and Kiki Shneider always smiled even when I distracted them with many questions and requests. Dr. Ben Bederson, Dr. Evan Golub, Bongshin Lee, Jinwook Seo, Bongwon Suh, Gene Chipman, Hyunmo Kang, Aaron Clamage, Hilary Hutchinson, Alex Aris, Jerry Fails, Amy Carlson, Adam Perer, Chunyuan Liao, Nicholas Y. Chen and other HCIL members have provided with a supportive, engaging, and often intellectually challenging research environment in HCIL.

Lastly, my special thanks go to my family members. My parents deserve my appreciation deep in my mind through their devotion to me since my birth. Their everyday diligence set an example for my life. Without their devotion and sacrifice, it could not have been possible for me to be what I am. I am very fortunate to have a loving husband Chi. He was always there for me during my journey in the U.S., cheered for my accomplishments, comforted me through adversities, and supported me as an intellectual colleague as well as a wonderful husband. My Ph.D. life simply could not be successful without his love and support.

Table of Contents

List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
Chapter 2 Related Work.....	7
2.1 Abstract Data	7
2.2 Abstract Data Visualization	10
2.3 Sonification.....	12
2.4 Non-speech Sounds.....	13
2.5 Non-visual Map Access	15
2.6 Auditory Perception.....	20
2.7 Data Sonification	22
2.7.1 Scientific Data Sonification	22
2.7.2 Abstract Data Sonification.....	23
2.8 Interacting with Data Sonification	27
2.9 Discussion.....	30
Chapter 3 An Action-by-Design-Component Framework.....	32
3.1 Auditory Information Seeking Actions (AISAs).....	32
3.1.1 Gist.....	33
3.1.2 Navigate, Situate, and Search	33
3.1.3 Filter.....	34

3.1.4 Details-on-demand and Adjust Information Level	34
3.1.5 Select.....	34
3.1.6 Brush	35
3.2 Design Components.....	35
3.2.1 Data View	36
3.2.2 Navigation Structure	37
3.2.3 Input Device.....	39
3.2.4 Auditory Feedback.....	44
3.3 Summary and Discussion.....	50
Chapter 4 iSonic and the Design Space for Geo-referenced Data	52
4.1 Explorative Designs and Evaluations	53
4.1.1 Geographical Data Pattern Recognition.....	53
4.1.2 Effect of Data View	55
4.1.3 Effect of Navigation Structure and Auditory Feedback.....	67
4.1.4 Lessons and Discussion	80
4.2 iSonic Design Components.....	81
4.2.1 Data Views.....	81
4.2.2 Input Devices and Navigation Structures	83
4.2.3 Auditory Feedback.....	85
4.3 iSonic Auditory Information Seeking Actions (AISA)	86
4.3.1 Gist.....	86
4.3.2 Navigate	88
4.3.3 Details-on-demand and Adjust Information Level	96

4.3.4 Situate and Search.....	96
4.3.5 Select.....	97
4.3.6 Brush.....	97
4.3.7 Filter and Search.....	98
4.4 iSonic Menu, Hotkeys, and Online Help.....	107
4.4.1 Menu and Hotkeys.....	107
4.4.2 Online Help.....	112
4.5 A Sample Task Solution.....	113
4.6 Discussion.....	117
Chapter 5 iSonic Implementation.....	120
5.1 Architecture Overview.....	120
5.2 Input Files.....	126
5.3 Map Preprocessing.....	132
5.4 Generate Map Spatial Sweep Orders.....	135
5.5 Generate Map Adjacency Matrices.....	142
5.6 Discussion.....	145
Chapter 6 Evaluations.....	147
6.1 Geographical Data Pattern Recognition.....	147
6.1.1 Effect of Input Device and Blindness.....	147
6.2 Case Studies.....	149
6.2.1 Subject Profiles.....	150
6.2.2 Material and Tasks.....	151
6.2.3 Procedure.....	153

6.2.4 Description of Seven Cases	154
6.2.5 Cross-case Task Completion and Error Analysis	156
6.2.6 Task Strategies and iSonic Usage Patterns.....	158
6.2.7 Discussion	164
6.3 Summary and Discussion.....	166
Chapter 7 Interactive Sonification of Abstract Data Graphs	169
7.1 Scatterplot.....	169
7.1.1 Dot Scanning.....	170
7.1.2 Dot aggregation.....	173
7.2 Line graph.....	181
7.3 Implementation	184
7.4 Discussion.....	186
Chapter 8 Contributions and Future Work.....	188
8.1 Contributions	188
8.2 Future work.....	190
Appendix A - 1: Description of Case One (P1)	196
Appendix A - 2: Description of Case Two (P2)	201
Appendix A - 3: Description of Case Three (P3)	207
Appendix A - 4: Description of Case Four (P4).....	213
Appendix A - 5: Description of Case Five (P5).....	219
Appendix A - 6: Description of Case Six (P6)	225
Appendix A - 7: Description of Case Seven (P7).....	231
Appendix B: Algorithms for Region Set Split.....	237

Appendix C: Algorithms for Finding Target Region in Map Adjacency Navigation	239
Bibliography	241

List of Tables

Table 2-1: A depiction of a case-by-variable Data Table (from [40], p18).....	8
Table 3-1: Input device categorization and examples.....	42
Table 3-2: Commonly perceived attributes of a single musical sound.....	46
Table 4-1: User key controls in the first study.....	56
Table 4-2: User key controls in the second user study. Only keys from the number pad plus the four arrow keys were used.....	69
Table 4-3: Average recognition accuracy for different general pattern types (n = 48). ...	77
Table 4-4: Some hotkeys in the iSonic implementation.....	111
Table 6-1: The seven test tasks for iSonic in the case studies.....	152
Table 6-2 : Task completion of all the seven subjects in the case studies.....	155
Table 6-3 : Task success rates in the case studies.....	155

List of Figures

- Figure 2-1: Examples of graphical representations of an abstract data set. (a) a case-by-variable table, (b) a line graph containing three cases, (c) a bivariate scatterplot, (d) a graph showing case dependency defined by a variable, and (e) a choropleth map showing the values of one variable for all cases (U.S. states). 11
- Figure 2-2 : The VTPlayer tactile mouse is equipped with two 4x4 mechanical matrix units that can be used to deliver a tactile sense of the screen. For example, as users explore a U.S. map, the tactile patterns change over different states. 17
- Figure 2-3 : An example of tactile maps augmented with speech description when mounted on a touch screen (Touch Graphics, undated): (a) sample map plate, and (b) the sample map plate in use with the talking tactile tablet (TTT). 18
- Figure 2-4 : The BATS sonic map helps those without sight to explore the map and locate map objects [127]. Real world audio icons (car noises, wave crashing, etc.) are played as the user navigates the map and appear clearly as coming from the left or the right. Volume increases as users get closer to the source of the sound. 20
- Figure 3-1: Design components (in red) involved in data interactions in an auditory interface. An Auditory Information Seeking Action (AISA) is an interaction loop that involves using an input device to issue a command to the system and listening to the auditory feedback..... 36
- Figure 3-2: Device transformation: (a) an IBM Thinkpad laptop keyboard is remapped to a 2-D display space and becomes a low resolution discrete absolute pointing device; (b) another method to transform a keyboard to an absolute pointing device is to recursively divide the 2-D space to 3x3 ranges and to map them to the 3x3 layout of the nine number keys on the number pad. 43
- Figure 3-3: Device transformation: a touch screen can be used to perform relative and discrete movements if areas are designated to activate predefined key actions, e.g.,

the number keys, and up, down, left, and right arrow keys.	44
Figure 4-1: Spatial choropleth map design: (a) spatial sounds tied to the map create the effect of a virtual half-cylinder shaped map surrounding the subject located at the center. The illustration does not reflect the real spatial parameters; (b) vertical map sweep order.	58
Figure 4-2: Five general pattern types	60
Figure 4-3: A sample of four visual pattern choices. A darker green presents a higher value. Values are categorized into five ranges.....	61
Figure 4-4: Overall accuracy of pattern recognition (n = 9). The answer of the pattern type after exploration was inferred from the pattern detail chosen after the exploration.	63
Figure 4-5: Overall speed of pattern recognition.....	63
Figure 4-6: User preference.	65
Figure 4-7: Two navigation methods on a choropleth map: (a) state-by-state column navigation; (b) cell-by-cell mosaic navigation.	70
Figure 4-8: Sample patterns for each pattern type.....	72
Figure 4-9: A sample of four visual pattern choices.....	74
Figure 4-10: Highly coordinated table and map views in iSonic. The result of any action in one view is always reflected in another. The screenshots show a sample data set for Maryland counties. The counties in the table view are in a descending order by “Population with a disability”. In both views, selected counties are highlighted in blue, and the county the user is currently at is highlighted in pink.	82
Figure 4-11 : Recursively map the 3x3 layout of the number pad to a 2-D space.....	84
Figure 4-12: Two spatial sweep orders on a U.S. state map: (a) horizontal sweep (b) and vertical sweep.....	86
Figure 4-13: Two table navigation modes: (a) in the cell mode, users press arrow keys to move left, right, up, down (shown in blue arrow lines), or to left edge, right edge,	

top edge, and bottom edge; (b) in the row/column mode, when users move from one column to another (blue arrow lines), the whole target column is swept from top to bottom (green dotted lines); (c) in the row/column mode, when users move from one row to another (blue arrow lines), the whole target row is swept from left to bottom (green dotted lines). 88

Figure 4-14: Two map relative navigation methods: (a) region-by-region navigation may have a reversibility problem: moving up from B usually goes to C. But if users start from A, move down to B, then move up, they will end up in C although users may expect to go back to A; (b) using a cell-by-cell relative navigation method on a mosaic version of the map, users may be able to sense the region size and shape, to avoid the reversibility problem, and possibly to have better location awareness. 89

Figure 4-15: Census regions and divisions of U.S..... 90

Figure 4-16: Two types of top level 3x3 partitions of a U.S. state map: (a) a manual partition based on U.S. Census Regions in Figure 4-15; (b) a partition automatically generated by iSonic according to regions' geometric center coordinates..... 91

Figure 4-17 : Zooming based on a fixed map partition: (a) press Ctrl + 9 to zoom into range 9 of the map partition in Figure 4-16 (a). States outside of range 9 are grayed out and disabled; (b) recursively explore range 9 according to an automatic 3x3 partition; (c) use arrow keys to move among adjacent states in range 9. 92

Figure 4-18: Zoom into the neighborhood of the specific region: (a) zoom into the neighborhood of Pennsylvania (highlighted in orange) from the map in Figure 4-16(a). States whose geometric center fall out of the area are grayed out and disabled; (b) recursively explore the neighborhood in the 3x3 style according to an automatic 3x3 partition; (c) use arrow keys to move among adjacent states in the neighborhood. 93

Figure 4-19: Touchpad exploration: (a) the touchpad is calibrated so that the map is mapped to its entire surface. The region under the finger position (highlighted with a bold pink border) is activated; (b) when users zoom into range 9, the touchpad is recalibrated so the current map range (range 9) is mapped to its entire surface,

providing higher exploration resolution. The region under the finger position (highlighted with a bold pink border) is activated.....	95
Figure 4-20: Coordination between views during zooming: (a) in the map view, zoom into range 9 of a U.S. state map. The state highlighted with a bold pink border is the last state visited; (b) after switching to the table view, only states in the zoomed range are shown in the table, and the last state visited in the map view becomes the current state in the table view.	96
Figure 4-21: Coordination between views for selection: any regions selected (highlighted with blue borders) in the table view are always selected in the map view and vice versa.	97
Figure 4-22: The filter control panel contains multiple two-thumb sliders, with each slider representing one variable. The current slider (the third slider) is highlighted in blue.	99
Figure 4-23: An example of filtering sequence applied to the map view: (a) no filtering; (b) increase the lower bound by two units. 17 states are filtered out and turn gray. A spatial sweep of the 17 states (illustrated in orange arrow lines) is played; (c) decrease the upper bound by two units. 1 more state becomes filtered. A spatial sweep containing only this state is played; (d) move the whole query range up by two units. 1 state becomes unfiltered and returns to the original green shade. 15 new states become filtered and turn gray. iSonic first plays the newly filtered states (illustrated in orange) then the unfiltered states (illustrated in a blue circular line).	103
Figure 4-24: Results for the same filtering sequence applied to the table view: (a) no filtering; (b) increase the lower bound by two units. The top 17 states are filtered out. A sweep of the 17 states (illustrated in arrow lines) is played; (c) decrease the upper bound by two units. The bottom state becomes filtered out and played; (d) move the whole query range up by two units. The bottom state becomes unfiltered. 15 new states become filtered. iSonic first plays the 15 newly filtered states then the unfiltered bottom state.	105

Figure 4-25: Filter by a variable different from the one displayed on the map.	106
Figure 4-26: The menu system associated with the map view. The menu can be accessed by a keyboard and displayed both visually and auditorily.....	108
Figure 4-27: The menu system associated with the table view.	109
Figure 4-28: The menu system associated with the filter control window.	110
Figure 4-29: iSonic online help contains a set of HTML documents. Users use arrow keys to navigate the hyperlinks. The current hyperlink is highlighted in pink.	112
Figure 4-30: iSonic online help: if the current hyperlink points to a prerecorded wave sound file, pressing ‘enter’ plays the sound file.	113
Figure 4-31: Steps to solve the task “what and where are the top 5 states with the highest employment rate for people with a disability”	117
Figure 5-1: A simplified diagram showing the core classes in iSonic and their interactions.....	124
Figure 5-2: A sample region-by-variable data file for U.S. states (.csv).....	128
Figure 5-3: (a) A sample decoding map of U.S. states and (b) the corresponding information file (.csv) defining valid regions on the map and the top zoom level manual 3x3 recursive map partition; (c) the mapping between the decoding map and the information file is established by encoding IDs in colors. The encoding rule is illustrated in Figure 5-4.....	130
Figure 5-4: The RGB color space and the relationship between IDs and the colors to encode the IDs ($nr = ng = nb = 4$). The colors at the grid vertices are actually used.	131
Figure 5-5: A sample file (.csv) defining a map spatial sweep order (horizontal sweep) for a U.S. state map (also see Figure 4-12 (a))......	132
Figure 5-6: Read and preprocess map shape data. ([*]: see [193] in Bibliography for details).....	132
Figure 5-7: Map digitization: (a) a partial decoding map; (b) the mosaic region grid after	

the decoding map is digitized.	133
Figure 5-8: Raster to vector map conversion: (a) obtain a vector representation of region #3 by walking the mosaic region grid; (b) the vector representation of region #3 (not all coordinates are drawn).	134
Figure 5-9: A shortest traveling path is not necessarily the straightest, nor does it minimize the degree of polygon interleaving. The whole traveling path in (a) is shorter than the whole path (the sum of the two paths) in (b). However, the path in (b) is straight and vertical, while the path in (a) is not.	136
Figure 5-10 : Two competing optimization goals in map spatial sweep order generation. For the same set of regions, the split in (a) does not cause any intrusion between the two subsets but produces travel paths that are less straight and vertical than the ones in (b). However, the split in (b) causes significant X-dimension intrusion between the two subsets.	137
Figure 5-11: Calculating the overlap ratio among two region subsets when the regions are split vertically at the circular orange-colored mass center. All dots are mass centers. Large sky blue dots are for regions in the left subset, and small lemon yellow dots are for regions in the right subset. Light green marks the intruding areas and dark green marks the remaining areas of regions that intrude into the bounding box of the other subset.	138
Figure 5-12: An automatically generated vertical spatial sweep order for a U.S. state map. The green numbers indicate the order of splits in the recursive splitting algorithm (per region intrusion ratio = 0.2). The orange lines indicate the sweep path. The yellow dots are region mass centers.	140
Figure 5-13: An automatically generated vertical spatial sweep order for a Maryland county map. The green numbers indicate the order of splits in the recursive splitting algorithm (per region intrusion ratio = 0.2). The orange lines indicate the sweep path. The yellow dots are region mass centers.	141
Figure 5-14: From region #2, moving upwards will land in region #1, and moving	

rightwards will cross the water to region #3. All dots represent region mass centers. Big orange dots are for regions in the adjacent region list of the source region #2 (marked with a big green dot).....	143
Figure 5-15: From region #1, moving to rightwards will land in region #4. All dots represent region mass centers. Big orange dots are for regions in the adjacent region list of the source region #1 (marked with a big green dot).	144
Figure 7-1: Different scatterplots may produce the same auditory display in the dot scanning relative navigation methods.	172
Figure 7-2: A bivariate scatterplot showing the correlation of “average household income” and “employment rate” for 24 Maryland counties.	174
Figure 7-3: A heat map at the resolution 9x9 grid for the scatterplot in Figure 7-2. The color indicates the number of dots in each grid cell. A darker green indicates a higher value.....	175
Figure 7-4: A gist of a scatterplot heat map and 3x3 exploration: (a) a horizontal sweep of a heat map, with each empty cell contributing a pause of a fixed duration to the gist; (b) divide the same heat map into 3x3 ranges and press 9 to sweep one range.....	176
Figure 7-5: 3x3 zooming: (a) zoom into range 6 of the scatterplot in Figure 7-2; (b) the 9x9 heat map for the partial scatterplot in (a).	177
Figure 7-6: Zoom into a scatterplot heat map cell: (a) zoom into the 3-dot cell of the scatterplot in Figure 7-2; (b) the corresponding 9x9 heat map.	178
Figure 7-7: Unselected dots fade into gray when users choose to view “selected regions only”	179
Figure 7-8: Heat map for the scatterplot in Figure 7-7.	180
Figure 7-9 : Extend the diagram in Figure 5-1 to incorporate the scatterplot view. Much code from the map view can be used in the scatterplot view with minor modification due to the high similarity of their sonification designs.	185

Chapter 1

Introduction

People collect and analyze data. Examples of such data include financial data, business information, government census data, and many other abstract conceptions. While information visualization for abstract data has produced numerous innovative techniques and interfaces that empower people to “use vision to think” [40], they are typically not usable by people with visual impairments.

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. Several interactive map visualization tools have been developed for exploring such data [10][48][107], but none of them are accessible to users with visual impairments.

About 7.4 million people in the European Union (1.92% of the total population) are estimated to have vision impairments [53]. In the US, there are about 1.3 million blind people [122]. Visual disability is one of the important factors contributing to the “digital gap” in accessing information resources provided by the global information revolution. To help bridge such gap and to promote the “universal usability” [3] of information systems, many countries and organizations have taken initiatives to achieve information accessibility for blind users [178], or even made it a legal requirement [155].

For people with visual impairments, sound is an important alternative or

supplemental information channel. A widely used accommodation for users with visual impairments to access digital information is to rely on screen readers, such as JAWS[®] [63], 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 [179]. For dynamic graphs, tabular data presentations are used instead (e.g., [189]), or textual summaries can be automatically generated from the data set (e.g., [166]).

Several problems exist in such text-based speech approaches. First, while a concise textual description is helpful, the data interaction that is a critical part of the data exploration process is lost. Automatic textual summarization techniques require predefined 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, although speech can accurately describe information, presenting information in speech tends to be longer in duration, harder to realize complex information (e.g., [27], p39).

The other direction, commonly referred to as “sonification” [94], is to present the information with non-speech sounds. Sonification has been effectively used for many application purposes (e.g., [120][28][174] [111]), including data presentations (e.g., [59][139]). For example, Ramloll et al. [139] found that using musical pitches to present values in a numerical table helped data comprehension, improved task success rate, reduced task time, and reduced users’ subjective mental workload. Other perceptual channels, especially tactile feedback, are often used together with auditory feedback to convey information, or to facilitate the users to activate the auditory feedback. Examples

of such multimodal systems, not necessary for data analysis, include systems for real world map information comprehension (e.g., TTT [96], BATS [127]), Audio-Haptic Graphs [138][192], and PC-Access [140]. While tactile feedback is important, it heavily relies on the availability of special devices, such as tactile graphics embossers or refreshable tactile displays. Except for using refreshable tactile displays (e.g., [105]), the displays are static and limited amount of information can be presented. These special devices are usually expensive, not widely available, and potentially limit the wide dissemination of the systems. Computer synthesized sound, on the other hand, has become almost a standard resource on many computers. Considering the goal of universal usability [3], it is important to provide effective audio-only solutions, and limit the requirement for special devices.

In exploratory data analysis, users often need to examine the data from various aspects. It is important that users interact with the data, not only just to respond to system events but to be at the center of the interaction and to initiate every interaction loop. Information visualization has produced design guidelines (e.g., [159]) and workspaces (e.g., [149][125]) for users to perform data exploration through unified interaction methods. For example, Shneiderman [159] described a visual information seeking mantra “overview first, zoom, filter, then details-on-demand”. It has been widely adopted in visualization designs and proved to be an effective guideline. Roth and colleagues [125] created Visage and Sage, providing a workspace that supports multiple visualization views. Some basic operations were supported across all the visualization views, such as to filter, copy, roll-up/drill down, and scale. Users could 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 [149] recognized the value of multiple coordinated visualizations and created a Snap-Together visualization interface that enables users to mix and match visualizations and coordinations without programming.

However, data interaction was typically missing from data sonification. Previous research on data sonification focused on data to sound attribute mapping. Some allowed basic user interactions with the data ([31][139]), but the interaction designs were done in an ad hoc way. Perez-Quinones et al. [129] 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 sonification (e.g., [82][152]), but they did not address task-oriented data interaction in the auditory mode without visual aids. They also assumed the data set can naturally be modeled as a physical environment, therefore may not be suitable for abstract data sets. The sonification and interaction methods emphasize spatial immersion effect, not analytical task needs.

I believe that task-oriented interactivity is a key element in auditory interfaces to support data exploration. I also believe that it has tremendous value to investigate whether those commonly used visualization techniques and guidelines can be applied to auditory interface designs and effectively used in the auditory mode. What adaptations are needed to make it applicable? Such investigation is not only valuable to designing effective auditory interfaces for blind users, but will 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 dissertation pioneers such investigation by trying 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 shall 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 shall the actions be designed? How would the choice of input device affect the interaction effectiveness? How shall information be encoded in the auditory feedback during the interaction?

There have been several data sonification toolkits for researchers to experiment with various data to sound attribute mapping designs (e.g., [100] [181]), but tools for blind users to explore data are missing. Additionally, there has been no previous work on sonifying geo-referenced abstract data for the purpose of exploratory data analysis. One of the goals of my research is to design an effective interactive sonification tool for blind users to explore geo-referenced abstract data.

After reviewing related work in Chapter 2, I describe in Chapter 3 an Action-by-Design-Component (ADC) framework to guide auditory interface designs for exploratory data analysis. The ADC framework includes a set of Auditory Information Seeking Actions (AISAs) that are inspired by and contrasted with the visual information seeking mantra. The framework also discusses general design considerations for several Design Components to support AISAs. In Chapter 4, the framework is applied to the interactive sonification of geo-referenced statistical data to guide the systematic exploration and

evaluation of its design space. The evaluation also partially validates the framework, and provides insights into human ability to perceive complex information from interactive sounds. The features of the resulting data exploration tool, iSonic, are described and illustrated. Chapter 5 explains the iSonic implementation. Chapter 6 reports more user evaluation of iSonic, including design option comparisons and the overall effectiveness of iSonic. The results further validate the utilities of the ADC framework. Chapter 7 demonstrates the extendibility and generalizability of the ADC framework by applying it to the interactive sonification of data graphs, including scatterplots and line graphs. Chapter 8 concludes with possible future work and contributions.

Chapter 2

Related Work

After defining abstract data, this chapter reviews related work from fields such as information visualization, non-visual map data exploration, auditory perception, and data sonification. The review identifies important findings from previous work that are valuable to the research in this dissertation. It also reveals many problems that have motivated my work, such as the lack of interactivity in sonification, the lack of general guidance for designing task-oriented user interactions in auditory interfaces, the lack of effective data exploration tools for blind users, and the lack of understanding of the applicability of visualization techniques to the audio mode.

2.1 Abstract Data

While methods in this dissertation may be applicable to all data, the focus is on abstract data. Abstract data, as defined by Card et al. [40], “is information – such as financial data, business information, collections of documents, and abstract conceptions – that, unlike scientific data, does not have any obvious physical space to map to”. Such abstract data can be presented as a Data Table in the form of cases by variables (Table 2-1). Examples of such abstract data include stock price data, merchandise sale records, company organizational chart, employment chain, film collection data, and government census data.

According to Card et al. [40], variables come in three basic types: nominal,

ordinal, or quantitative. A nominal variable N is an unordered set, such as film titles {Goldfinger, Ben Hur, Star Wars}. A nominal variable is usually only = or != to other values, although nominal variables can be sorted (e.g., lexicographically) to create ordinal variables. An ordinal variable O is an ordered set that obeys a < relation, such as film ratings <G, PG, PG-13, R>. A quantitative variable Q is a numeric range, such as film lengths [0, 360] minutes. Arithmetic can be done on quantitative variables.

According to Cohen [45], a quantitative variable can have an interval or ratio scale. An interval scale possesses the interval property which means that a measurement unit at one part of the scale is always the same size as the unit at any other part of the scale (e.g., the inch). A ratio scale possesses the ratio property which means that when the measurement scale tells you that you now have twice as many units of the variable as before, you really do have twice as much of the variable. All ratio scales have the interval property, but some interval scales may not have the ratio property. For example, the Celsius and Fahrenheit temperature scales are interval scales but not ratio scales. The reason is that one cannot say that something that has a temperature of 40 degree is twice as hot as something that has a temperature of 20 degree, because the zero points on the two scales do not represent absolute zero temperature – a total lack of heat.

	Case _i	Case _j	Case _k	...
Variable _x	Value _{ix}	Value _{jx}	Value _{kx}	...
Variable _y	Value _{iy}	Value _{iy}	Value _{ky}	...
...

Table 2-1: A depiction of a case-by-variable Data Table (from [40], p18)

In addition to the three basic types of variables, there are subtypes that represent important properties of the world associated with specialized representative or operational conventions [40]. Examples include strings representing the names of administrative geographical regions (e.g., USA, France, or P.R. China) that imply geographical shapes and sizes, numbers presenting geophysical coordinates, and temporal variables with associated transformations (e.g., collecting days into weeks, months, or years).

While a case-by-variable table is a basic representation format of an abstract data set (Figure 2-1 (a)), graphs are very often used to visualize such data and the relation in the data items. In a line graph (Figure 2-1 (b)), the variables that define the x-axis are usually ordinal or quantitative by themselves (e.g., time stamps). The variable values of each case are plotted as Y-axis coordinates. A line graph is often used to visually show the value change of one (or more) cases over the period defined by the variables. Visually, it is often more effective than a case-by-variable tabular representation, but such declaration will need to be investigated in the auditory mode. In a bivariate scatterplot (Figure 2-1 (c)), each case is plotted on a 2-D space with the coordinates decided by the values of the two axial variables. The values of the two axial variables usually are ordinal or quantitative. In a node-link graph (Figure 2-1 (d)), each case is a node. The links connecting the nodes show the case dependencies defined by a variable. When the cases represent some geographical objects, such as USA administrative states, a choropleth map (Figure 2-1 (e)) can be used to show the geographical distribution pattern of a variable. A choropleth map is a map with predefined regions, each colored to show how a variable differs from one region to another. Usually one variable is displayed

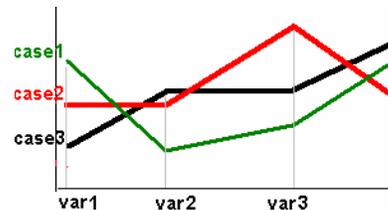
on a choropleth map, i.e., a univariate choropleth map. Sometime, two or more variables can be displayed on one single choropleth map, by using different texture patterns other than colors.

2.2 Abstract Data Visualization

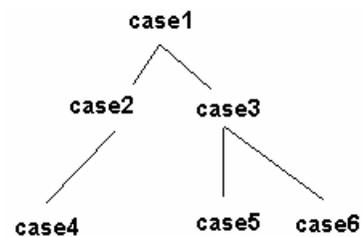
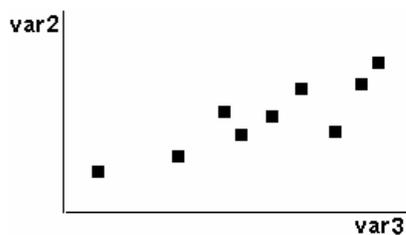
Information visualization is distinguished from scientific visualization by Card et al. [40] as the visualization of abstract data. Researchers of information visualization have long used various graphs, including the ones mentioned above, as powerful data representations to support information discovery. Mackinlay [109] described the automatic design and generation of a variety of graphical presentations of relational information. Card et al. [40] analyzed numerous visualization systems and provided a set of data to visual attribute mapping recommendations for effective visualization design. All these work have provided substantial influences in the field of information visualization.

	case1	case2	case3	...
var1				
var2				
var3				
...				

(a)



(b)



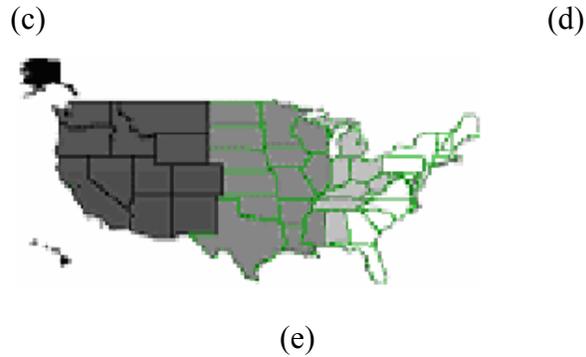


Figure 2-1: Examples of graphical representations of an abstract data set. (a) a case-by-variable table, (b) a line graph containing three cases, (c) a bivariate scatterplot, (d) a graph showing case dependency defined by a variable, and (e) a choropleth map showing the values of one variable for all cases (U.S. states).

Interactivity is a key element in visualization. Several design guidelines and workspaces were already produced that allow users to perform data exploration through unified interaction methods. For example, Shneiderman [159] described a visual information seeking mantra “overview first, zoom, filter, then details-on-demand”. It has been widely adopted in visualization designs and proved to be an effective guideline. Roth and colleagues [149] created Visage and Sage, providing a workspace that supports multiple visualization views. Some basic operations were supported across all the visualization views, such as to filter, copy, roll-up/drill down, and scale. Users could 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 [125] recognized the value of multiple coordinated visualizations and created a Snap-Together visualization interface

that enables users to mix and match visualizations and coordinations without programming.

However, as revealed in the remaining part of this chapter, previous sonification research has focused on data to sound attribute mapping but not interactivity. There is also a lack of understanding of whether common visualization techniques can be applied in the audio mode. Although the data to sound mapping is an essential part of the interface design, the focus of this dissertation is not on how to automatically generate such mapping, but on interactivity design instead.

2.3 Sonification

Kramer et al. [94] defined sonification to be “the use of non-speech sounds to convey information”, or more specially, “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation”. Research into sonification has developed rapidly in recent decades. It brought together interests from various fields, such as sound synthesis (e.g., [50][34][157]), computer music (e.g., [114]), acoustic psychology, and human computer interaction (e.g., [131]). Sonification has been used for many application purposes, both as a supplementary information channel to enhance visual representation and as an alternative in the absence of visual feedback. Examples of application domains include data mining (e.g., [54][79]), exploratory data analysis (e.g., [172]), GUI access/navigation (e.g., [22][65][66][120][28][191]), Web access (e.g., [86]), information presentation and interaction on mobile devices (e.g., [97][180][132][148]), graph and image presentation (e.g., [104][111][47][6][151][89][88]), mathematics access (e.g.,

[167][168]), algorithm presentation and computer program debugging (e.g., [35][175][174]), map exploration (e.g., [85][154]), and collaborative environments (e.g., [136][118]).

2.4 Non-speech Sounds

Non-speech sounds can be categorized into three types: real-life natural sounds (based on naturally occurring, nonmusical events), communications sounds (based on traditional electronic signals), and musical sounds (abstract musical signals). Gaver [65] first described auditory icon, the use of real-life sounds, such as door closing or bird chirping, to present events in a computer. The concept has since been explored by many researchers (e.g., [65][119]). Real-life sounds have the ability to convey complex messages in a single sound, and do not require mapping learning, provided that the sound is easy to identify and is a good conceptual mapping of the event or operation. Yet users reported them to be the least pleasant and least appropriate for use in computer interfaces [161][143]. Additionally, it will not work if there is no everyday equivalent to the event or operation. It can also be misleading if the users' interpretation of the sound is different from the designers [37]. Sikora and colleagues [161] conducted a series of experiments that asked subjects to map a set of interface functions to a set of sounds consisting of natural sounds, communication sounds, and musical sounds. The experiments also asked the subjects to rate the sound pleasantness, the mapping appropriateness, and their preferences of which sound to use for a given function. The results showed that although real world sounds mapped most predictably to the functions, they were never chosen as the preferred sounds. Subjects consistently preferred musical sounds to the real world

sounds, giving it the highest rating of pleasantness and mapping appropriateness.

Communications sounds varied in subjective ratings and occasionally was chosen to be the best selection for a given interface function.

In recent years, researchers have been exploring using the highly structured nature of musical sounds to convey information. Blattner and colleagues [22] first developed earcons, the use of abstract, synthetic tones in structured combinations to create auditory messages that are used in the computer/user interface to provide information to the user about some computer object, operation or interaction. As icons are visually recognizable computer objects, earcons also target at individual recognizable entities. Brewster and colleagues further studied the structures of earcons and extensively used earcons in various applications, such as to enhance GUI widgets ([27][30]), menu hierarchy navigation ([28][29]), mathematics access ([167]), information presentation or navigation on small devices ([180][97]). Although earcons do not possess an intuitive mapping as auditory icons do, therefore have to be learnt, they also do not have many disadvantages that auditory icons have. Users found earcons in general more appropriate for applications [143].

A musical sound is made up from three basic components: pitch, timbre, and loudness. Pitch is related to the frequency of the tone, and decides the order of sounds on a musical scale [8]. Loudness is the perceived intensity of a sound, and is often referred to as the sound volume. Timbre is the “quality” of a sound, through which a user can distinguish two sounds similarly presented and having the same loudness and pitch [7]. In another word, Timbre is that which allows a listener to distinguish between a piano and a violin playing the same note.

Sounds arising from real world sources carry spatial location information. Sounds available from a computer traditionally are mono or stereo panned. Recently, researchers have been able to simulate such spatial localization cues for computer-synthesized sounds. Currently there are two ways to generate sounds that have spatial locations similar to sounds arising from real world sources. Some systems vary the signals presented from two (or more) loudspeakers to simulate sources from different locations (e.g., [64][158]). Other systems are headphone-based in which sounds are processed using Head Related Transfer Function (HRTF) to simulate the normal auditory localization cues (e.g., [188][41][202]). The later approach is often preferred because it allows precise stimulus control and retains good simulation effect regardless of listener movements. However, in order to achieve good source location perception, particularly the source elevation, individualized HRTFs need to be used ([187][24]). In practice, individualized HRTFs can be directly measured using moving speaker and in-the-ear microphone. It is the most accurate but a very time-consuming process that requires precise measurement and careful calibration [41]. A number of studies have sought to perform simple HRTF personalization. For example, Middlebrooks [112] used frequency scaling of non-individualized HRTFs to improve localization. Zotkin et al. [203][204] showed that localization accuracy improved from matching the listener's certain anthropometric ear parameters with the HRTF database, and incorporation of a low-frequency "head-and-torso" model.

2.5 Non-visual Map Access

Current assistive technology research is exploring a number of techniques to help blind

users benefit from the spatial awareness provided by maps. For navigation in the real world, GPS-based talking maps have been developed [67][68]. One example product is a talking U.S. nationwide digital map consisting of most addresses and street intersections [156]. Users can navigate the map using the arrow keys and listen to speech-synthesized descriptions of the map and directions. Trekker [177] is another GPS-based application that helps the blind to navigate. Maps are also important to learn an area beforehand and chose a route. Schneider and Strothotte [154] have designed a tangible interface using physical building blocks that users manipulate to promote constructive exploration of the map.

When using maps to learn geo-referenced data, it is a standard accessibility practice to provide a 'd link' mark for image descriptions that visually impaired users are able to search for. OptiMaps [46] employs such a mark below each choropleth map to provide a hyperlink to a textual version of the data, generated automatically with the dynamic map. Examples are available in AtlasPlus [121]. New vector graphic file formats, such as SVG, permit embedding text descriptions with the graphic information and should simplify this process as a result. Unfortunately reading data values still does not give an adequate feeling for the spatial relationships between areas that only spatial techniques can provide. Tactile and haptic techniques (and strategies for signifying data haptically) have been devised [70].

The Opticon reading machine for the blind invented in 1970s [1] was a versatile device not only for haptic reading of text prints, but could also be used for reading image including maps. As users slide the camera on the paper, signals are sent from the camera to 144 tiny pins under the finger. These pins then raise and form a tactile representation

of the symbol under the camera. The darker the symbol is, the higher the pins are. The drawbacks include that the device was very bulky and expensive. Effective use of such device needs years of training. The device was widely used in 1980s but today has been largely replaced by OCR (Optical Character Recognizer) and other electronic reading aids.



Figure 2-2 : The VTPlayer tactile mouse is equipped with two 4x4 mechanical matrix units that can be used to deliver a tactile sense of the screen. For example, as users explore a U.S. map, the tactile patterns change over different states.

VirTouch's VTPlayer [176] provides an example of map exploration using a tactile mouse. The mouse is equipped with two 4x4 mechanical matrix units, which can be used to deliver a tactile sense of the screen (using Braille and other alternatives), as shown in Figure 2-2. As users explore the map, the tactile patterns change over different states or areas. One of the limitations of the mouse is that it is an input device providing only relative positioning so absolute position on the map has to be provided by audio feedback.

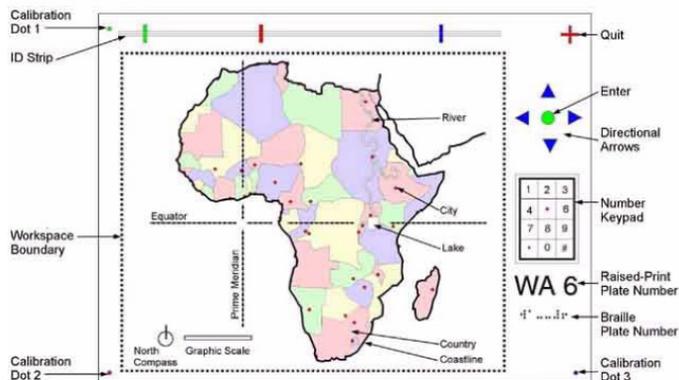


Figure 2-3 : An example of tactile maps augmented with speech description when mounted on a touch screen (Touch Graphics, undated): (a) sample map plate, and (b) the sample map plate in use with the talking tactile tablet (TTT).

Another example developed by TouchGraphics and their research colleagues [96] augments standard printed tactile maps with speech, but does not use sonification. Tactile maps are secured on top of a touch screen that provides the location of each touch (Figure 2-3). The TouchGraphics Atlas has five operational modes. Users can simply explore by touching a physical tactile map and hear names spoken of the places touched. They can also select a destination from an index, and listen to directions that are updated as a user's finger gets closer to the destination. The distance between two points can be calculated, and descriptions of the areas can be listened to. This work was inspired by Nomad [128]. 3D maps can benefit from the use of haptic devices. For example, the Phantom device provides 6 degrees of freedom input and 3 degrees of freedom output to explore the virtual sound space and augment it with haptic feedback providing the same sensation as moving a single finger over a physical 3D map. Tactile maps can be augmented with

abstract audio output [57][95]. For example, users can hear a series of graduated pitches proportional to elevations above sea level as they explore the map. Jeong [87] compared the effectiveness of and preferences for using auditory feedback (volume of sound), haptic feedback (extent of vibration) or both in the tasks of identifying the highest or the middle valued state on a partial static choropleth map of the US. The experiment showed that overall performance is most successful when using haptic feedback alone but users preferred having both haptic and audio feedback. This result may be attributed to the fact that haptic devices provided spatial cues while the standard sound output did not.

Unfortunately tactile and haptic techniques require special hardware, from embossed maps to specialized input/output devices, and hardware is usually a limiting factor for wide dissemination, as users may not have access to the technology or the financial means to purchase it. It is of significant practical importance to invent techniques that rely solely on audio. Such techniques have the potential to be used by a larger population as blind users often already rely on headphones when using screen readers.

The BATS project [127] uses simple spatial audio (mostly stereo effects) to sonify maps as shown in Figure 2-4. Real world auditory icons are played as the user moves the cursor over the map. For example, car noises are heard over cities and birds and crashing waves are heard over forests and beaches. The sound becomes louder as the user gets close to the source of the sound. The BATS interface helps blind users explore the map to locate map objects, but does not target at assisting abstract data comprehension, such as to discover the population density distribution on the map.

North Carolina

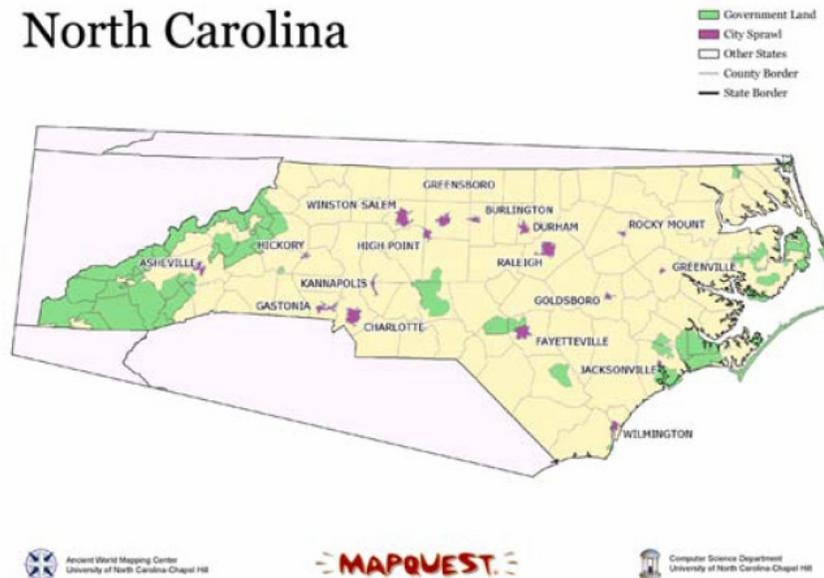


Figure 2-4 : The BATS sonic map helps those without sight to explore the map and locate map objects [127]. Real world audio icons (car noises, wave crashing, etc.) are played as the user navigates the map and appear clearly as coming from the left or the right. Volume increases as users get closer to the source of the sound.

2.6 Auditory Perception

A rich history of research has provided valuable insights into the perceptual and cognitive aspects of general auditory perception. Examples regarding speech and relatively simple auditory tones include intensity (loudness), frequency (pitch), and temporal discrimination of static sounds ([75][113][115][116]), the determinants of pitch and loudness [72], and auditory localization abilities [24]. Investigations have also been done into the perception of more complex, dynamic auditory patterns in speech and music

([26][74][106][182]). The perceptual interferences among sound properties and cross modalities have also been studied, such as the interferences between pitch and localization [147], pitch and loudness [123], and visual and auditory perceptions [135][171]. Much of the work is very relevant for designing data sonification.

Much research has shown that visually impaired people have better hearing and sound localization ability than normal sighted people. For example, visually-impaired people showed greater locomotion ability than blindfolded sighted people, using natural auditory cues for guidance [169]. They also displayed better sound localization ability than blindfolded sighted people [98][145]. While early visual experience benefits spatial hearing ability of people who became blind later on [1], congenitally or early-blind people can replace vision with audiomotor feedback to calibrate their auditory space [14][99]. Studies have also shown evidence that early vision deprivation leads to cross-modal cerebral reorganization, enabling totally blind individuals to have better auditory ability than sighted people, in order to compensate for the loss of vision [145][11].

There has been consistent evidence that congenitally totally blind people were able to construct and operate mental spatial imagery and pictorial imagery, but performed poorer than sighted people in demanding spatial tasks, such as mental rotation and scanning (e.g., [77][173][4]). Explanations in the literature of such deficit take two sides: representation explanation and processing explanation. Representation explanations emphasize that congenitally blind people and sighted people use different spatial representation formats (e.g., [201][9]). For example, the most extreme version of the explanation account proposed that the blind use “abstract semantic representations” [201] whereas the sighted use analogue visuo-spatial representations. Heller [77] suggested that

blind people's visuo-spatial representations differ from sighted people's because the blind lack experience of external frame or reference cues (e.g., knowledge of the horizontal and vertical) that aid the construction of a more manipulable object-centered representation [146]. Processing explanations argue that the performance difference is due to processing deficiencies in the stages of constructing, integrating, and manipulating, since the serial processing of sound and touch is slower and more prone to errors than the parallel processing of vision (e.g., [4]). Evidences have been presented suggesting that most blind visuo-spatial deficits are the result of processing differences between the sighted and the blind. For example, Vecchi [173] found that the blind were equally able to remember spatial patterns when they were engaged in a task that did not require manipulation or elaboration through active processing.

2.7 Data Sonification

2.7.1 Scientific Data Sonification

As scientific visualization has been used by Card et al. [40] to differentiate visualization of scientific data from abstract data, I use scientific data sonification to refer to sonification of scientific data. There have been numerous case studies of using sounds to present scientific data, for revealing interesting patterns and helping understand underlying structures. For example, Blattner et al. [23] sonified turbulence information. Axen and Choi [16][17] described using audio to present structures of complex and large data sets such as flow patterns, turbulence of fluids, and complex topological representations. Fitch and Kramer [58] sonified body electric. Lodha et al. [101] used

music melodies to assess protein structural alignments.

2.7.2 Abstract Data Sonification

There has also been research using sonification to present abstract data, both for monitoring purposes and for analytical purposes. Among the numerous examples are some pioneer work by Kramer and others in the early 90's on sonifying high-dimensional data ([90][93][92]). Previous abstract data sonification work largely focused on data to sound mapping design and data comprehension through passive listening. Some have limited support for user interaction but interactions were designed without systematically considering task needs. Some interactions used control devices, such as computer mice, that are not appropriate for blind users.

Smith et al. [164] described a method to sonically probe an iconographic display of high dimensional data. The data is depicted as small icons on a 2-D scatterplot representing two dimensions. The remaining dimension values are visually represented as the shapes and colors of the icons, and decide the properties of sounds produced by the icons when the icons are swept by the user's mouse movements. The iconographic pattern and the speed and direction of the movement determine how many notes and different timbres play simultaneously, thus producing various "auditory textures" of the data display.

Hermann et al. [79] showed that principal curve sonification can facilitate the detection of the "main" structures in high-dimensional data. Principal Curve (PC) of a data set is a trajectory that passes through the "middle" of the data. While moving along the principal curve in the high-dimensional virtual space, the listener hears a changing

auditory scene presenting the PC properties, the data points, and the locally averaged data density. To be more specific, the PC properties are presented in an audio stream that resembles the sound of a moving vehicle. Large curvatures result in a higher pitched sound, while the volume of the sound corresponds to the velocity of the listener. Each data point emits a tick sound whenever the listener position passes its projecting point onto the PC. The tick's loudness, pitch, and localization are determined by the relative position of the data point to the listener. The locally averaged density of the data is presented by a time-variant oscillator whose frequency is determined by the density. The amplitude of the oscillator is modulated by the local average distance of the data.

Some researchers have examined the ability of human with normal sight to perceive facts in abstract data set by listening to sonified graphs, such as line graphs, scatterplots, and diagrams. Much of the work focused on passive auditory perception of the graphs, but did not examine the role of task-oriented user interactions. For example, Flowers and colleagues examined univariate line graphs [59] and bivariate scatterplots [60], showing that most people can understand trends, clustering, outliers, correlations, and other simple statistical features of a data set just as well by listening to a quick sonified overview of the graph as they could by reading the visual graph. Bonebright et al. [25] further studied sonified graph comprehension and found that most people can understand multivariate scatterplots except for those with widely dispersed data points and without easily discernable “envelopes”. Franklin and Roberts [61] evaluated five sonification designs of pie charts. Brown and Brewster [32] found that people can possibly interpret line graphs containing two data series. From data sonification practices and user evaluations, some guidelines have also been extracted and summarized, such as

the guidelines for sonifying line graphs and tables [33]. Much of the above work could have benefited from a rich set of user interactions, as discussed in Chapter 7.

For general data to sound mapping, various experimental assessments have been done to find effective or natural mappings from data attributes to sound attributes. For example, Kramer [91] summarized some organization principles for representing data with sounds. Walker and Kramer conducted a series of experiments on the general psychophysical scaling, polarity and magnitude estimation [183][184][185].

Sound is time-sensitive stimuli. Human auditory perception is far less synoptic than visual perception. It limits the number of individual data items that can be presented in parallel to the user. As a result, multiple data items need to be presented in a sequenced manner. For data that contains the time dimension (such as a time-series data collection), a sequencing order is already naturally implied. In fact, many sonification projects have made use of that implicit order. For data that does not imply any natural data to time dimension mapping, an appropriate data dimension needs to be picked for sequencing. Saue [152] defined temporization as the process of mapping from a non-temporal to a temporal domain. Saue's definition of temporization implies both sequencing during automatic sonification of the data items (automatic sequencing), and sequencing according to the order of the user's movement among the data items (manual sequencing).

The need for automatic sequencing was already recognized by researchers in designing auditory information presentations. For example, Alty and Rigas [6] used various scanning orders (top down, central scan, and ascending by object size) to present the layout of a set of simple 2-D graphical objects. Franklin and Roberts [62] proposed a

tour based model to unify various methods of sequencings that exist in previous research, and to create new sequencing designs in a 2-D or 3-D presentation space. In their model, automatic sequencing is predefined by using a selectable grid to define paths with spans, span angles and span envelopes. The path describes the order in which data is to be automatically played. The span, span angles and span envelopes associated with each path point in the path define its neighborhood. All data items in the neighborhood are played in parallel. The definition is done by selecting and connecting points on a selectable grid. Predefined sequencing can be loaded by the user to automatically scan the 2-D or 3-D presentations of general data sets. When the resolution of the data and those of the path points are different, the data presentation space needs to be registered to the path points by, i.e., scaling, sampling or averaging. Some sample paths were described, such as a scan line path, a raster scan path, a zigzag path, a spiral path, and a water droplet path.

Barrass [18] recognized the importance of tasks in sonification design, and proposed a TaDa method for auditory information design that is “Task-oriented and Data-sensitive”. The method includes design process that integrates task analysis through scenario description and requirement analysis, a database of sound examples, a rule-based mapping design aid, and interactive sound design tools. But the resulting sonification does not support user interaction with the information.

Various data sonification toolkits have been developed to help researchers experiment with different sonification designs. Examples of toolkits that accept general data sets in tabular forms include Clarity Sonification Toolkit ([93], Listen [190], MUSE [100], SonART [21], Sandbox [181], and Interactive Sonification Toolkit [130]. Ramloll

and Brewster also described AudioCave, an environment for studying sonified graphs [137]. However, no tool is yet available to support users in exploratory data analysis, especially for users with visual impairments.

2.8 Interacting with Data Sonification

While there have been much research on data sonification design, the support for user interaction with the data set, i.e., moving around in the data set and examining various parts, is missing from most systems and toolkits.

In many systems and toolkits (as in the previous section), the process of sonification is done through parameter mapping [153], which involves taking data attributes and converting them into sounds. The user is allowed to control this mapping process (e.g., to change the scaling or mapping). After the user scales the data and specifies the parameter mapping, the entire data set is converted into a sound file, which is then listened to non-interactively, rather like a CD. The emphasis of those systems and toolkits is on creating sonification for experimenting with various sonification mappings. Although in some of these tools, the user can dynamically change the mapping during the playback of the sonification, such as adjusting the panning of individual channels, as in [130], the interaction is done to the mapping function but not with the data. While such systems and toolkits may satisfy the needs for experimenting with different sonification mappings or monitoring tasks, they are not satisfactory for the purpose of exploratory data analysis.

Some researchers have recognized the importance of interaction with data in sonification, and provided some supports for the users to move around in the auditory

scene. For example, Smith et al. [164] allows user-controlled mouse movements to “sonic probe” the iconographic display (a 2-D scatterplot) of high dimensional data. Sounds associated with the icons (data items) on the iconographic display are activated when the icons are swept by the mouse movement. Similarly, Fernstrom and Brazil [55] described a sonic browser to view a collection of multimedia data files. They used an “active cursor” to move around in the starfield presentation of the data files (sonic objects). The objects closest to the cursor are activated and the associated sound files start to play. In the work by Ramloll et al. [139] to use non-speech sounds to improve visually impaired users’ access to 2-D tabular numerical information, users are allowed to use a keyboard to move up, down, left, and right among the 2-D table cells. In the sonification of line graphs containing one or two data series, the user can use a numerical keypad to move along the X-axis to activate a note presenting the Y-value of each individual data point visited [31]. In the Interactive Sonification Toolkit [130], multiple data streams (in the form of a matrix where each column is a data series evolving over time) can be loaded. The data is presented in a rectangle interaction area (with time portrayed from left to right). The user can click and drag the mouse along the X-axis over the area to instantly listen to the sound of each data array visited (all data along the Y-axis belongs to an array). However, many of these user interaction designs have not systematically considered task needs, or investigated whether the design can be effectively used without visual feedback, thus may not be suitable for users with visual impairments.

In order to unify the design of interaction with data sonification, some models have been proposed. But they assume the data set has natural physical presentations (like in scientific visualization), and thus may not be appropriate for abstract data. Additional,

some models emphasize using special devices to issue the interaction, and none of the models take into account users' task needs, or the suitability for users with visual impairments.

For example, Hermann and Ritter [80] proposed a Model Based Sonification (MBS) framework for data analysis. In a sonification model in MBS, the data is not used by the user to map to sound attributes but is used to parameterize a virtual sounding object, i.e., to instantiate a virtual object with specific surface material or 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. [82] emphasized that interaction with sonification should be as close to real-world acoustic interaction as possible, in order to promote the 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 [78], or other tangible interfaces, rather than just using standard computer input devices, such as mice or keyboards. Based on MBS, they developed several models for different data sets, such as the Data-Solid Sonification Model for high-dimensional data [31].

Saue [152] presented SoundWalker, a model for sonification of large spatial data sets based on ideas from ecological acoustics. The data set is modeled as spatially distributed objects in a scene that mimicks a real world physical scenario. Each object can be associated with several variables. The objects are sonified as parameterized auditory icons carrying information about the object, such as the size, and attribute

values. The user “walks” through the scene by moving an active listener (controlled by mice or similar devices), listens to the sound produced by the nearby data items rhythmically (ObjectSounds). Sound spatialization is used to present the position and distance of these objects relative to the user’s listening position in the scene. Besides the surrounding objects, the user will also hear the sound confirming the walking (MoveSounds), and at the distance the sound of objects in a more global scope (AmbientSounds). The goal of SoundWalker is to “virtually bringing the data set back to its physical origin, e.g., to use a subterranean cave model to present a seismic data set, a human body model to present a medical data set”. The model mentioned analysis tasks but did not show how tasks can actually be performed by walking in the model.

Similarly, user controlled activation of object sounds in virtual environments have been suggested, such as sonic probes by mice [164], virtual microphones [71], control paths [43], auditory cursor and radar interactor [144]. None of these have considered the needs of data interaction without visual aids.

2.9 Discussion

Review of past research on data sonification has revealed that the following aspects have not been adequately addressed regarding abstract data sonification for exploratory data analysis.

Previous abstract data sonification applications typically lack support for users to interact with the data. Some have limited support but the interactions have been designed without systematically considering task needs.

Existing frameworks or models to unify interactive data sonification assume the data

set can naturally be modeled as a physical environment, therefore may not be suitable for abstract data sets. Their sonification and interaction methods emphasize spatial immersion effect, not analytical task needs.

Many applications and frameworks use computer mice as the interaction control device. Some emphasize the use of special devices that are hard to obtain. The existing work rarely considered interaction needs without any visual aid, such as a traditional computer mouse is not suitable for users with visual impairments.

There have been several sonification tools developed for researchers to experiment with various sonification mapping, but no tool is yet available to support users in exploratory data analysis, especially for users with visual impairments.

While sonification has been used to assist map exploration and locating objects on the map, there has been no previous work on sonifying geo-referenced abstract data for the purpose of exploratory data analysis.

Chapter 3

An Action-by-Design-Component Framework

Data interaction has been extensively investigated in visualization systems. But previous data sonification research typically lacks support for task-oriented data interactions. I believe it is important to investigate whether techniques in visualizations can be translated for use in auditory data exploration without visual aids, and to understand what design implications are involved. In this section, I describe a framework to help guiding auditory interface designs for exploratory data analysis. The framework includes a set of Auditory Information Seeking Actions (AISA) that are contrasted with visual actions, and a set of Design Components with general design considerations for supporting AISAs. The validity, benefit and limitation of the framework will be examined in later chapters.

3.1 Auditory Information Seeking Actions (AISAs)

Shneiderman [159] described a Visual Information Seeking Mantra “overview first, zoom and filter, then details-on-demand” that has been widely used as a principle in information visualization designs. It is also a common technique to use multiple coordinated visualizations in the visual mode. Inspired by the work in visualization, I propose the following set of Auditory Information Seeking Actions (AISA) for accomplishing exploratory data analysis tasks in the auditory mode. Many of the actions resemble those in the Visual Information Seeking Mantra but involve different cognitive processes and present special design challenges due to the highly transient nature of

sounds. Other researchers can further extend this list with more actions.

3.1.1 Gist

Obtaining a gist is to experience the overall data trend or patterns 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 Section 3.2.4) because human auditory perception is much less synoptic than visual perception. The term gist is used instead of auditory glance (as used by some researchers [167]) to differentiate from “seeing” by eyes.

3.1.2 Navigate, Situate, and Search

Navigation refers to the user moving around to examine portions of the data set by listening to a sub-gist of that portion. Navigation is not just to play, stop, resume or rewind along the time dimension of the data set’s sound scape. It needs to follow paths that are natural to the data relations. When navigate is done in a fast tempo along the sequencing path of the gist, the effect of the sound scape produced by navigation is similar to the gist.

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 the 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.

3.1.3 Filter

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 need to be given after filtering is done instead of continuous display updates during the filtering process.

3.1.4 Details-on-demand and Adjust Information Level

Users can select an item or group to get details. While sonification emphasizes the use of nonspeech sound, speech is an effective presentation at the details-on-demand level.

It is hard to understand a data element without the appropriate context. On the other hand, too much detail slows down the sequential presentation and can be overwhelming. Multiple information detail levels that incorporate different combinations of speech and non-speech sounds are needed.

3.1.5 Select

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.

3.1.6 Brush

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, have them plotted on a map to view their geographical locations.

3.2 Design Components

In visual interfaces based on graphical displays, data is often presented as visual properties of objects in the virtual environment. Interaction with the data is typically done by using 2-D pointing devices to directly manipulate the visual objects, and possibly changing the visual appearance of the virtual environment as the result, that is displayed through the 2-D computer screen. It is enabled by the continuous, sustained feedback empowered by today's graphics display techniques and by a human's highly synoptic visual perceptual ability.

However, this typical process in visual interfaces cannot be directly applied to the auditory mode, because sounds are perceived by human as transient/evanescent, time-sensitive, sequential stimuli. As a result, users must continuously construct and maintain some mental representations of the data space in order to interact with the data through

the auditory interface.

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 (Figure 3-1).

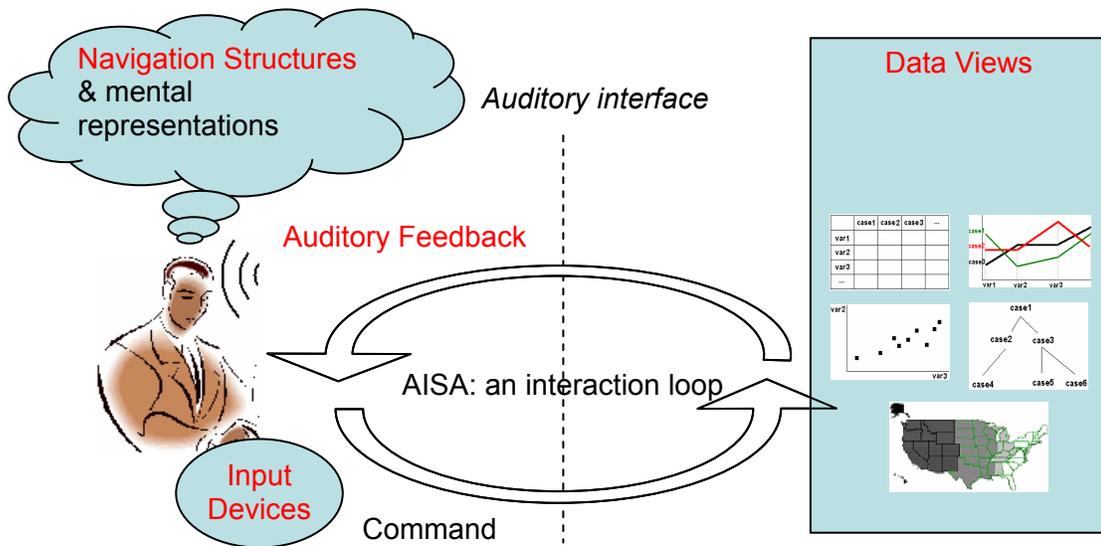


Figure 3-1: Design components (in red) involved in data interactions in an auditory interface. An Auditory Information Seeking Action (AISA) is an interaction loop that involves using an input device to issue a command to the system and listening to the auditory feedback.

3.2.1 Data View

A data view is a form of presenting the data items and their relations. Examples include a

table, a map, a line graph, a node-link graph, and a scatterplot (Figure 3-1). The choice of data views should be based on the structure of the data set, and to display the data relation that is most useful for the current task needs. In visual interfaces, an external representations that “fits” the task can provide a memory aid and allows the task to be performed more easily [199]. In auditory interfaces, Flowers et al. [60] suggested that to display the correlation relation of bivariate data, a sonified bivariate scatterplot presentation is more effective than a line graph sonified as two pitch-coded sound streams (one for each variable) in different spatial channels. On the other hand, the line graph can be useful to convey similarity in functional relationship between the two variables and time, such as phase relationship in periodic. There is also other evidence [20][194] that choosing the right data view for a given task dramatically influences performance in auditory interfaces. While visually impaired users traditionally access data in a linear manner, researchers have shown that they can construct a mental representation of the spatial layout of the information and benefit from the spatial layout (e.g., [150]).

3.2.2 Navigation Structure

A data view provides navigation structures that allow the users to efficiently move around in the data set, searching for a data item or a group of items that they are interested in. In some previous data sonification work, users used a mouse or other input devices to move around in the 2-D or 3-D data space using a “torch metaphor” [49]. The data items falling into the neighborhood of the current listener position produce sounds whose acoustic properties depend on the value of the data item and the data item’s

distance from the current listener position. Such a navigation method could be useful for some data views, e.g., a scatterplot, but may be inefficient for others, e.g. a node-link diagram.

Different data views have different navigation structures that are natural to them. For example, in a case-by-variable table, users can do cell-by-cell navigation, i.e., moving from one table cell to an adjacent cell. In other words, users can simply move left/right to get the value of each case for the same variable, or move up/down to get the value of each variable for the same case. Users can also use column-by-column navigation or row-by-row navigation. In a column-by-column navigation, users can move from one column to another, and listen to the target column's sub-gist, produced by all the values in that column played automatically. In a line graph, users can do variable-by-variable navigation, i.e., moving from one variable to another along the X-axis, to listen to the Y-axis values of all the cases. Users can also do case-by-case navigation, i.e., moving from one case to another, to listen to all the variable values for the current case. In a scatterplot, users can jump from one case to the nearest case in a particular moving direction (up, down, left, right, diagonally), move a sweep line along the X-axis to activate all cases encountered, recursively partition the 2-D space and listen to the cases in each partition as a group, or move the cursor freely in the 2D space to activate the data items falling into the neighborhood of the current listener position. In a node-link graph, it is natural to follow the node connection links. In a choropleth map, users can jump from one geographical region to its adjacent ones, lay an even-size grid to generate a mosaic version of the map and move among adjacent grid cells, recursively partition the 2-D space and listen to regions in each partition as a group, or move freely in the 2-D

space and activate the geographical region falling into the neighborhood of the current listener position.

The choices of navigation structures should reflect the data relations in the data view and match the current task needs. The importance has already been demonstrated in previous work about diagram understanding [20]. The work showed that hierarchically presented information facilitated hierarchical tasks, but if the tasks were navigational the information was best presented with an emphasis on connections. In the next few chapters, many of the navigation structures described above are illustrated, examined and compared.

3.2.3 Input Device

The choice of input device needs to consider both effectiveness and universal availability.

For universal usability, input devices that are easily available are preferable. This reduces visually impaired users' dependence on special devices and allows them to perform tasks in a variety of situations without the need to carry around a special piece of bulky equipment. General purpose input devices that are inexpensive are preferable than those designed for special purposes that are often more expensive and not always available. For example, a keyboard is a part of the standard hardware configuration of a computer. For success by users with visual impairments, a system should provide interactions optimized for keyboard-only operations.

For effectiveness, in auditory mode without sustained feedback, one important effectiveness measure should be how well the device helps users obtain their sense of orientation and position in the navigation structure, and how well and fast it allows them

to accurately move to the next desired target. There have been studies showing that the choices of input devices affect users' spatial task performance in virtual environments ([13][163]. Physical input devices are often better than speech input devices for tasks involving spatial representations, because in both real and virtual worlds, kinesthetic feedback is used together with sensory (e.g., visual) information to construct a mental spatial representation of the information (e.g. [44][162]. For visually impaired users, choosing input devices that can maximize the kinesthetic feedback is often very important because the lack of visual feedback requires them to rely even more on other sensory channels to keep oriented.

The mouse barrier refers to the problem that visually impaired users cannot effectively handle the mouse, a 2-D relative pointing device at the default configuration. Comparatively, absolute positioning input devices allow users to move their finger around while at the same time keeping track of their position to a certain extent through the kinesthetic feedback. Zheng et al. [200] showed evidence that using absolute mode devices increased users' ability to develop an accurate mental spatial representation of a virtual environment, compared to relative mode devices. Researchers also pointed out the advantage of using absolute pointing devices over relative pointing devices for blind users (e.g., [133][134]), although some research [88] showed that the absolute positions of the tablet, without any other feedback, is not sufficient to precisely determine the location. To be more specific, absolute positioning input devices (such as a touch-sensitive screen), together with a 3-D immersive virtual sound environment or a virtual sonic grid, allow blind users to construct a mental representation of the spatial layout of abstract objects, enforce their orientation and mobility skills (e.g.,

[89][150][103][102][88]). However, such auditory location feedback needs to be designed carefully because using sound to present the movement may reduce the auditory bandwidth left for sonifying other information.

The importance of matching input device properties to manipulation structures in visual interfaces has been demonstrated by Jacob and Sibert [84] in their experiment. It showed that integral tasks (changing object position and size) were performed better with an integral device (Polhemus) and separable tasks (changing object position and grayscale) were performed better with a separable device (mouse). In auditory interfaces, input devices should be chosen to match the navigation structure of the current data view and task needs.

While research has shown absolute positioning devices are preferable over relative positioning devices for visually impaired users, absolute positioning may not always outperform relative positioning, especially when none of the data attributes or data relations have been mapped as the exact item locations on the 2-D space. Similarly, continuous movement devices are generally preferable for 2-D GUI interaction. But discrete movement devices could be more suitable for some navigation purposes. For example, to explore the spatial layout of a map, an input device that supports continuous movements and absolute positioning on the map (e.g., a touchpad calibrated to the full map range) may be more suitable than a standard keyboard which typically provides discrete and relative movements on a 2-D space through the use of arrow keys. On the other hand, to move among rows and columns in a table, a device that provides discrete movements and relative positioning (e.g., keyboards) may be satisfactory. In fact, such devices may be better than a continuous and absolute positioning device because it allows

the user to precisely land on a row, a column, or a cell, on each keystroke.

Each input device has been designed to naturally support some types of movements and positioning. Card et al. [40] categorized physical input devices by their physical manipulation properties, such as whether they are linear or rotary, absolute or relative, and by position or by force. Table 3-1 lists some input devices and categorizes them by their default movement and positioning type.

	Absolute	Relative
Continuous	Touch screen, Touchpad	Mouse, Touchpad, Trackball
Discrete	Keyboard	Keyboard (arrow keys)

Table 3-1: Input device categorization and examples

Of course, the categorization is not definite. By changing a device's physical-property-to-input-space transformation function, the device can be remapped to support movements and positioning other than what they have been originally designed for. For example, by designating each key to a predefined area in a 2-D display space, a keyboard can be transformed into a low resolution discrete absolute pointing device (Figure 3-2 (a)). Another method to transform a keyboard to an absolute pointing device is to recursively divide the 2-D display space to 3x3 ranges and to map them to the 3x3 layout of the nine number keys on the number pad (Figure 3-2 (b)). On the other hand, a touch screen can be used to perform relative and discrete movements if areas are designated to activate predefined key actions (Figure 3-3). Such method was in fact used in the product of Touch Graphics [96].

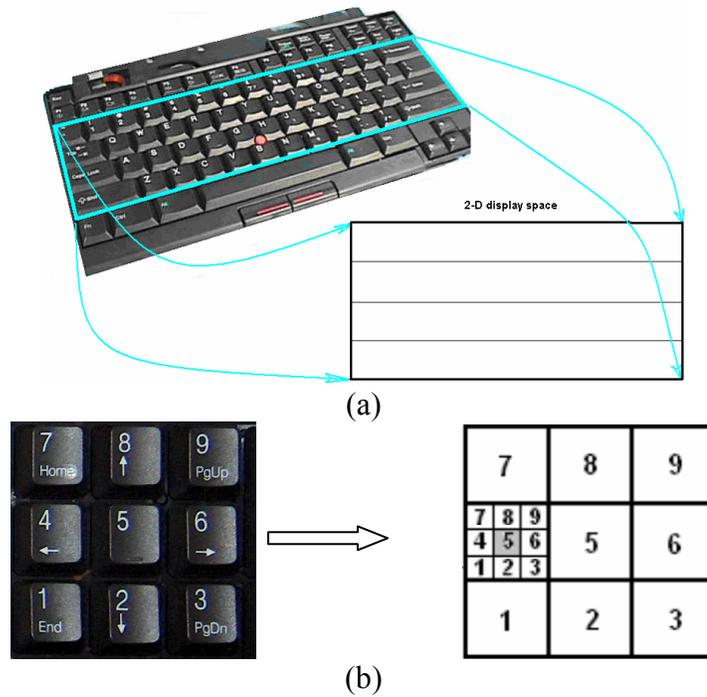


Figure 3-2: Device transformation: (a) an IBM Thinkpad laptop keyboard is remapped to a 2-D display space and becomes a low resolution discrete absolute pointing device; (b) another method to transform a keyboard to an absolute pointing device is to recursively divide the 2-D space to 3x3 ranges and to map them to the 3x3 layout of the nine number keys on the number pad.

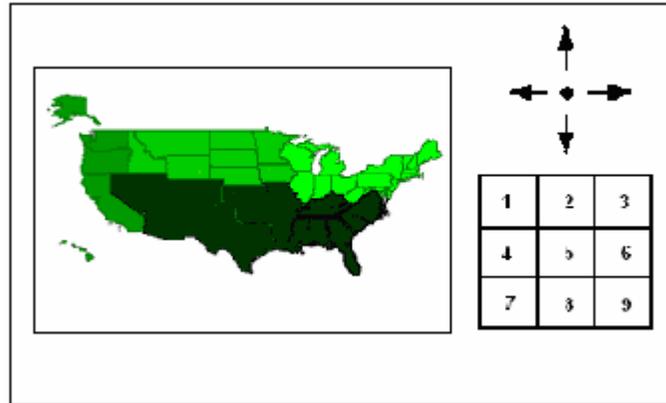


Figure 3-3: Device transformation: a touch screen can be used to perform relative and discrete movements if areas are designated to activate predefined key actions, e.g., the number keys, and up, down, left, and right arrow keys.

3.2.4 Auditory Feedback

Data and data relation are represented in auditory messages as feedback to the users. In order for users to decode the information carried by the messages, the encoding of data and data relation into sound attributes has to be faithful and informative. While the design of auditory feedback needs to consider many factors, in general, auditory feedback should have the following features.

- (1) Low latency: low latency feedback (e.g., less than 100 milliseconds) is important in every interactive system. It increases user engagement, and allows users to quickly refine their control activities in the exploration process.
- (2) Short: audio is transient. Processing auditory feedback is a short-term memory task. Human's short-term memory (STM) has a limited capacity and decays within 10 seconds unless actively maintained by attention/rehearsal [15][142]. An auditory feedback should be short, maybe up to 10 seconds. A message that does not fit into

the STM should include pauses to allow midpoint STM processing. It is also important to be able to explicitly pause and resume the auditory feedback at any time.

- (3) Synchronized with displays in other modalities: while auditory display is important for users with no residual vision, it can also benefit partially sighted users for whom the visual display is still a significant information resource. Low vision users often wish to be able to use their residual vision. A visual display can be overburdened ([117], p322) and sighted users can become effectively visually disabled by being overloaded with visual information ([38], p3). By synchronizing auditory display with displays in other modalities, especially visual display, partially sighted users may combine the information perceived through multiple channels. Synchronization between auditory display and visual display also makes the collaboration easier between sighted users and visually impaired users.

3.2.4.1 Sound Attributes and General Data to Sound Mapping

As described in section 2.1, variables come in three basic types, nominal, ordinal, and quantitative. Similarly, sound attributes can also be categorized as nominal, ordinal, and quantitative. Table 3-2 lists some of the commonly perceived attributes of a single musical sound.

The choice of what sounds to use and how the data and data relation should be mapped to sound attributes should be based on evidence drawn from scientific psychoacoustic experiments, not just based on whether it “sounds good” or “makes sense” to the designers. While aesthetics is also important in audio design, the choice of data to sound mapping should target at truthful, informative, and intuitive auditory

feedback. There have been many studies on human auditory perception abilities and effective data to sound attribute mapping, as reviewed in section 2.1. Those results can be used as guidelines in auditory feedback design.

Attribute	Type	Examples
Timbre	Nominal	Piano, chime
Pitch	Quantitative	Middle C
Volume	Quantitative	Db
Duration	Quantitative	1 second
Location	Quantitative	45° azimuth and 0° elevation
Onset	Quantitative	After 500 milliseconds

Table 3-2: Commonly perceived attributes of a single musical sound.

Generally speaking, the data to sound mapping should match variable types to equivalent sound attribute types. Within the same type, the effectiveness of different sound attributes in conveying data is different. Furthermore, some sound attributes may not be suitable for conveying data if multiple data items are involved. For example, although psychoacoustic experiments show that the duration of a single sound or the tempo of a sound sequence is good in conveying values, such mapping has drawbacks, especially if the feedback involves multiple data items. Using duration elongates the auditory feedback. If the series of data evolve naturally in time, the duration mapping destroys the analogical relationship between the data and the time. Moreover, multiple data series can not be played in parallel because they may occupy different time scales.

Sound location can carry information. While a stereo sound only carries location information on a 1-D space (left/right), a spatial sound gives a 3-D impression (left/right,

up/down, and near/far). However, sound location is not a reliable way to convey data and data relation, but should rather be used to assist localization and orientation, and to create a rich auditory representation. This is due to several reasons. First, even with sounds arising from real world sources, the spatial resolution of human hearing is very coarse. The ear has angular resolutions approximately 10 times more coarse than the eye across the sensorily richest regions [81]. Second, the current techniques to generate spatial sounds require special resources that are often not convenient to obtain. The loudspeaker-based techniques require a listening environment carefully setup with multiple loudspeakers. The simulation effect is affected by the listener movements. The HRTF-based techniques are very computing intensive, thus requiring fast machines. Additionally, in order to achieve good source location perception, particularly the source elevation, individualized HRTFs are needed. However, individualized HRTFs are unlikely to be available for the general public.

3.2.4.2 Sequencing, Aggregation and Zooming

While humans are able to simultaneously monitor more than one stream of information and recognize subtle changes [12][42], the amount of accurate information that humans are able to extract and obtain from simultaneously presented sound streams is limited [74]. In another word, auditory perception is much less synoptic than visual perception. Multiple sounds representing multiple items often need to be arranged along the time dimension to be played in a sequence instead of all at once. Sound is inherently a temporal medium, and its benefits depend crucially on changes over time [94]. As human eyes are very good at detecting spatial position difference, human ears are very sensitive to changes in the sound evolving along the time dimension. As a result, the time

dimension is one of the most expressive sound attributes. Some data relations can be naturally mapped to the time dimension of the sound. For example, for a time-series data collection such as the price change history of a stock, the most intuitive data dimension to be mapped to the time dimension is the time dimension in the data collection. When the data collection does not involve a time dimension, the data or data relation to be mapped to the time dimension needs to be chosen carefully, so as to best reflect the user's interest and to best help with the user's task.

Researchers have adopted sound sequencing for various graphs, such as bar charts [192], pie charts [61], line graphs [32], 2-D geometric shapes [20], and bit map images [111]. To unify various sequencing methods, Franklin and Roberts [62] described a path based sonification model in which a selectable grid is used to predefine sequencing orders for data points in a continuous 2-D or 3-D space. The model can be useful for data views where the spatial positions of data items carry the most important data relation information. But it does not address the sequencing needs for data representations where spatial positions are of less concern compared to other data relations such as logical connections among data items, nor does it consider the semantic structure in abstract objects, or connect various sequencing methods to specific task needs.

By sequencing sounds, the auditory feedback message can sometime be very long. In order to keep it short, the data can be aggregated before being encoded in the message. For example, while the price history of a stock in one day at 1 hour interval can be presented as one sound per value, the stock price history over the period of 20 years will need to be aggregated and presented at 1 month interval or even 1 year interval. What aggregation to use depends on the task needs. For example, if the user is interested

in the change of average stock prices, then an average function can be used. If the user is interested in the degree of fluctuation in each period, then the aggregated value should be the maximum value minus the minimum value in each period. On the other hand, when the amount of data items to be encoded is small, e.g., one data item, more detailed information can be presented. Speech is often a good choice because it can give precise description about the data item.

While aggregation is necessary for encoding a large data collection into a short auditory message, disaggregation is needed when the user narrows down the attention to a smaller portion of the data collection. The smaller the portion is, the more detailed information can be presented. For sonification that models the physical real world, Axen and Choi [17] describe auditory zooming as going from a simple to a more complex sonification when approaching an object. For abstract data sonification, the term auditory zooming can be used to describe going from aggregated group representations to disaggregated individual representations.

3.2.4.3 Context and Orientation

Sometime the whole data set may need to be represented to serve as a context for the user to better understand the relation of the focus data items to the whole collection. For example, providing the information of the stock price range in the past 10 years can help the users to understand the changes in the past five years. Smith and Walker [165] found that adding some context sound to the sonification of single variable line graphs, in the forms of such as tick marks on the X-axis and beeping reference tone (max/min values) on the Y-axis, helped sighted subjects to understand the data trend and identify specific data points.

Auditory feedback should also provide information to help the user keep oriented during navigation. This is especially important because the user does not have persistent feedback, as is available in the visual display, which tells the user the current position in the data space. Depending on the data view and the navigation structure, various orientation cues can be provided. For example, in a table, when the user moves from one cell to another, the current cell position can be encoded as audio stereo panning [139]. The case or the variable name can be spoken together with the value sound to remind the user of its meaning. If the case or variable names are ordered, the spoken name can help the user judge the current position.

3.3 Summary and Discussion

The Action-by-Design-Component (ADC) framework characterizes exploratory data analysis as a set of auditory information seeking actions (AISA): gist, navigate, situate, search, filter, details-on-demand and adjust information level, select, and brush. The ADC framework also describes design considerations of the Design Components involved to support such actions. The framework can be used to guide the design of interactive data sonifications that are helpful to exploratory data analysis tasks. Many actions and components in the framework are analogous to those in visualization systems but there are many unique design challenges resulting from the highly transient nature of sounds. While there has been no previous work examining whether interaction methods in visualization can be transformed for use without visual aid, I believe that it is important to identify such analogies and examine their validity through the application to specific data sets and user evaluation of the resulting design (see the next few chapters).

The ADC framework also emphasizes the synchronization of visual and auditory presentations and interactions, which allows partially sighted users to effectively combine information perceived through visual and auditory channels. Such synchronization is also helpful for sighted users to communicate with visually impaired users.

Chapter 4

iSonic and the Design Space for Geo-referenced Data

Guided by the Action-by-Design-Component (ADC) framework, I explored the design space for geo-referenced statistical data and designed iSonic, a data exploration tool for visually impaired users. Geo-referenced statistical data was chosen as the focus design example because it is a legal requirement [155] for any statistical data published on a U.S. federal government web site to be accessible to users with visual impairments. Such data is often related to predefined administrative or census geographical regions, such as population distribution by U.S. states. Such data is often presented as choropleth maps that typically use colors to show the value of each map region.

Multiple design iterations were carried out. Two users without residual vision were closely involved as regular consultants. Guided by the ADC framework, many design options were investigated, along the dimensions of data view, navigation structure, input device, and auditory feedback.

This chapter first presents several early exploratory designs to support geographical data distribution pattern recognition, with focus on the role of data view, navigation structure, and auditory feedback encoding. Two controlled experiments with sighted users examined those earlier designs and partially the utility of AISAs.

Then the remaining of this chapter describes the current iSonic design, which was based on lessons learnt from those earlier explorations. The current iSonic interface also incorporates other AISAs and design components to support general data exploration

tasks besides geographical pattern recognition. The description focuses on the table and map views, leaving details of the scatterplot view to Chapter 7. The auditory display is highly synchronized with the visual display in iSonic, so this chapter also shows many visual figures of the interface.

Readers can try iSonic or viewing the video demonstration by visiting www.cs.umd.edu/hcil/iSonic.

4.1 Explorative Designs and Evaluations

The exploration of geo-referenced data design space started with several earlier explorative designs to support geographical data distribution pattern recognition. Guided by the ADC framework, the exploration examined the dimensions of data view, navigation structure, and auditory feedback encoding. Two controlled experiments were conducted to evaluate those design options, and to partially validate the utility of the AISAs. Another purpose of the experiments was to gain insight into the users' ability to perceive data patterns from interactive sounds.

4.1.1 Geographical Data Pattern Recognition

Geographical data distribution pattern recognition was chosen as the task because geo-referenced data analysis often involves geographical contexts. In the visual mode, a picture is often said to be worth a thousand of words. A quick glance at the geographical distribution pattern of the data often gives users very valuable information. The geographic distribution pattern of geo-referenced data involves three dimensions, where the data points scatter on a 2-D map plane with the values of the data points as the third dimension. Research has shown that users can interpret a quick sonified overview of 2-D

line graphs containing a single data series [59], and two data series [25][32], and bivariate scatterplots [60]. Research has also shown that users can recognize 2-D graphical shapes by listening to sounds tracing the border of the shapes [6].

However, there have been few observations about the ability to recognize data distribution patterns with more than two dimensions in the auditory mode. Meijer's work [111] aims to let visually impaired users "see" with hearing. It translates an arbitrary image into a time-multiplexed sound representation that is the superposition of the sound of multiple image pixels. The effectiveness of this approach remains to be established. Wang and Ben-Arie [186] found that sighted users can recognize simple shapes on binary images of 9 x 13 resolutions where the pixels are raster-scanned slowly. Jeong [87] showed that sighted users can locate the minimum or maximum value on a simplified choropleth map with up to nine geographic regions, with the values presented as different sound volumes.

The two controlled experiments investigated whether subjects could perceive geographical distribution patterns of a five-category data on a 51 region U.S. state map (not grids or simplified maps). There are 50 U.S. administrative states plus the District of Columbia. They are referred to as 51 states in this dissertation. Subjects perform the tasks by using a series of Auditory Information Seeking Actions (AISAs) including gist, navigate, and details-on-Demand. So the two experiments also provided early observations into the utility of the ADC framework.

The first controlled experiment is a pilot study to gain some initial insights into whether subjects are able to perceive such patterns. It also compared the effect of data views, i.e., a choropleth map and a table enhanced with geographical knowledge [194].

The second controlled experiment further examined subjects' pattern recognition ability and compared the effectiveness of two navigation structures and two auditory feedback encodings on a choropleth map [195]. In each experiment, the subjects performed a set of geographical pattern recognition tasks. Sighted subjects were used and they need to choose the matching pattern from a set of visual patterns.

4.1.2 Effect of Data View

The hypothesis of this first study was that a map view was better than a table view for geographical data distribution pattern recognition tasks. The study was a within-subject controlled experiment between a table-based design and a map-based design. The data in the study is simulated, and categorized into five value ranges.

4.1.2.1 Two Sonification Designs in the Study

Enhanced Table

In the enhanced table, data was mapped to sounds as follows: Five value categories are mapped to five string pitches (pitches from a string instrument) from an increasing scale of CEGCE starting from the middle C on a piano keyboard. A lower pitch indicated a lower value category. The value pitch of each U.S. state played for 200 milliseconds. All sounds are stereo without panning

Gist: The gist was a column sweep of all the 51 U.S. states that lasted for about 30 seconds, with three continuous bells indicating the end. For each state, the state name was spoken while the string value pitch was played. The sweep went from the top to the bottom in the table.

	Enhanced table	Spatial choropleth map
<Up>	Go to & play previous state in table	Go to & play previous state in the sweep order
<Down>	Go to & play next state in table	Go to & play next state in the sweep order
<Pg Up>	Go to & play first state in table	Go to & play north-most state in current sweep column
<Pg Dn>	Go to & play last state in table	Go to & play south-most state in current sweep column
<Left>	N/A	Go to & play nearest state to the left of current state
<Right>	N/A	Go to & play nearest state to the right of current state
<Home>	N/A	Go to & play state in first sweep column that is nearest to current state
<End>	N/A	Go to & play state in last sweep column that is nearest to current state
<Enter>	Play gist of all states	
<R>	Play sub-gist starting from current state	
<Space>	Request detail of current state	
<L>	Play audio legend	
<Any key>	Stop playing the gist, set current state to be the state just played in the gist	

Table 4-1: User key controls in the first study.

The order of states in the table reflected a spatial sweep from west to east and north to south according to the states' geographical locations on the real map, and was exactly the same as the sweep order in the spatial choropleth map design (see the next section for details). While a traditional table is often alphabetically ordered by state names, it was expected to be hard even for users with excellent knowledge of the map to catch the data distribution pattern because they have to mentally switch between locating a state from the state name and processing the state's value. Meanwhile, it was expected that a table-based design can be enhanced by adding extra geographic information about

a state in the form of a data column, such as west part, middle part, and east part. The enhanced table design in this study used the same sweep order as in the map-based design to maximize geographic location cues and to minimize users' mental load. I believed that such enhancement made the comparison to map-based designs more meaningful.

Navigate: Subjects used a keyboard to navigate the soundscape in the table order, such as moving up/down the table to play one state on one keystroke, jumping to the first/last state, and so on. The playback tempo depended on how fast subjects pressed the keys. Table 4-1 lists all the key functions.

Details-on-demand: Anytime subjects pressed the 'space' key, the name and value of the current state were spoken.

Spatial Choropleth Map

In the spatial choropleth map design, all sounds, including the speech, were spatial sounds synthesized using KEMAR mannequin HRTF. Spatial sounds were tied to the map to create the effect of a virtual half-cylinder shaped map surrounding the subject located at the center (Azimuth range: $-90^{\circ} \sim +90^{\circ}$, elevation range: $-31^{\circ} \sim +63^{\circ}$) (Figure 4-1). Since non-individual HRTF has shown to result in poor elevation perception [187], a piano pitch was used in addition to indicate the vertical position of each state. The piano pitches ranged from about one octave below middle C to about two octaves above middle C. Lower pitches indicated states to the south. Five value categories were mapped to five string pitches, same as in the enhanced table.

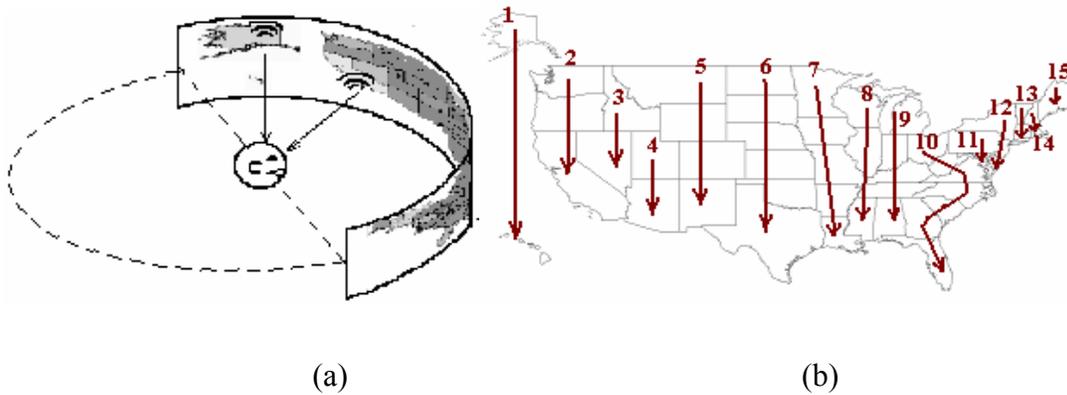


Figure 4-1: Spatial choropleth map design: (a) spatial sounds tied to the map create the effect of a virtual half-cylinder shaped map surrounding the subject located at the center. The illustration does not reflect the real spatial parameters; (b) vertical map sweep order.

Gist: The gist was a vertical spatial sweep of all the 51 U.S. states, following the fifteen sweep columns shown in Figure 4-1 (b), from north to south in each column and from column one to column fifteen. For each state, a value pitch was played for 200 milliseconds followed by a vertical position pitch lasting for 100 milliseconds. A bell rang for 570 milliseconds at the end of each column. The gist lasted about 25 seconds, including a 100- millisecond pause between two adjacent columns and three continuous bells indicating the end.

Navigate: Users used a keyboard to explore the soundscape simply following the sweep order. For example, pressing the <Up>/<Down> key went to and played the previous/next state along the sweep order. Within one column, this was the same as going to the state to the north/south of the current state. When two columns were involved, a bell sound was played before jumping to the most southern state of the previous column or the most northern state of the next column. Furthermore, subjects could do 2-D map

navigations. For example, pressing the <Left>/<Right> key went to and played the nearest state to the left/right of the current state. Table 4-1 lists all the key functions.

The fact that states have irregular shapes and sizes on real maps introduced extra difficulties in both defining a good sweep order for the gist and defining an effective 2-D map navigation grid. In the one-state-per-key-stroke type of navigation using only four directions of up, down, left and right, it was not easy to choose the target state there were multiple adjacent states in the direction of movement. For example, in Figure 4-1 (b), on a <Left> keystroke from the top state in column seven, which one of the top two states in column six should become the current state? In this study, the map navigation was based on the sweep columns. For example, a <Left> keystroke took the subject to the state in the previous column whose geographical center was nearest to the current state. An associated problem with this was that sometimes a series of continuous movements in one direction, such as to the left, might actually take the subject in a direction shifted northern or southern from the straight left. Alternative navigation design options will be discussed in Section 4.1.2.3 . *Details-on-demand*: Anytime subjects pressed the ‘space’ key, the name and value of the current state were spoken.

4.1.2.2 Study procedure

Nine sighted University of Maryland students/staff (2 females and 7 males, ages from 19 to 40) were paid to participate in the study. Eight were intensive computer users. Two had some music training in high school. A pre-test showed six subjects had excellent knowledge of U.S. state locations, while the others did not know the accurate location of some states, especially states in the middle region. The pre-test also tested subjects’

ability to perceive the spatial sound used in the spatial choropleth map interface. The test showed that all subjects were able to point with arm with very good approximation the five azimuth locations tested (approximately -90° , -45° , 0° , 45° , and 90°), but were often confused about the three elevation locations tested (approximately -31° , $+34^\circ$, $+63^\circ$), all using the same string sound. Subjects could easily tell the timbres used and distinguish the five value pitches well. Subjects were told before the study that the best performer would receive extra monetary rewards.

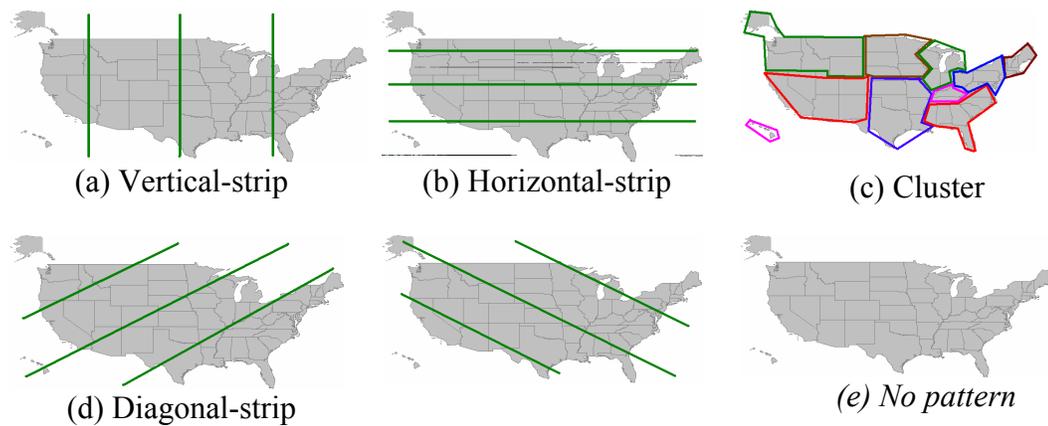


Figure 4-2: Five general pattern types

The assignment of subjects to the two design conditions was counter-balanced. Before the test of each design, subjects learnt the sound design and interface controls, and practiced one training task following the same procedure as in the real test. Subjects performed four tasks in each design. The study lasted for a little over one hour on average.

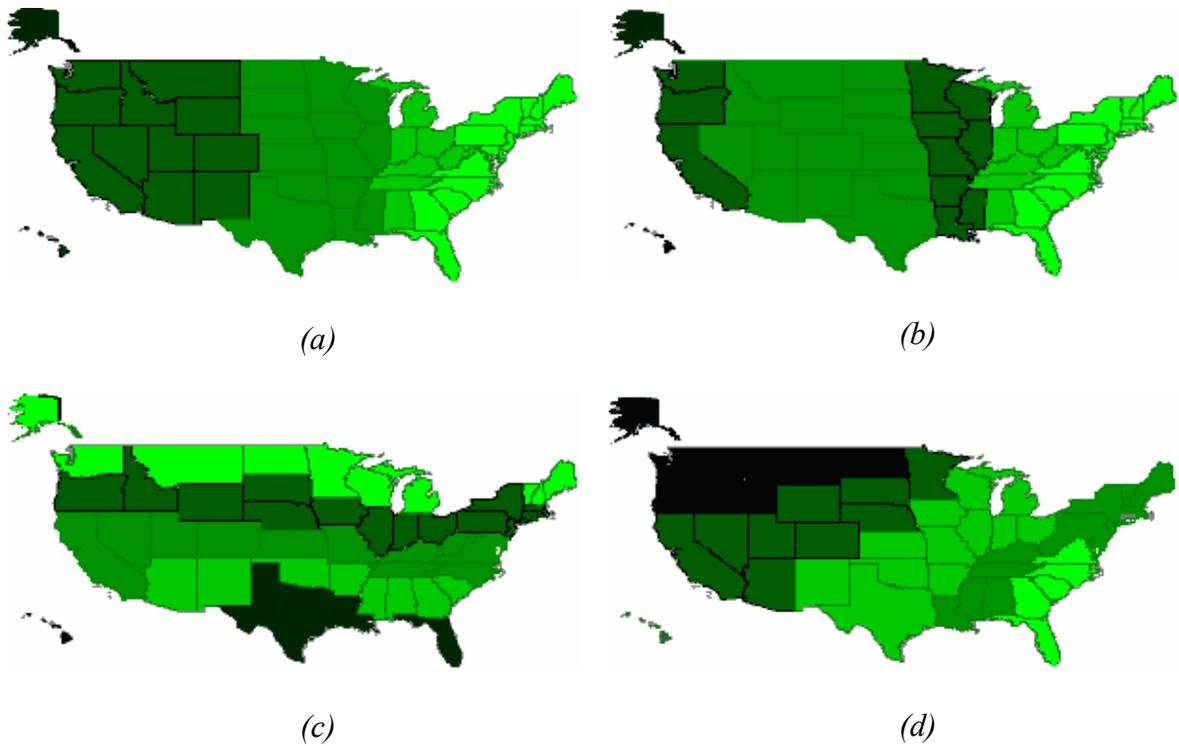


Figure 4-3: A sample of four visual pattern choices. A darker green presents a higher value. Values are categorized into five ranges.

Each task was carried out in three steps. First, subjects listened to the gist of the data once, and were asked whether they perceived any pattern in the data by choosing from the five pattern types shown in Figure 4-2. Second, subjects explored the sound scene by using the key controls listed in Table 4-1 for as long as they needed or up to 3 minutes. Subjects then chose the matching pattern from four visual pattern choices, and chose their confidence level about the answer based on a 10%-break scale (Note: the confidence level data is not used in the analysis). Figure 4-3 shows a sample of such four visual pattern choices. Pattern (b) was the pattern played. Pattern (a) was considered to be similar to (b). Pattern (d) was less similar to (b) and pattern (c) was the least similar. The

patterns in each task were designed in a way that the correct pattern in an earlier task did not appear in the choices for any later tasks. Third, the subjects were told the correct answer and had the chance of exploring the sound scene again for up to 1 minute. Subjects were timed for their exploration and time to choose the answer. The post-test questionnaire asked subjects about the overall experience with the two designs, their pattern recognition strategies, and so on.

4.1.2.3 Results and Discussion

In the visual mode, such pattern recognition tasks would take only seconds. This study result showed that it was much harder in auditory mode, but novice users were still able to do it with overall good accuracy in reasonable time, as shown in Figure 4-4 and Figure 4-5. After listening to the gist once, the subjects recognized the pattern type with an overall accuracy of 56% in both designs. It was significantly higher than the chance accuracy of 20% (for table, $t(8)=3.9$, $p=0.005$, for map, $t(8)=6.1$, $p<0.001$). After exploration, subjects were able to obtain pattern details and choose the correct pattern out of a few similar choices with good accuracy (67% in table and 75% in map). The task accuracy was significantly higher than chance accuracy of 25% (for table, $t(8)=5.8$, $p<0.001$, for map, $t(8)=12.0$, $p<0.001$). Subjects spent an average of 110.7 seconds exploring the table and 110.6 seconds the map. The average task time for recognizing pattern details was 130.0 seconds for the table and 134.2 seconds for the map. After exploration, subjects recognized patterns better in the map than in the table, but no statistically significant difference was obtained.

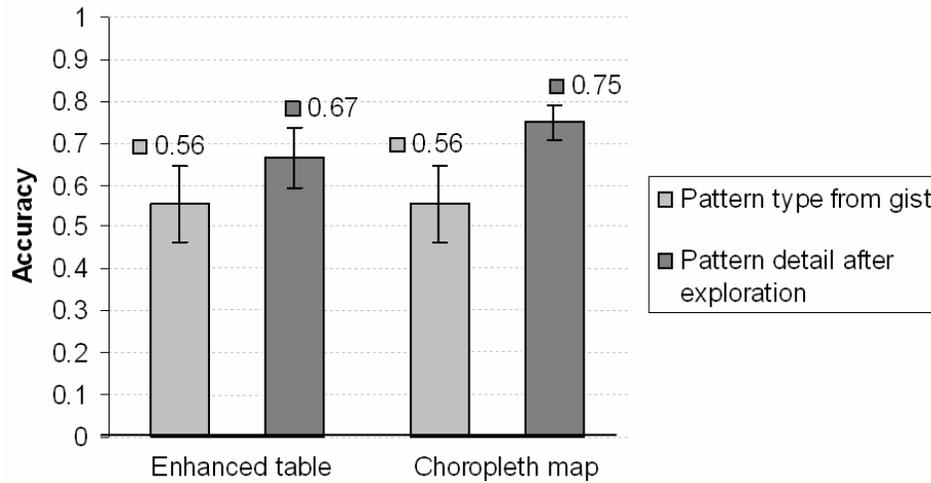


Figure 4-4: Overall accuracy of pattern recognition (n = 9). The answer of the pattern type after exploration was inferred from the pattern detail chosen after the exploration.

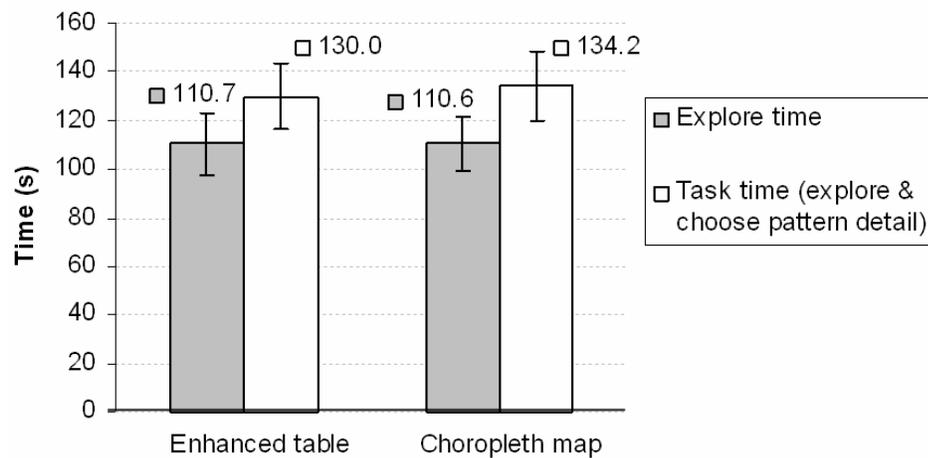


Figure 4-5: Overall speed of pattern recognition.

Subjects used different task strategies. A common strategy used was “the gist gives a general idea of the pattern type, I explore to confirm my guess, find where the value changes happen, try to ‘visualize’ the pattern, and remember the pitches of a few states/clusters/areas to eliminate the wrong patterns”. The result showed that subjects

were able to obtain the “big picture” from a short gist, which guided their exploration, which in turn confirmed their initial impression and provided more detail for choosing the correct pattern. This showed evidence for the utility of AISAs in the ADC framework.

Since the pattern type recognition accuracy was the same in the table and the map before the exploration, neither the use of map view nor the sound positions in map seemed to have helped during the passive gist presentation. It was not too surprising that there was no difference between the map and table interfaces before the exploration, because everything was the same except for the change of sound position in the map interface. However, after the exploration, subjects achieved higher accuracy in the map than in the table. Most subjects used west-east navigation on map and commented that it gave them the flexibility of moving in the directions they want in order to confirm the pattern, thus helping them to remember the pattern. As for the sound position, most subjects reported that the left-right position cue from the spatial sound in map helped them to orientate and “picture” the data distribution pattern. It seemed that sound stereo panning was an important feature during the exploration and might be one of the reasons why the map interface was better. This indicated that the usefulness of stereo was brought out through the interactivity. These results confirmed the importance of choosing an appropriate data view and that interactivity is a key element in the design of auditory interfaces. The lack of a statistically significant difference between the map and table interfaces might be due to the small numbers of subjects, and the inefficient map design used in the study.

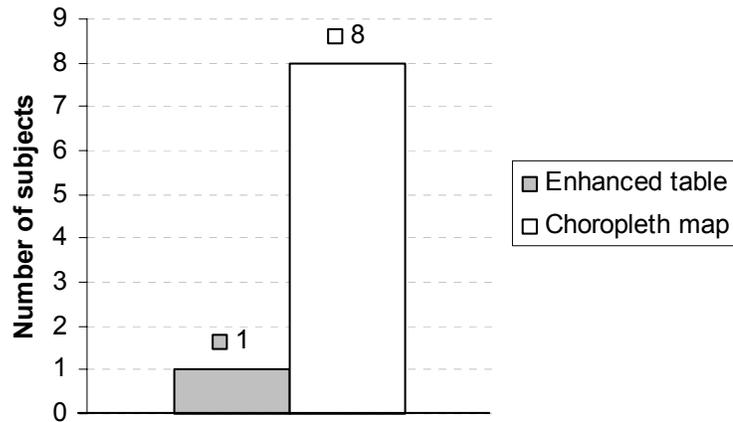


Figure 4-6: User preference.

Most subjects (8 out of 9, Figure 4-6) strongly preferred the choropleth map and commented that they could probably do better and faster in the map with more practice. Analysis of the error cases and subjects' feedback showed the following reasons why the map design used in the study did not perform as good as expected.

(1) Subjects needed time to better understand the sweep column definition and adjacency in 2-D map navigation. The irregular state shape and size in the real map imposed special difficulty in placing states in the sweep grid, and sequencing the states. Sometimes the sweep order and navigation order did not meet the subjects' intuitive expectation. For example, subjects expected the beginning of every column to start at the northern boundary of the map. But in fact sweep column #4 started from the middle. The arrow key adjacency navigation conformed to the sweep column definition and sometimes was very different from the actual adjacency. The subjects were quickly

shown the sweep columns before the test and had the chances to learn more during the learning stage at the end of each task. However, due to the short experiment time, there was not enough tasks/time to ensure the enforcement of the knowledge.

The importance of familiarizing subjects with column definitions was also confirmed by my experience with one blind design partner. Before using a tactile map to learn the sweep columns, he could only tell west-east trends. After brief learning, he could describe the patterns in high details, in both the west-east and north-south directions.

(2) The vertical position piano sound interfered with the value string sound. The piano sound was provided to aid north-south localization of a state, as a supplement to the fact that the non-individual HRTF spatial sound technique did not provide sufficient vertical position cues. This was also expected to solve the problem in (1). But in fact, all subjects reported that two short pitches with one (piano pitch) following immediately the other (string pitch) were too overwhelming for them to process both, especially when multiple states were played continuously. This was consistent with McGookin and Brewster's result [25] about users' identification ability of concurrently presented earcons. As a result, all subjects chose to focus on the value sound and ignore the position sound when they listened to the gist, occasionally shifting quickly to a piano sound to confirm the current location. Some subjects seemed to be able to filter out the second sound, but some subjects reported that the second sound was so distracting that it caused their misinterpretation of the value pattern. All subjects reported that it was much easier to locate the current state through the between-column bell sound, the navigation direction using the arrow keys, and the general west-east, north-south sweep order they

already know.

Comparatively, the verbal presentation of state names in the table was not rated as being distracting as the piano sound in the map. Some subjects reported that it was difficult to switch between state names and value sound frequently, but they also commented that there was no need to pay attention to every state name. Instead, they followed the west-east, north-south order in which the states were played, searching for some specific states as their position anchors. The anchor states usually were the states the subjects were familiar with, and/or the north-boundary/south-boundary states that indicated the sweep column change. With the accurate position cue provided by the anchor states, the subjects then placed other states in between following the west-east, north-south sweeping order.

The role of state names was clearly demonstrated by subjects' keystroke patterns during the exploration. They often pressed arrow keys fast to skip the playing of state names, and slowed down for state names once they encountered a change of the value pitches or they expected a change in columns. This was also demonstrated in the map where the requests for state names mostly happened upon change of value pitches and change of columns.

4.1.3 Effect of Navigation Structure and Auditory Feedback

Although no statistical significant difference was obtained in the first study, the results, observations, and user comments supported the advantage of a map view over a table view for geographical pattern recognition tasks, and indicated the importance of interactivity in the design of auditory interfaces. Based on observations and subjects'

comments in the first study, I expected that the map-based design could be significantly improved. The following two observations led to the design of the second user study:

- (1) Some subjects reported the sound that indicated the state vertical position distracted them from the value.
- (2) Irregular state shapes and sizes impose difficulty on defining a good state-by-state navigation matrix. It often caused the actual navigation direction to drift away from the subjects' expectation, possibly causing misinterpretation.

4.1.3.1 Study Design

The second user study was a 2 x 2 controlled experiment with two between-subject factors. There was also a within-subject factor which was the repeated measurement of the general pattern type recognition before and after the exploration. Each subject was assigned to one of four interfaces, defined by the two between-subject factors.

The first factor was the presence or absence of a vertical position sound (VPS). There were two treatments. In the treatment without the VPS, for each state, only a 200-millisecond string pitch was played to indicate the state's value. In the treatment with the VPS, for each state, a 100-millisecond piano pitch followed the value pitch to indicate the vertical position of the state. The ranges of the value pitches and VPS pitches were the same as those in the first user study.

The second factor was the navigation method (NM). There were two navigation methods, namely, the column and mosaic interfaces. The column interface used state-by-state navigation (see Figure 4-7 (a)) similar to the navigation method in the first user study. The differences were that there was no automatic jumping between adjacent sweep columns when the top or bottom of a column was reached, and the control keys were

remapped using only the keys in the number pad plus the four arrow keys on a standard keyboard to control the interface. Table 4-2 lists all the controls available to the subjects.

Key	Mosaic navigation	Column navigation
8	Go to closest valid cell* above current cell, play sound**	Go to & play previous state in the current sweep column
2	Go to closest valid cell* below current cell, play sound**	Go to & play next state in the current sweep column
4	Go to closest valid cell* left to current cell, play sound**	Go to & play state in the previous sweep column that is nearest to current state
6	Go to closest valid cell* right to current cell, play sound**	Go to & play state in the next sweep column that is nearest to current state
Up, Down arrow	Go to North- (South-) most valid cell*, play sound**	Go to & play North- (South-) most state in current sweep column
Left, Right arrow	Go to West- (East-) most valid cell*, play sound**	Go to & play state in first (last) sweep column that is nearest to current state
1, 3, 7, 9	Go to closest valid cell* diagonally adjacent to current cell, play sound**	
Enter	Play gist of all states	
0	Play sub-gist starting from current state	
5	Request state name and value pitch of current state	
+	Request spoken value of current state	
Any key	Stop playing the gist, set current state to be the state just played in the gist	
* Valid cell: a cell that is part of a state		
** Play Sound: play a percussion sound for each background cell crossed, then play the target state if it is different from current state		

Table 4-2: User key controls in the second user study. Only keys from the number pad plus the four arrow keys were used.

The mosaic navigation method was based on a map created by laying a grid over the regular map (see Section 5.3). Using the number pad, subjects navigated the cells in eight directions. When subjects moved from a cell to another cell in a different state, the sound(s) of that state were played. When the movements were within one state, no sound was played. The purpose of this was to allow the subjects to sense the size of a state. When a movement crossed the background (e.g., an ocean, which had no data values) to get to a state, a series of percussion sounds were played before the target state's sound(s) were played. The number of percussion sounds played equaled the count of background cells crossed. All the key controls were on the number pad plus the four arrow keys.

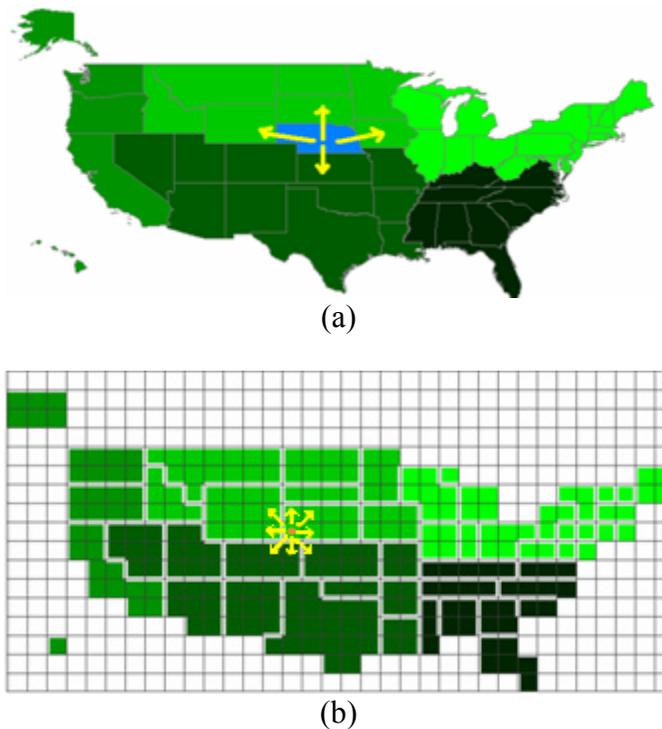


Figure 4-7: Two navigation methods on a choropleth map: (a) state-by-state column navigation; (b) cell-by-cell mosaic navigation.

Whenever a border was reached and subjects could not go further in that direction, they were reminded by a synthesized female speech that they were already at the boundary. During the sweeping and the navigation, the visual display was synchronized with the auditory presentation. This was to help the subjects understand the interfaces better during the interface explanation. During the training task and experimental tasks, the display was hidden from the subjects, and could be seen only by the experimenters. In the column navigation, the current state was highlighted in a blue color. In the mosaic navigation, the current cell was always marked by an orange dot. The dot also moved across white space when a national border or ocean was crossed.

All four interfaces used the same sweep order as in the first user study (vertical sweep order, see Figure 4-1 (b)). Since all subjects from the first user study reported they could not tell the vertical positions of the non-individual HRTF spatial sounds used, stereo panning (0 ~ 127) was simply used in this study to indicate left-right sound positions. The sound to indicate the end of each sweep column and the end of a sweeping was changed from a bell sound in the first study to a percussion sound in this study. The bell sound was used to indicate the end of the sweeping.

Forty-eight subjects from introductory Psychology courses participated in the study to earn extra credit. Ages ranged from 18 to 51 with a median of 20. There were 37 female and 11 male participants. All subjects reported using computers at least one hour per week, and 44 reported using them at least 5 hours per week. Fourteen subjects reported having had professional music training for a year or more. None of these factors were significantly correlated with performance on any outcome measure at the 0.05 level. Subjects were randomly assigned to one of four interface conditions.

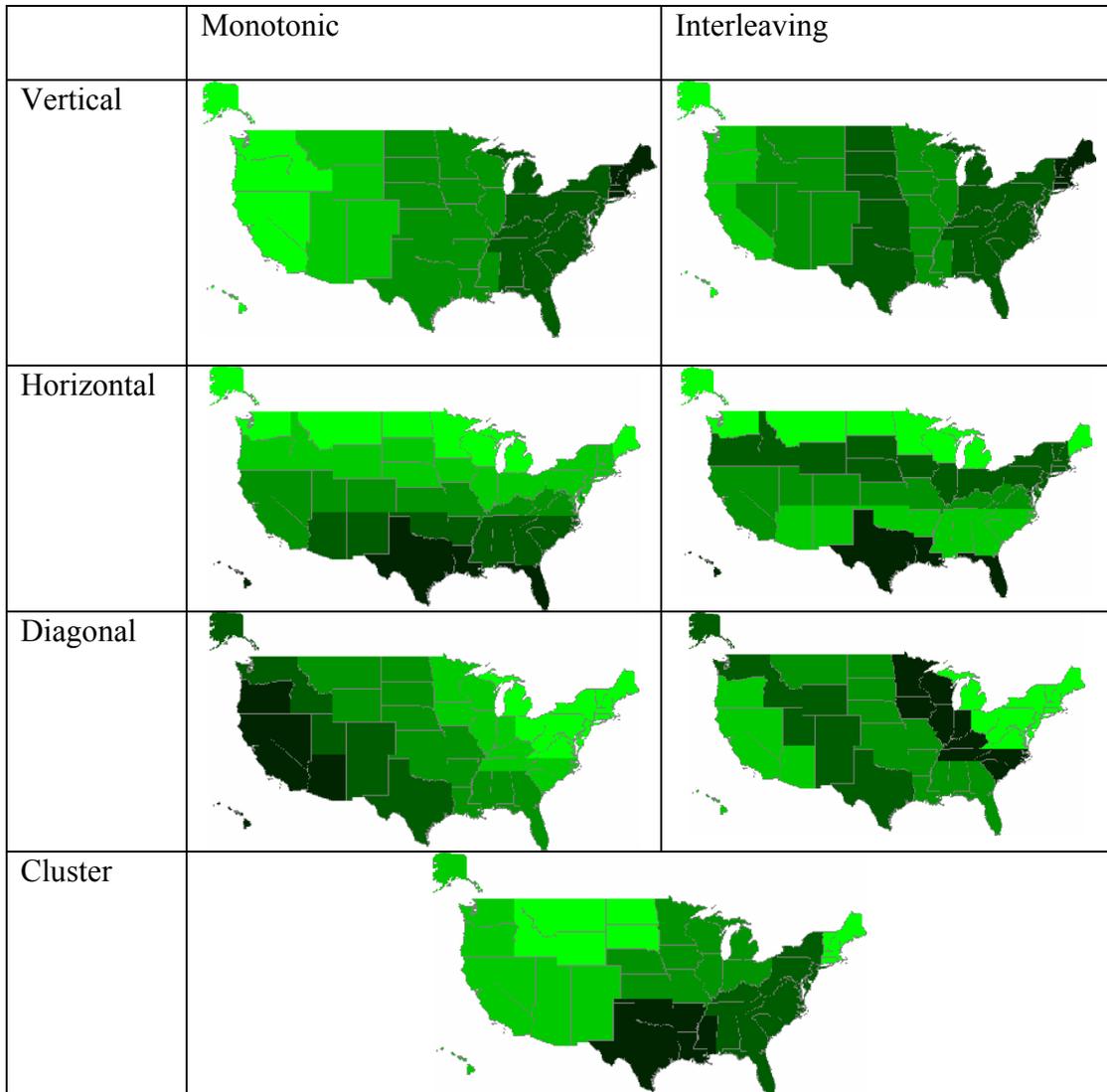


Figure 4-8: Sample patterns for each pattern type.

First, each subject was tested on the ability to distinguish the instruments, pitches, stereo panning, and on geographic knowledge, similar to the pre-test in the first study. Subjects were then taught how to use the interface they had been assigned. (Two experimenters, neither of whom knew the focal hypothesis, ran the experiment; there were no significant differences in performance for subjects run by each experimenter.)

The subjects learned the sound design and interface controls while viewing the display. The display was then hidden from the subjects, and they practiced by performing a training task, with a monotonically horizontal-strip pattern, following the same procedure as in the real test.

Each subject performed 6 pattern matching tasks. Figure 4-2 and Figure 4-8 illustrate the pattern types and patterns used. There were two tasks with vertical-strip patterns (one monotonic and another interleaving), two tasks with diagonal-strip patterns (one monotonic and another interleaving), and two tasks with cluster patterns. The task order was counter-balanced using a Latin square. The orders were set up such that no subject ever had the same general map type twice in a row. Each subject was randomly assigned to one of the six resulting task orders. Subjects were notified that both accuracy and speed would be measured, but accuracy was more important.

The task procedure was similar to the first user study but with a few changes. Each task was carried out in three steps. First, subjects listened to the gist of the data once, and were asked whether they perceived any pattern in the data by choosing from the five pattern types as shown in Figure 4-2. Subjects also chose their confidence level about the answer based on a 10%-break scale. Second, subjects explored the map by using the key controls listed in Table 4-2 for as long as they needed or up to 3 minutes. Subjects then chose the pattern type again, and their confidence level. Third, four maps were presented to the subjects, each with the same pattern type. One of the four maps was the actual map that the subjects had been exploring, and thus the general pattern was not necessarily the same as the pattern chosen by the subject. Subjects then chose the matching pattern from the four visual patterns, with their confidence level (Note: none of

the confidence level data is used in the analysis). Figure 4-9 shows a sample of such four visual pattern choices. At the end of the six tasks, the subjects were given a post-test questionnaire. The entire experiment took less than an hour per subject.

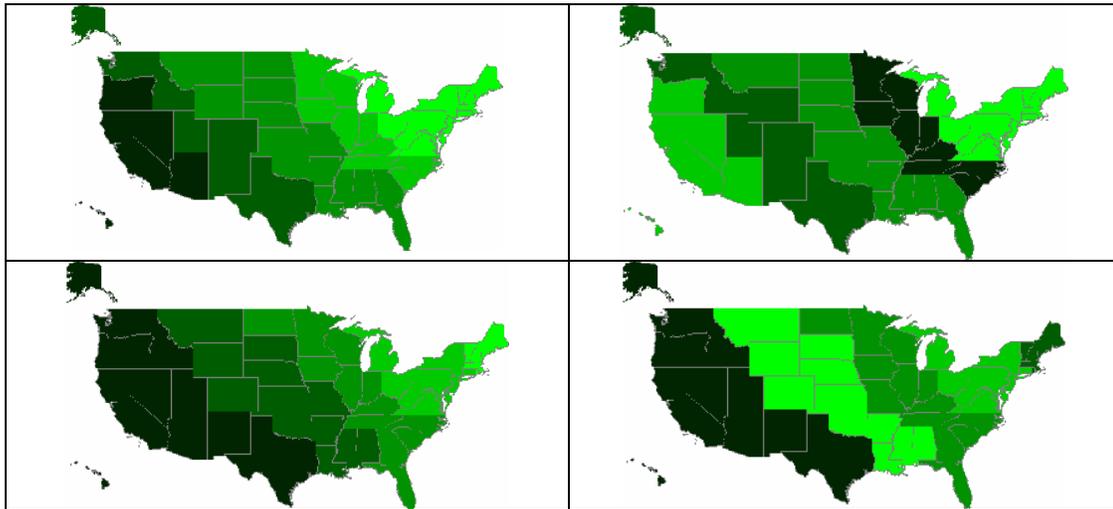


Figure 4-9: A sample of four visual pattern choices.

4.1.3.2 Results

Subjects were able to perceive five-category value distribution patterns on a 51-state real map, although overall the tasks were difficult. The average pattern type recognition accuracy was 50.7% after the gist but before exploration. It was significantly higher than the chance accuracy 20% ($t(47)=8.5$, $p<0.001$). After exploration, the pattern type recognition accuracy increased slightly to 55.2% which was significantly higher than the chance accuracy 20% ($t(47)=8.723$, $p<0.001$). The specific pattern recognition accuracy was 48.7%, significantly higher than the chance accuracy 25% ($t(47)=6.311$, $p<0.001$). There are at least two likely explanations for the low accuracy for this study compared to that in the first study. First, the tasks in this study were more difficult. After exploration,

subjects had to choose the general pattern type explicitly, rather than having their choice inferred from their selection of a specific pattern. Similarly, subjects had to choose a specific map from four similar choices of the same pattern type, rather than from a set of choices of different pattern types. Second, the differences between the subject populations possibly accounted for some of the difference. In this study, subjects were not paid and were not given performance incentives, as subjects in the first study were. Since the tasks were difficult, the subjects' motivation played an important role.

The general pattern type recognition data was submitted to a 2x2x3 mixed ANOVA, in which the two interface factors (presence vs. absence of VPS, column vs. mosaic NM) served as the between-subject variables and the exploration (before vs. after) served as the within-subject variable. None of the main effects were found significant. There were also no significant 3-way or 2-way interactions at the 0.05 level, except that the interaction between the exploration and navigation method approached significance ($F(1, 44)=3.126, p=0.084$). Subjects who used the column navigation method did significantly better in general pattern type recognition ($t(23)=-2.532, p=0.019$) after exploration, while subjects who used the mosaic navigation actually did slightly worse after exploration.

No statistically significant difference was found in task performance across the four interfaces. However, subjects' performance was found to be significantly affected by three correlational factors:

- (1) Geographical knowledge. Subjects' knowledge of U.S. geography was positively correlated with performance on identifying both the general ($r = 0.31, p<0.05$) and specific ($r= 0.36, p<0.05$) patterns after the exploration period.

- (2) Pitch differentiation ability. All subjects were able to distinguish between three pitches on our pretest, but seven subjects needed a second try to get the answer correct. These subjects generally did worse on all outcome measures, and significantly worse on the general pattern type after exploration ($r = 0.45$, $p < 0.05$).
- (3) Task strategy. The post-test questionnaire asked subjects to describe what strategies they used. Their strategies were compared by identifying common words and phrases in the descriptions. Some strategies appear to have been more effective than others. Subjects who reported listening for changes did particularly well on the two outcome measures taken after the exploration period. (For the general pattern type after exploration, $r = 0.36$, $p < 0.05$; for the specific pattern type, $r = 0.31$, $p < 0.05$). Subjects who reported trying to visualize the map did particularly well identifying the general pattern type after the gist ($r = 0.49$, $p < 0.05$). Subjects who reported paying attention to the piano sound (which indicated the vertical position of the states) did significantly worse identifying specific maps ($r = 0.30$, $p < 0.05$). The most common strategies reported were moving around to find particular states, and visualizing the map.

The tasks varied in difficulty, and each had only a single pattern type, so there is no way to separate the difficulty of identifying patterns or maps from the difficulty of the specific tasks given to the subjects. However, we can report that the easiest tasks were the vertical patterned maps, and the most difficult map was a diagonal one. Surprisingly, the only task for which identification of the general pattern type improved significantly after the exploration period was the vertical, monotonic map ($t(47) = 2.72$, $p < 0.05$). See Table

4-3 for details. This may have suggested that first impression, which was conveyed by gists in the study, was very important.

	Horizontal (train)	Cluster 1	Cluster 2	Diagonal Interleaving	Diagonal Monotonic	Vertical Interleaving	Vertical Monotonic
Gen1	0.44	0.44	0.54	0.50	0.50	0.63	0.44
Gen2	0.42	0.50	0.50	0.44	0.52	0.67	0.69
Spec	0.42	0.52	0.52	0.29	0.46	0.63	0.50

Table 4-3: Average recognition accuracy for different general pattern types (n = 48).

Gen1 indicates identification of the general pattern type after the gist. Gen2 indicates identification of the general pattern type after exploration. Spec indicates identification of the specific map. Chance performance levels are 0.20 for Gen1 and Gen2, and 0.25 for Spec. The horizontal pattern was the training task and was always first. The other three patterns are averages of two tasks and were counterbalanced throughout the rest of the study.

The post-test questionnaire asked subjects to rate the task difficulty, and the interface difficulty. Subjects in all conditions found the task hard (2.8 on a 7-point Likert scale, with 1 being very difficult and 7 being very easy). Subjects using the column navigation interface found it easier to use than subjects using the mosaic navigation interface. On the same scale, column navigation subjects rated the interface 6.0, mosaic subjects rated it 5.2. This difference was significant ($F(1, 44)=5.23, p<0.05$).

Two questions on the post-test questionnaire asked about the stereo sound. The first asked whether the stereo sound helped subjects locate states. The responses were just above neutral (4.5) on a 7-point Likert scale from 1 being very distracting to 7 being very helpful. The second question asked whether the stereo sound helped them picture the data distribution. Subjects that did not have the vertical position sound gave higher responses (5.2) to this question than subjects that had the vertical position sound (4.4) ($F(1,44)=5.21, p<0.05$).

Two questions asked whether the vertical position sound was helpful for locating states or picturing the data distribution (only subjects in one of these conditions were asked this question). The average responses were very close to neutral, 3.8 to the first and 4.2 to the second, on the same scale as that for the stereo sound.

A question asked subjects how good of a sense they had of where they were on the map. On a 7-point scale, subjects gave a response of 4.6, which was just better than “Some sense” (4) but considerably lower than “Good Sense” (7). There were no significant differences between conditions for this question.

Finally, subjects were asked about the tempo of the sounds. The average response was 4.5 on a 7-point scale, close to “The right tempo” (4), on the side of “too fast”.

4.1.3.3 Discussion

Because there were no significant differences in performance between the four interface conditions, caution must be taken in drawing any conclusions. However, it can be concluded that the vertical position sound seems to have been unhelpful at best, and subjects that reported paying attention to it did worse than those that did not. It also seems to have taken away from utilization of the stereo sound. There does not appear to

be any advantage to using a vertical position sound, and combined with the added complexity of such a sound, we decided not to use it in future sonifications.

The column navigation interface was ranked significantly easier to use than the mosaic navigation interface, although there were no significant differences between the interfaces in terms of performance. It is not exactly clear what subjects preferred about the column navigation, but it was somewhat simpler to use, with fewer keystrokes necessary, and auditory feedback given after every keystroke. This suggested a possible improvement to the mosaic navigation interface, which is to provide sound feedback after every keystroke. In the mosaic navigation, the sound of a state was played only when the subject entered the state. No feedback was given when the subject moved within the state unless the key <5> or <+> was pressed. The design was intended to give the subjects some sense of the state size. In a new mosaic navigation design, both design considerations can be met by providing state value feedback on every keystroke, and playing a special sound before the state value sound to indicate whether the user has entered a new state.

Some of the difficulties subjects had in identifying patterns and maps may have been due to the experimental conditions, and not the interface itself. Subjects in the study had a few minutes to learn the interface, but might have benefited from more learning time. Visually impaired users would likely spend considerable amounts of time learning and using such systems, so more training time is reasonable. It might also be useful to suggest certain strategies to users during training, such as visualizing the map and trying to listen for changes, as these strategies were helpful for other users.

4.1.4 Lessons and Discussion

The results from the two user studies led to the following lessons about auditory interface design and 5-category geographical data distribution pattern recognition on 51-region real maps (not simulated maps):

- (1) Geographical pattern recognition through audio is hard but novice sighted users can do it.
- (2) A map view, which conveys the geographical relation of data items and allows 2-D navigation in the geographical context, should be used instead of a table view, for such a task. This also illustrates the importance of choosing an appropriate data view to fit the task need. To support a variety of tasks, an auditory interface may need to provide multiple data views to choose from.
- (3) A quick passive gist presentation allows pattern recognition. It gives the user an important first impression of the data that guides further exploration. This observation partially validated the utility of AISAs in the ADC framework.
- (4) Interactivity is a key element in auditory interface design. Interaction increases pattern recognition accuracy, and also brings out the benefit of other interface techniques such as the use of sound stereo position.
- (5) It is hard to tell the vertical location of non-individual HRTF spatial sound, at least for novice users. Designers shall not solely rely on the position cue of such sound to convey vertical location information.
- (6) Using an extra pitch to encode the vertical position of each geographical region is unhelpful at best. To improve vertical location awareness, we need to investigate

other means, such as to use better virtual spatial sound techniques with individual HRTF, or to provide more position cues in the navigation structure.

- (7) Sound stereo position helped pattern recognition and location awareness, but did not seem to be sufficient with the current keyboard navigation methods. To improve localization on maps and other data views in which data item positions are important, we need to try other means, such as to give position cues through other navigation methods.
- (8) Training on geography and sweep order seems to benefit pattern recognition. For unfamiliar maps, tactile maps can assist such training. It is also important to enable users to acquire the geographical knowledge through effective auditory map designs.
- (7) It is important to provide sound feedback on each user action.
- (9) Users benefited from “good” strategies. Teaching such strategies could be a part of the training.

4.2 iSonic Design Components

Based on lessons learned from those early explorations, I carried out more design iterations with regular feedback from the two blind users. The current iSonic design incorporates the effective design elements identified and addresses the weaknesses and needs revealed in the early design explorations.

4.2.1 Data Views

Early design explorations showed the importance of using the appropriate data view that fits the task needs. iSonic provides three highly coordinated data views – a region-by-

variable table, a choropleth map, and a bivariate scatterplot. This chapter focuses on the table and the map views (Figure 4-10). The scatterplot view was done as an extension after all user evaluations. It is discussed in Chapter 7.

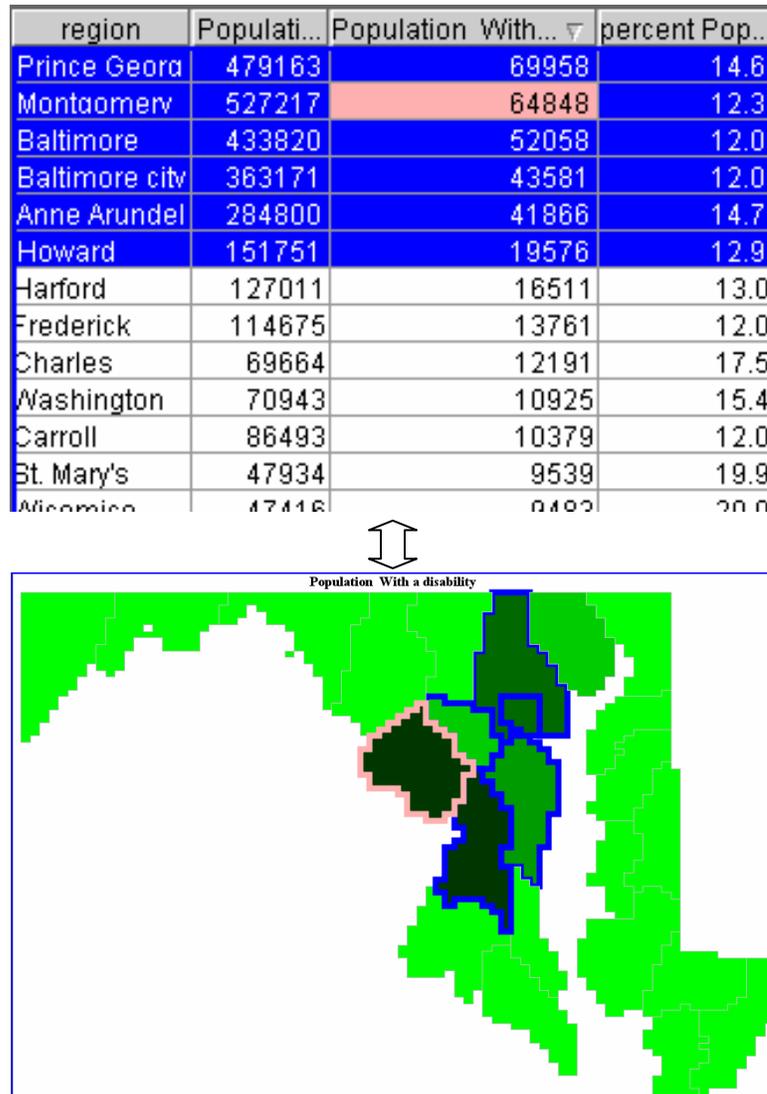


Figure 4-10: Highly coordinated table and map views in iSonic. The result of any action in one view is always reflected in another. The screenshots show a sample data set for Maryland counties. The counties in the table view are in a descending order by “Population with a disability”. In both views, selected counties are highlighted in blue, and the county the user is currently at is highlighted in pink.

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 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 strongly preferred a map over a table for discovering geographical value trends and performed better on pattern recognition tasks with a choropleth map than with a geographical knowledge enhanced table [194] (also see Section 4.1.2 and Section 4.1.4). Other views, such as scatterplots or line graphs, can be helpful for some analytical tasks. They were introduced after the iterative design process and evaluation because I wanted to first examine how users could operate multiple coordinated auditory views. Auditory and visual displays are synchronized to allow easy communication between sighted and blind users.

4.2.2 Input Devices and Navigation Structures

The earlier design explorations revealed the needs to improve users’ location awareness and to support map knowledge acquisition through means other than the auditory feedback encoding. I believe such needs can be addressed by choosing the appropriate input devices and designing effective navigation structures. When choosing input devices, I 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 or 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.

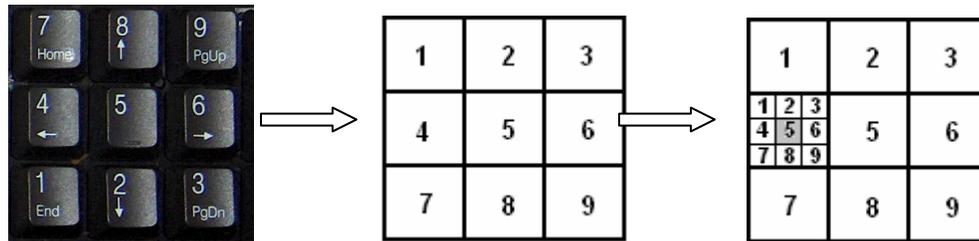


Figure 4-11 : Recursively map the 3x3 layout of the number pad to a 2-D space.

iSonic works with a keyboard alone. A keyboard is available on most computers and blind users are very comfortable using it. I use the arrow keys as natural means for relative movements in the left, right, up, and down directions. As mentioned in Section 3.2.3, 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, the 2-D display space is recursively mapped to the 3x3 layout of the number pad (Figure 4-11).

iSonic works with a touchpad or touch screen. Touchpad and touch screens are relatively common. A 14" touch screen 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 (e.g., as in [96]), but I chose not to rely on them because they need to be changed when the map changes and tactile

printers are expensive and rarely available. A mid-point solution is to use a generic grid with subtle tactile dots instead as a position and direction aid.

4.2.3 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. 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. iSonic can be configured to either generate stereo MIDI sounds or virtual spatial sounds by communicating to an external virtual spatial sound server. With MIDI sounds, the stereo panning effects indicate a region's azimuth position on the virtual auditory map. With the external virtual spatial sound server, region locations are mapped to Head Related Transfer Function (HRTF) virtual spatial sound locations. The left/right sound location is also used in the table to indicate the column order, with a left column sounds more to the left. 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. The use of HRTF virtual spatial sounds can harm the universal usability of the system, so I also explored other methods to present region vertical positions when only stereo MIDI sounds are available. I tried to play a piano pitch after

each violin value pitch to indicate the region's vertical position. Unfortunately, such extra sounds were found to be unhelpful at best according to the earlier design exploration.

4.3 iSonic Auditory Information Seeking Actions (AISA)

The earlier explorative designs and evaluations confirmed the importance of interactivity in auditory interfaces and validated the utility of gist, navigate, and details-on-demand.

To support general data exploration tasks besides geographical pattern recognition, iSonic also supports other 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. The two views are highly coordinated, so the result of any action in one view is always reflected in the other, including but not limited to navigation, zoom, and select.

4.3.1 Gist



Figure 4-12: Two spatial sweep orders on a U.S. state map: (a) horizontal sweep (b) and vertical sweep.

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 (green lines in Figure 4-13 (b) or (c)). 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 (Figure 4-12 (a)). Vertical sweeping is also an option (Figure 4-12 (b)). Users can press the ‘*’ key on the number pad to switch between the two sweeping orders. Generally, horizontal sweeping can take advantage of the stereo panning effect better, and is used by default. But which sweep order is more effective for the geographical data distribution pattern recognition can also depend on the map pattern (see Section 4.1.3.2). When the end of a 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 details in the gist, thus controlling its duration. For example, when the information level is set to “musical sound only”, a sweep of the entire U.S. state map containing 51 regions lasts for 9 seconds.

A choropleth map is not a perfect grid, so it is difficult to define the spatial sweep orders. The spatial sweep orders can be predefined manually and fed to iSonic as a part of the input data. The two manually defined orders shown in Figure 4-12 try to balance the number of regions in a sweep row/column, the number of sweep rows/columns, and how close it reflects how a user would visually scan the map. When no manually defined sweep orders exist in the input data, iSonic automatically generates them (see Section 5.4 for details).

4.3.2 Navigate

Table navigation is done by using arrow keys to move up, down, left, right, and to top, bottom, left and right edges. Users press ‘u’ to switch between two navigation modes. In the cell mode, the target cell is played (Figure 4-13 (a)). In the row/column mode, a sub-gist of the whole target row or column is played, depending on the direction of the movement. As illustrated in Figure 4-13 (b), when users move from one column to another (shown as blue arrow lines), the whole target column is swept from top to bottom (shown as green dotted arrow lines). When the user moves from one row to another, the whole target row is swept from left to right (Figure 4-13 (c)).

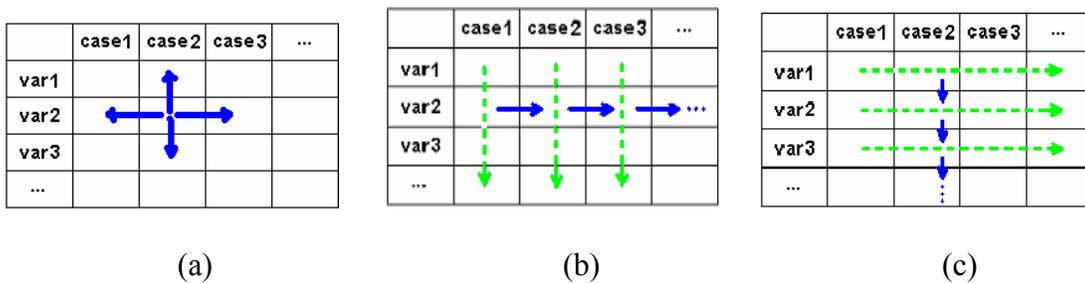


Figure 4-13: Two table navigation modes: (a) in the cell mode, users press arrow keys to move left, right, up, down (shown in blue arrow lines), or to left edge, right edge, top edge, and bottom edge; (b) in the row/column mode, when users move from one column to another (blue arrow lines), the whole target column is swept from top to bottom (green dotted lines); (c) in the row/column mode, when users move from one row to another (blue arrow lines), the whole target row is swept from left to bottom (green dotted lines).

While it is easy to navigate a 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 all the studies (see Chapter 6) 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. This is especially true when the adjacency matrix has to be automatically generated for any map (see Chapter 5.5 for details on how it is done in iSonic). 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.

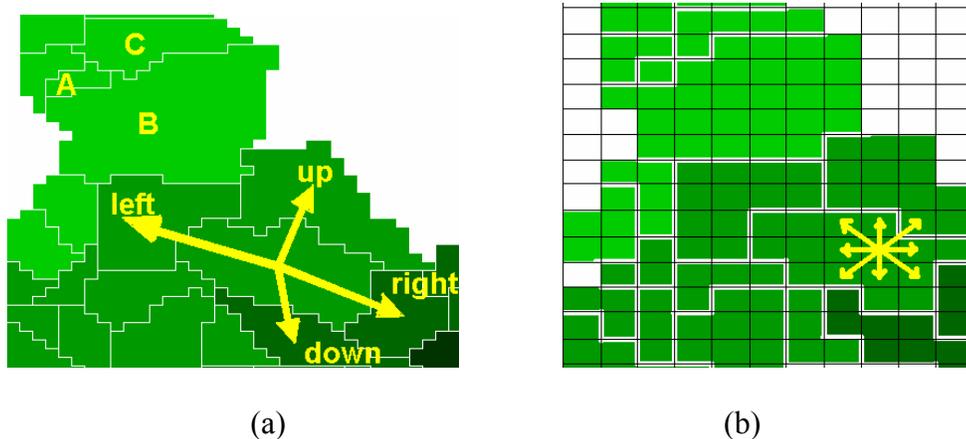


Figure 4-14: Two map relative navigation methods: (a) region-by-region navigation may have a reversibility problem: moving up from B usually goes to C. But if users start from A, move down to B, then move up, they will end up in C although users may expect to go back to A; (b) using a cell-by-cell relative navigation method on a mosaic version of the map, users may be able to sense the region size and shape, to avoid the reversibility problem, and possibly to have better location awareness.

For example, Figure 4-14 (a) shows three regions A, B, and C. Moving up from B usually goes to C. But if users start from A, move down to B, then move up, they will end up in C although users may expect to go back to A. To tackle some of the problems and to convey regions sizes and shapes, a cell-by-cell movements on a mosaic version of the map was tested (Figure 4-14 (b)). But no evidence was obtained that it improved users' location awareness. Instead, it was much less preferred because it required more keystrokes to move around [195] (also see Section 4.1.3).

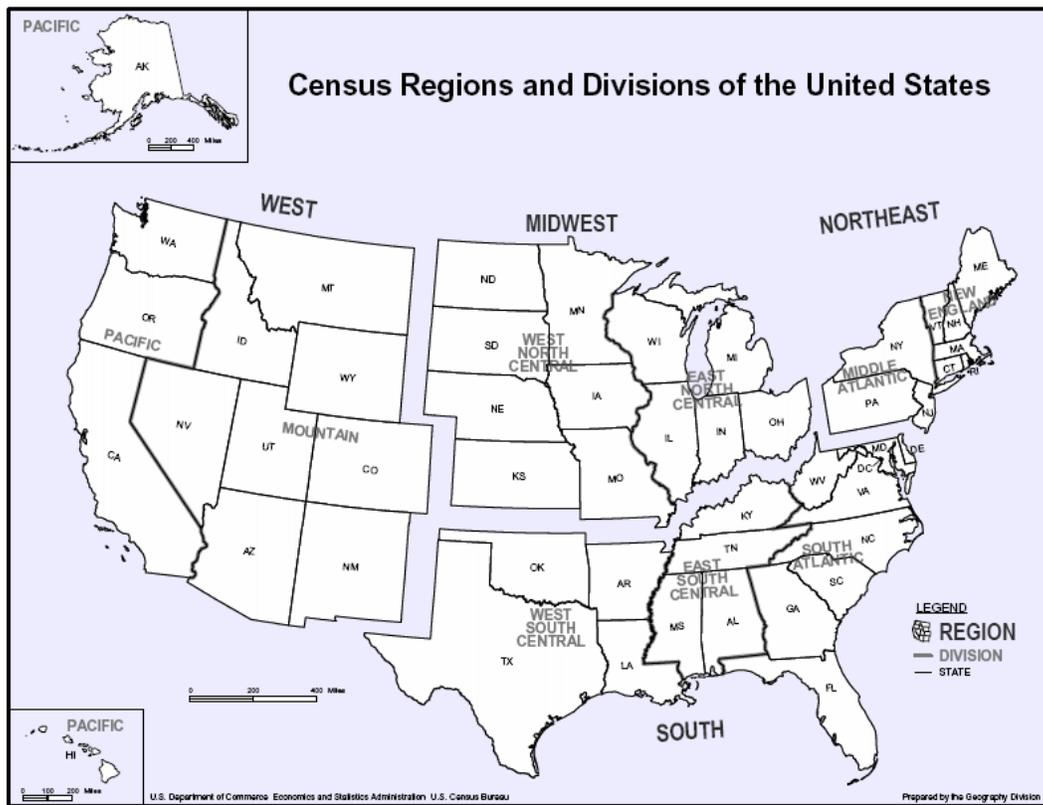


Figure 4-15: Census regions and divisions of U.S.

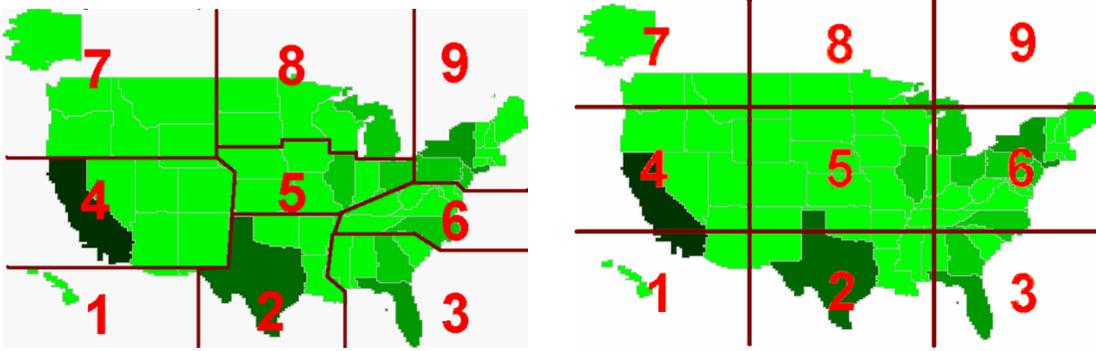


Figure 4-16: Two types of top level 3x3 partitions of a U.S. state map: (a) a manual partition based on U.S. Census Regions in Figure 4-15; (b) a partition automatically generated by iSonic according to regions' geometric center coordinates.

To improve absolute position awareness, iSonic supports absolute navigation. The map is divided into 3x3 ranges (Figure 4-16) and users use the 3x3 number keys on a 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. Predefined manual partitions can be fed to iSonic. For example, U.S. states are typically divided into four census regions by the U.S. Census Bureau: northeast, south, midwest, and west (Figure 4-15). We can further divide these census regions to get a 3x3 partition to be mapped to the 3x3 layout of the number pad. Figure 4-16 (a) shows a manual partition of U.S. state map that is consistent with the census tradition and reflects the spatial 3x3 layout. When no manual partitions are defined, iSonic automatically generates the partition according to regions' geometric center coordinates, such as the one shown in Figure 4-16 (b). Studies showed that blind users can realize the map spatial layout by listening to each range (see Chapter 6).

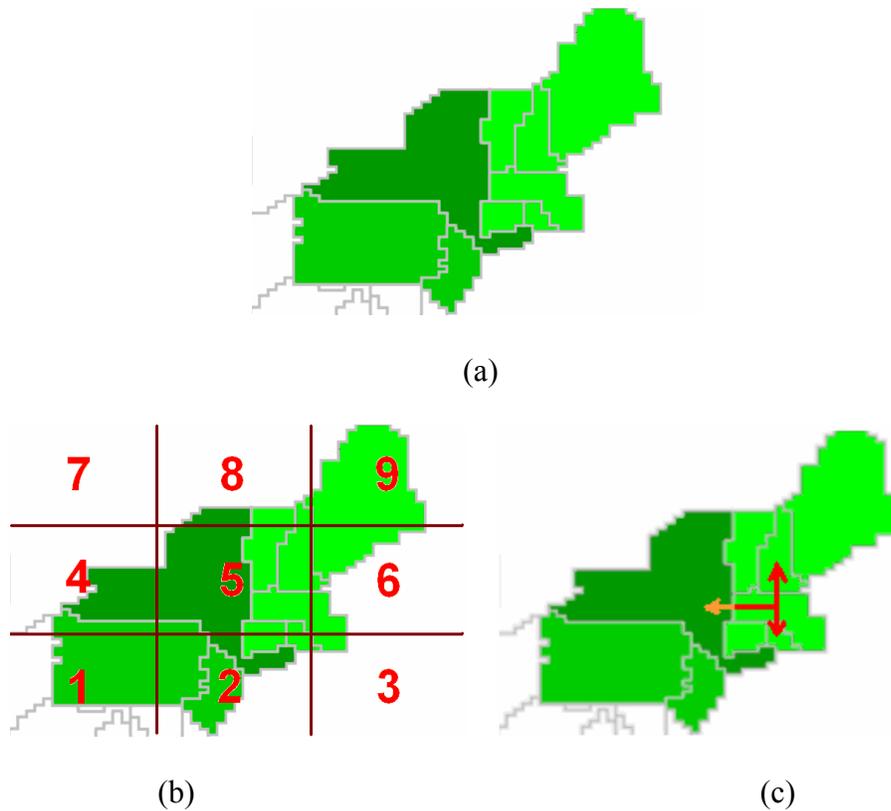


Figure 4-17 : Zooming based on a fixed map partition: (a) press Ctrl + 9 to zoom into range 9 of the map partition in Figure 4-16 (a). States outside of range 9 are grayed out and disabled; (b) recursively explore range 9 according to an automatic 3x3 partition; (c) use arrow keys to move among adjacent states in range 9.

Users can zoom into a part of the map for further exploration. There are two types of zooming. In zooming based on a fixed 3x3 partition (either a manual or automatic partition), users can use Ctrl+[number] to zoom into any of the 9 ranges, within which they can recursively explore using the 3x3 pattern or use arrow keys to move around. For example, Figure 4-17 (a) shows the display after the user presses Ctrl+9 to zoom into range 9 of the manual map partition in Figure 4-16 (a). States outside of range 9 are

grayed out and disabled in the zoomed view. Figure 4-17 (b) shows recursive 3x3 exploration of range 9 according to an automatic partition. For 3x3 exploration at any zoom level, iSonic checks whether a manual partition is defined for that level, and uses an automatic partition if it does not exist. By using recursive 3x3 exploration, users could get a sense of the spatial layout of the map (see Chapter 6). Pressing ‘0’ sweeps the current zoomed map range or the whole map if no zooming has been performed.

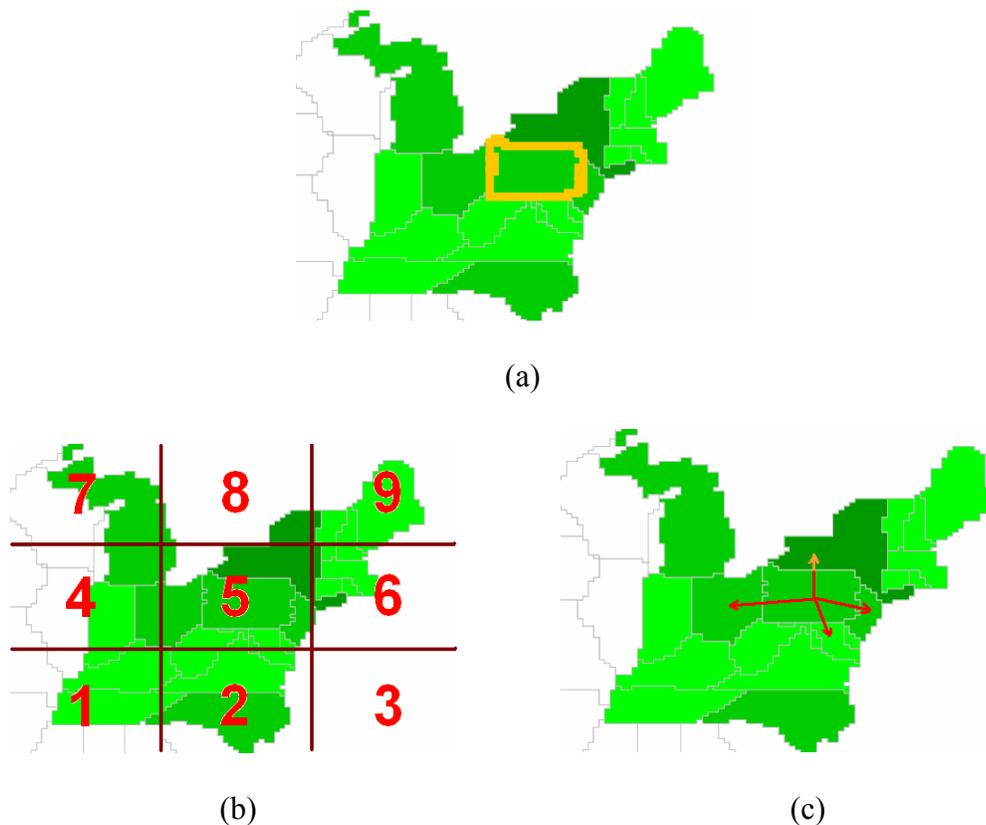


Figure 4-18: Zoom into the neighborhood of the specific region: (a) zoom into the neighborhood of Pennsylvania (highlighted in orange) from the map in Figure 4-16(a). States whose geometric center fall out of the area are grayed out and disabled; (b) recursively explore the neighborhood in the 3x3 style according to an automatic 3x3 partition; (c) use arrow keys to move among adjacent states in the neighborhood.

Zooming based on a fixed partition may separate adjacent regions, making it hard to explore the neighborhood of a particular region. To solve the problem, iSonic users can instead press “Enter” to zoom into the neighborhood of the current region. The neighborhood is an area centered at that region’s geometric center and has a size of 1/9 of the previous zoom level. All regions whose geometric centers fall out of the area are grayed out and disabled at the zoomed view. For example, Figure 4-18 (a) shows the neighborhood of Pennsylvania. Within that neighborhood, users can recursively explore in the 3x3 style (Figure 4-18 (b)), or zoom further into a neighborhood, or use arrow keys to move around (Figure 4-18 (c)).

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 (Figure 4-19). Stereo or virtual spatial 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. When users zoom into a range, such as range 9, touchpad exploration will focus on only that range with a resolution nine times higher than the previous zoom level (Figure 4-19 (b)). User observations suggested that both the keyboard and touchpad navigations allow users to gain geographical knowledge (see Chapter 6).

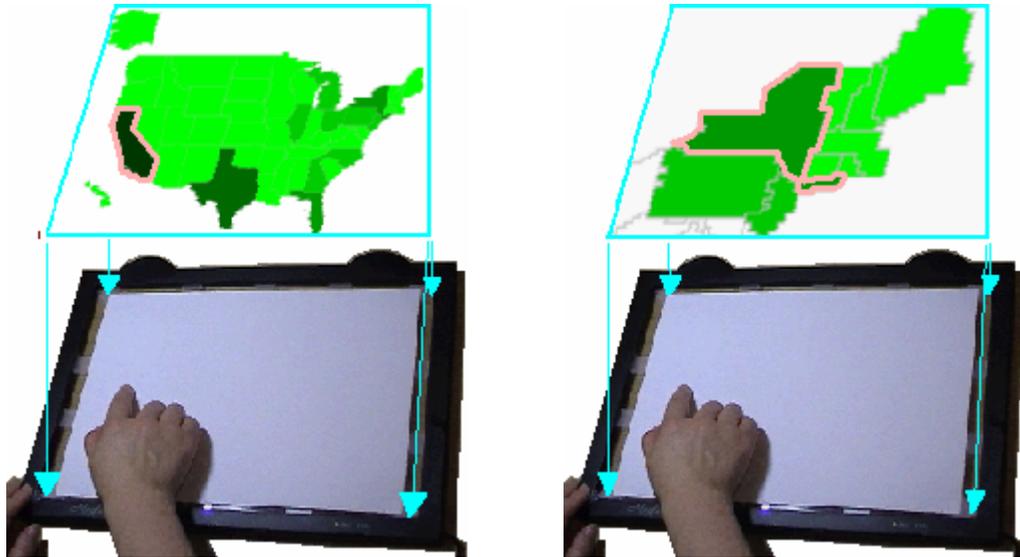


Figure 4-19: Touchpad exploration: (a) the touchpad is calibrated so that the map is mapped to its entire surface. The region under the finger position (highlighted with a bold pink border) is activated; (b) when users zoom into range 9, the touchpad is recalibrated so the current map range (range 9) is mapped to its entire surface, providing higher exploration resolution. The region under the finger position (highlighted with a bold pink border) is activated.

As mentioned before, the map and the table views are highly coordinated, so when users zoom into a range on the map and switch to the table view, only regions in that map range are displayed in the table view. In the example in Figure 4-20, users zoom into range 9 of the U.S. state map in the map view. The state highlighted with a bold pink border is the last state visited. After switching to the table view, only states in the zoomed range are shown in the table, and the last state visited in the map view becomes the current state in the table view (highlighted with a pink cell background).

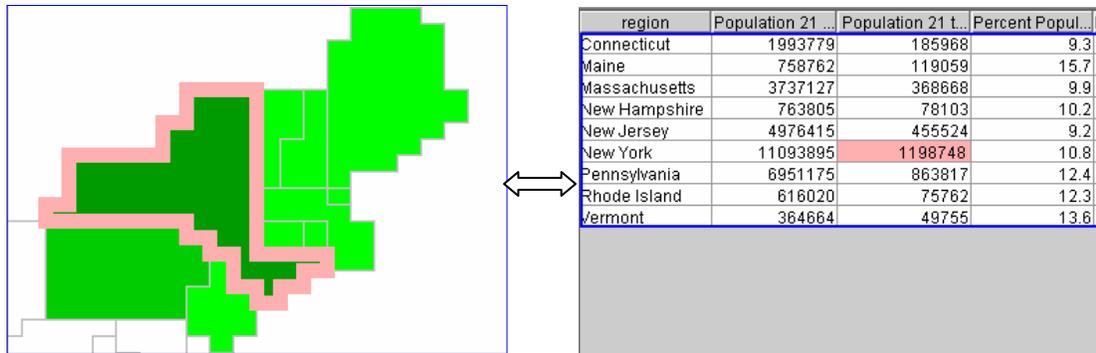


Figure 4-20: Coordination between views during zooming: (a) in the map view, zoom into range 9 of a U.S. state map. The state highlighted with a bold pink border is the last state visited; (b) after switching to the table view, only states in the zoomed range are shown in the table, and the last state visited in the map view becomes the current state in the table view.

4.3.3 Details-on-demand and Adjust Information Level

In both map and table views, pressing ‘space’ plays details of the current region. Another way to get details is to increase the information level to the maximum level in which details of a region are given by default when that region is activate.

4.3.4 Situate and Search

When users press ‘I’ (as for ‘Information’), iSonic speaks the current interface 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.

Searching by region name is obviously helpful but is not implemented yet.

4.3.5 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. The two views are coordinated during selection too. When a region is selected in one view, it is automatically selected in another. Selected regions are highlighted with blue borders by default (Figure 4-21).

When users are in “selected regions only” in one view and switch to another view, they are in “selected regions only” in that view.

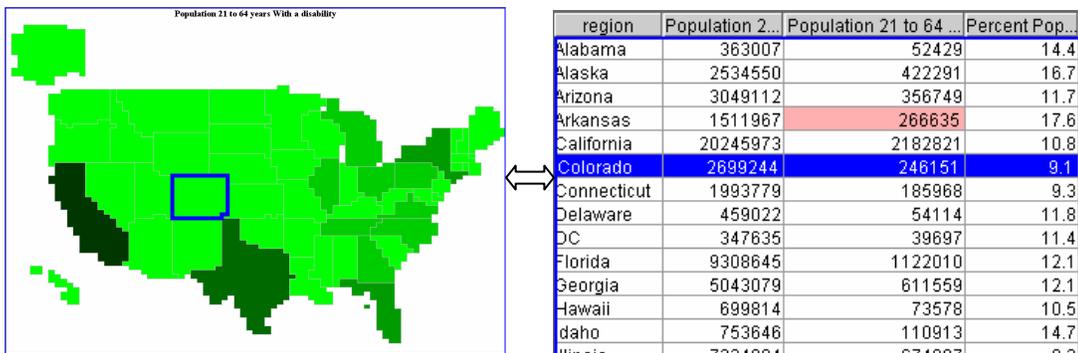


Figure 4-21: Coordination between views for selection: any regions selected (highlighted with blue borders) in the table view are always selected in the map view and vice versa.

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

map.

4.3.7 Filter and Search

Filtering is done by adjusting the query range through two-thumb sliders. It is logically the same as the slider-based filtering in the map-based geo-referenced data visualization [124] but all controls are done through keyboards instead of mice. The design challenges are to provide an intuitive set of keyboard commands to support two-thumb slider operations, to provide appropriate auditory feedback to convey the filtering results, as well as the query being applied. In visualization, slider-based dynamic query provides rapid, revertible, and continuous query with the goal to continuously update the visual display in less than 100 milliseconds [20]. In the auditory mode, such a short time is usually not enough to present a gist of changes. In iSonic, results are presented discretely, with a gist of the changes played after each query range adjustment.

Users can perform filtering anytime by pressing ‘f’ to launch the filter control panel (Figure 4-22). The filter control panel contains multiple sliders, with each slider representing one statistical variable in the data set. All slider operations can be activated through the associated menu system or pressing hotkeys that are also explained in the menu system. For example, users use the left/right/home/end keys to move to the previous/next/first/last slider accordingly. The current slider is highlighted in blue. When a new slider becomes the current slider, or users press ‘space’ to request information about the current slider, iSonic speaks the current slider name, then plays four percussion tones to present the minimum value, current lower query bound, current upper query bound, and maximum value. The four tones are played in sequence, with the length of the

pause between two adjacent tones representing the proportional distance of the two corresponding values on the slider. After the four tones, iSonic speaks the query range, and the minimum and maximum values of the slider.



Figure 4-22: The filter control panel contains multiple two-thumb sliders, with each slider representing one variable. The current slider (the third slider) is highlighted in blue.

Users use the 'a' key to cycle among "left thumb", "right thumb", and "both thumbs" of the current slider as the active thumb(s). Pressing the up/down arrow keys moves the active thumb(s) and changes the query range of the current slider. When the active thumb is "left thumb", pressing the up arrow key increases the lower bound of the query range, potentially causing some regions to be filtered out. Pressing the down arrow

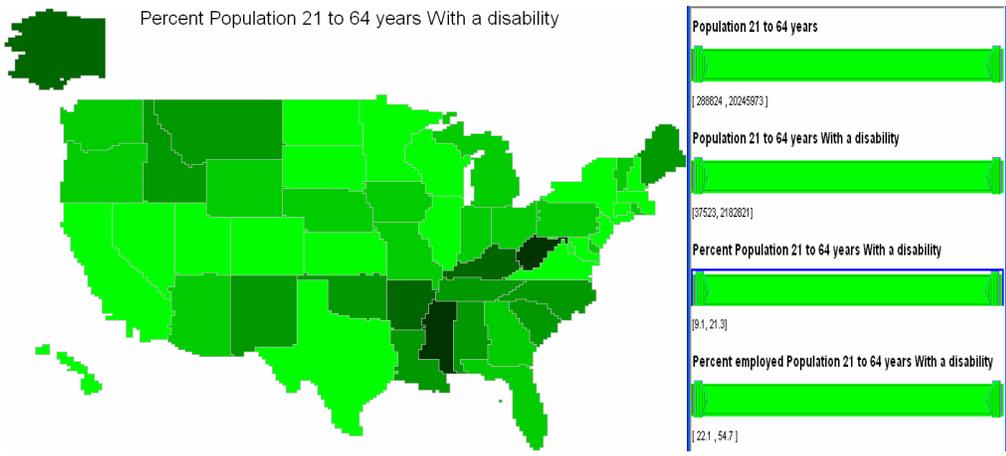
key decreases the lower bound of the query range, reverting the filtering. When the active thumb is the “right thumb”, pressing the down arrow key decreases the upper bound, potentially causing some regions to be filtered out. Pressing the up arrow key increases the upper bound, reverting the filtering. When the current thumb is “both thumbs”, pressing the up arrow key increases both the upper and lower bounds, and pressing the down arrow key decreases them. As the result, some regions become filtered out and some become unfiltered. Regions that are filtered out are grayed out in both the map (e.g., Figure 4-23 (d)) and table views (e.g., Figure 4-24 (b)). The gray shade of the filtered region in the map view reflects the original color shade of that region.

Users use the ‘u’ key to switch between two types of filtering units: by distribution and by value unit. When filtering by distribution, each arrow keystroke causes regions with the same value (usually one region) to change the filter status. It only applies when the left or the right thumb is being adjusted. When filtering by value unit, each arrow keystroke causes the corresponding query bound or both bounds to increase/decrease by a preset value unit (two by default).

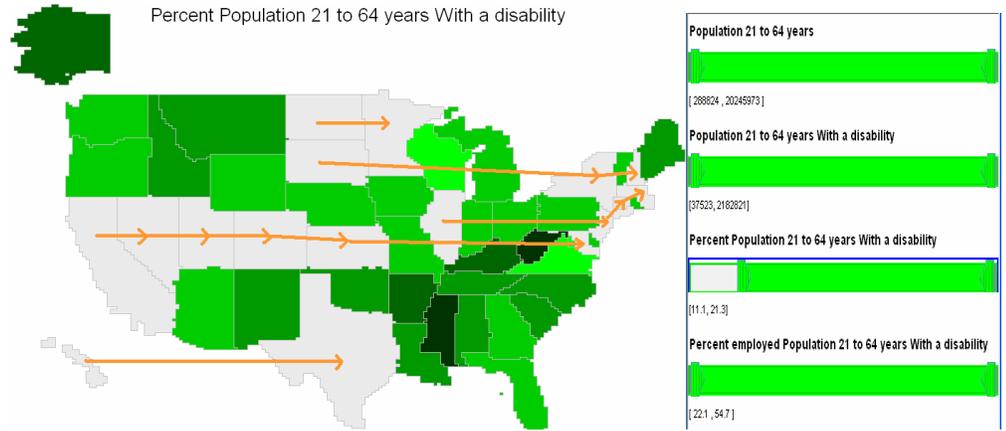
For each adjustment action, iSonic changes the visual display of the affected regions (from green to gray if being filtered out or from gray to a green shade if being unfiltered). Meanwhile, iSonic speaks the number of affected regions then plays a gist of those regions. Region sequencing in the gist depends on the view from which the filter control has been launched and is consistent with the gist of other AISAs. If it is the map view, the gist is a spatial sweep on the map. If it is the table view, the gist is a sweep by the regions’ order in the table (column sweep from top to bottom). Two timbres distinguish filtered regions from the remaining. By default, a violin pitch represents an

unfiltered region, and a piano pitch represents a filtered region.

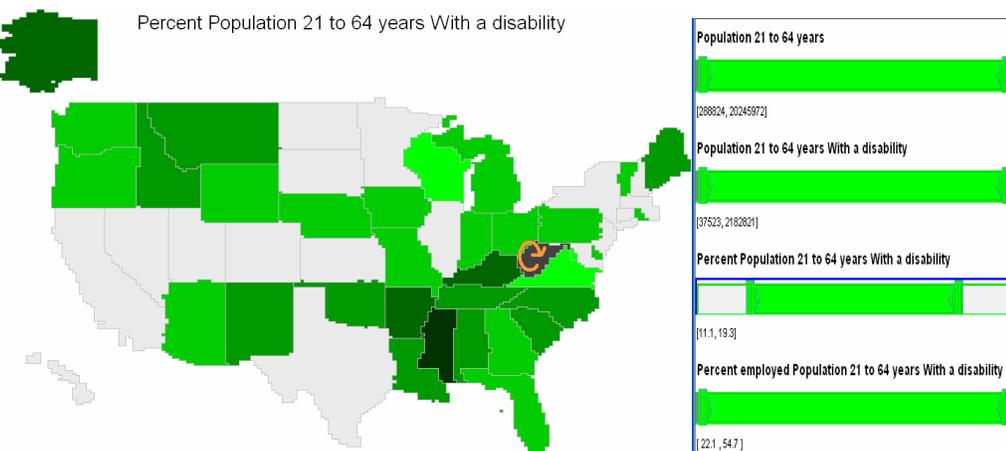
Figure 4-23 illustrates a sequence of filtering. The user launches the filtering panel from the map view and goes to the slider “percent population 21 to 64 years with a disability”. The map is showing the same variable. No filtering is applied originally (Figure 4-23 (a)). By default, the left thumb is the active thumb. By pressing the up arrow key once, the user increases the lower thumb by two units (default value unit is set to two), thus changing the query range of the slider variable from [9.1%, 21.3%] to [11.1%, 21.3%]. As the result, 17 states whose values for the slider variable are not in the query range become filtered out and turn gray. At the same time, a spatial sweep of the 17 states (illustrated by orange arrow lines) is played (Figure 4-23 (b)). By pressing the ‘a’ key once, the right thumb becomes the active thumb. The user presses the down arrow key once to further filter out states with high values for the slider variable. One more state becomes filtered out and turns gray. A spatial sweep containing only this state is played (Figure 4-23 (c)). By pressing the ‘a’ key again, both thumbs become active. The user presses the up arrow key once to move the whole query range up by two units. As the result, 15 states become newly filtered out due to the increment of the lower query bound, and 1 state becomes unfiltered and returns to the original green shade due to the increment of the upper query bound (Figure 4-23 (d)). iSonic first plays the 15 newly filtered states (illustrated by orange arrow lines) then the one unfiltered state (illustrated by the blue circular arrow line). The level of details about each state in the gist depends on what information level the user has chosen before launching the filter control.



(a)



(b)



(c)

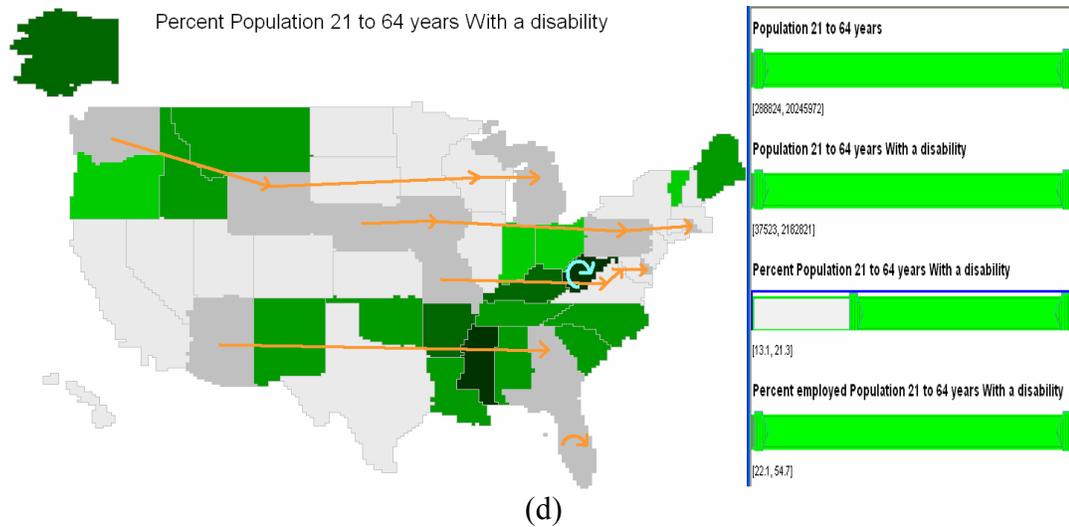
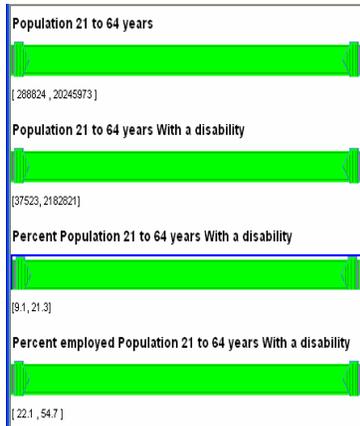


Figure 4-23: An example of filtering sequence applied to the map view: (a) no filtering; (b) increase the lower bound by two units. 17 states are filtered out and turn gray. A spatial sweep of the 17 states (illustrated in orange arrow lines) is played; (c) decrease the upper bound by two units. 1 more state becomes filtered. A spatial sweep containing only this state is played; (d) move the whole query range up by two units. 1 state becomes unfiltered and returns to the original green shade. 15 new states become filtered and turn gray. iSonic first plays the newly filtered states (illustrated in orange) then the unfiltered states (illustrated in a blue circular line).

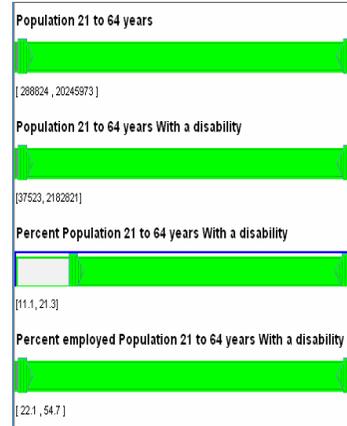
Figure 4-24 illustrates the results of the same filtering sequence if the filtering panel is launched from the table view. The table is ordered ascending by the slider variable so the top table records are filtered out when the left thumb increases and the bottom table records are filtered out when the right thumb decreases.

region	Population ...	Population 21...	Percent ...	Percent emp...
Colorado	2699244	246151	9.1	48.3
Minnesota	2921129	267524	9.2	48.8
New Jersey	4976415	455524	9.2	29.6
Connecticut	1993779	185968	9.3	22.7
Illinois	7234804	674997	9.3	39.2
South Dakota	409763	37905	9.3	54.7
Utah	1249122	120419	9.6	49.3
Massachusetts	3737127	368668	9.9	26.7
New Hampshire	763805	78103	10.2	29.4
Hawaii	699814	73578	10.5	27.7
Nevada	1301666	137276	10.5	40.8
Maryland	3193664	341909	10.7	22.1
California	20245973	2182821	10.8	36.4
New York	11093895	1198748	10.8	35.1
North Dakota	349000	38075	10.9	48.5
Kansas	1515346	166730	11.0	42.2
Texas	12356358	1361880	11.0	39.7
Wisconsin	3123585	351938	11.3	43.9
DC	347635	39697	11.4	23.1
Virginia	4250549	485563	11.4	29.9
Arizona	3049112	356749	11.7	38.9
Delaware	459022	54114	11.8	23.3
Florida	9308645	1122010	12.1	27.6
Georgia	6042070	641550	10.6	24.0
Montana	1000000	200000	10.0	20.0
Maine	758762	119059	15.7	22.3
Alaska	2534550	422291	16.7	27.5
Arkansas	1511967	266635	17.6	39.8
Kentucky	2370784	426972	18.0	35.6
Mississippi	1581751	306058	19.3	39.4
West Virginia	1046335	222594	21.3	24.9

region	Population ...	Population 21...	Percent ...	Percent emp...
Colorado	2699244	246151	9.1	48.3
Minnesota	2921129	267524	9.2	48.8
New Jersey	4976415	455524	9.2	29.6
Connecticut	1993779	185968	9.3	22.7
Illinois	7234804	674997	9.3	39.2
South Dakota	409763	37905	9.3	54.7
Utah	1249122	120419	9.6	49.3
Massachusetts	3737127	368668	9.9	26.7
New Hampshire	763805	78103	10.2	29.4
Hawaii	699814	73578	10.5	27.7
Nevada	1301666	137276	10.5	40.8
Maryland	3193664	341909	10.7	22.1
California	20245973	2182821	10.8	36.4
New York	11093895	1198748	10.8	35.1
North Dakota	349000	38075	10.9	48.5
Kansas	1515346	166730	11.0	42.2
Texas	12356358	1361880	11.0	39.7
Wisconsin	3123585	351938	11.3	43.9
DC	347635	39697	11.4	23.1
Virginia	4250549	485563	11.4	29.9
Arizona	3049112	356749	11.7	38.9
Delaware	459022	54114	11.8	23.3
Florida	9308645	1122010	12.1	27.6
Georgia	6042070	641550	10.6	24.0
Montana	1000000	200000	10.0	20.0
Maine	758762	119059	15.7	22.3
Alaska	2534550	422291	16.7	27.5
Arkansas	1511967	266635	17.6	39.8
Kentucky	2370784	426972	18.0	35.6
Mississippi	1581751	306058	19.3	39.4
West Virginia	1046335	222594	21.3	24.9



(a)



(b)

region	Population ...	Population 21...	Percent...	Percent emp...
Colorado	2699244	246151	9.1	48.3
Minnesota	2921129	267524	9.2	48.8
New Jersey	4976415	455524	9.2	29.6
Connecticut	1993779	185968	9.3	22.7
Illinois	7234804	674997	9.3	39.2
South Dakota	409763	37905	9.3	54.7
Utah	1249122	120419	9.6	49.3
Massachusetts	3737127	368668	9.9	26.7
New Hampshire	763805	78103	10.2	29.4
Hawaii	699814	73578	10.5	27.7
Nevada	1301666	137276	10.5	40.8
Maryland	3193664	341909	10.7	22.1
California	20245973	2182821	10.8	36.4
New York	11093895	1198748	10.8	35.1
North Dakota	349000	38075	10.9	48.5
Kansas	1515346	166730	11.0	42.2
Texas	12356358	1361880	11.0	39.7
Wisconsin	3123585	351938	11.3	43.9
DC	347635	39697	11.4	23.1
Virginia	4250549	485563	11.4	29.9
Arizona	3049112	356749	11.7	38.9
Delaware	459022	54114	11.8	23.3
Florida	9308645	1122010	12.1	27.6
Ukraine	1905993	230604	13.3	40.4
Maine	758762	119059	15.7	22.3
Alaska	2534550	422291	16.7	27.5
Arkansas	1511967	266635	17.6	39.8
Kentucky	2370784	426972	18.0	35.6
Mississippi	1581751	306058	19.3	39.4
West Virginia	1046335	222594	21.3	24.9

region	Population ...	Population 21...	Percent...	Percent emp...
Colorado	2699244	246151	9.1	48.3
Minnesota	2921129	267524	9.2	48.8
New Jersey	4976415	455524	9.2	29.6
Connecticut	1993779	185968	9.3	22.7
Illinois	7234804	674997	9.3	39.2
South Dakota	409763	37905	9.3	54.7
Utah	1249122	120419	9.6	49.3
Massachusetts	3737127	368668	9.9	26.7
New Hampshire	763805	78103	10.2	29.4
Hawaii	699814	73578	10.5	27.7
Nevada	1301666	137276	10.5	40.8
Maryland	3193664	341909	10.7	22.1
California	20245973	2182821	10.8	36.4
New York	11093895	1198748	10.8	35.1
North Dakota	349000	38075	10.9	48.5
Kansas	1515346	166730	11.0	42.2
Texas	12356358	1361880	11.0	39.7
Wisconsin	3123585	351938	11.3	43.9
DC	347635	39697	11.4	23.1
Virginia	4250549	485563	11.4	29.9
Arizona	3049112	356749	11.7	38.9
Delaware	459022	54114	11.8	23.3
Florida	9308645	1122010	12.1	27.6
Georgia	6043079	611569	12.1	34.8
Iowa	1649586	203502	12.3	48.2
Rhode Island	816020	75762	12.3	26.8
Nebraska	969114	120212	12.4	48.4
Pennsylvania	6951175	863817	12.4	35.6
Missouri	3218268	401766	12.5	38.3
Michigan	5734026	722693	12.6	36.4
Washington	3585533	457957	12.8	40.4
Wyoming	288824	37523	13.0	52.3
Oregon	2082399	274957	13.2	38.7
Indiana	3483326	463402	13.3	37.6
Ohio	6169000	867677	14.2	27.6
Ukraine	1905993	230604	13.3	40.4
Maine	758762	119059	15.7	22.3
Alaska	2534550	422291	16.7	27.5
Arkansas	1511967	266635	17.6	39.8
Kentucky	2370784	426972	18.0	35.6
Mississippi	1581751	306058	19.3	39.4
West Virginia	1046335	222594	21.3	24.9



(c)



(d)

Figure 4-24: Results for the same filtering sequence applied to the table view: (a) no filtering; (b) increase the lower bound by two units. The top 17 states are filtered out. A sweep of the 17 states (illustrated in arrow lines) is played; (c) decrease the upper bound by two units. The bottom state becomes filtered out and played; (d) move the whole query range up by two units. The bottom state becomes unfiltered. 15 new states become filtered. iSonic first plays the 15 newly filtered states then the unfiltered bottom state.

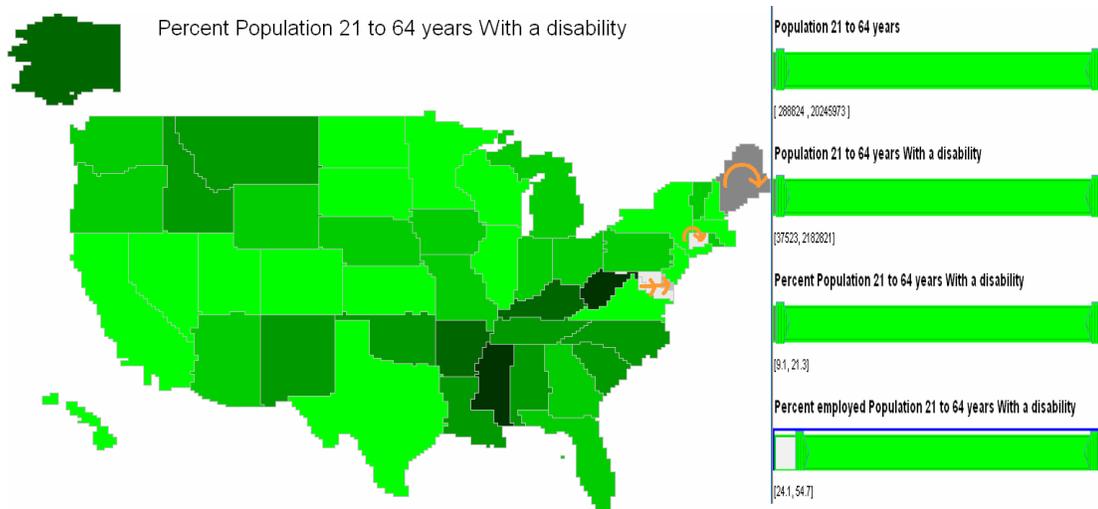


Figure 4-25: Filter by a variable different from the one displayed on the map.

The variable played in the filtering gist is the variable shown on the map if it is a map view, or the current table variable if it is a table view. Filtering can be done on variables other than the one displayed and played. For example, in Figure 4-25, users adjust the left thumb of the fourth slider “percent employed population 21 to 64 years with a disability”. But the variable conveyed in the gist is “percent population 21 to 64 years with a disability” because it is the variable displayed on the map. In slider-based dynamic queries coupled with visual displays, users can easily perceive from such filtering the underlying relation (such as correlation) between the displayed variable and the manipulated variable [124]. It will be interesting to find out if it also works in the auditory mode.

By adjusting multiple sliders, users can produce conjunctive queries. Users can request the current filter result, such as the number of regions unfiltered, by pressing the ‘s’ key (stand for ‘Situat’). The user can press the ‘r’ key to reset the query range of the

current slider or Ctrl + r to reset all sliders.

Filtered regions remain filtered after users close the filtering control panel and return to the map or table view. When a region is activated by other AISA actions, the timbre of its value pitch represents whether it is filtered or not. When performing other AISAs, users can switch between “filtered regions”, “unfiltered regions” and “all regions”, just like they are able to switch between “selected regions” and “all regions”. Combining the two status dimensions, a region can have 6 statuses. Currently iSonic allows users to choose among 6 region sets for focused exploration: “selected regions only”, “unfiltered regions only”, “filtered regions only”, “selected and unfiltered regions only” and “all regions”.

4.4 iSonic Menu, Hotkeys, and Online Help

Whenever the iSonic application window becomes active, a greeting message is played. It rapidly plays the musical sounds as a legend, and notifies the user how to activate the menu and the online help in speech.

4.4.1 Menu and Hotkeys

From the table view, the map view, or the filter control window, pressing F10 takes the user to the associated menu system. The speech prompts the user to use left/right arrow keys to choose and open a menu group, and use the up/down arrow keys to choose a menu item in the group. When the user moves among the menu groups and menu items, the speech reads the name of the menu group or menu item, the associated hotkey, and a brief message explaining what the menu item does. Pressing ‘enter’ or ‘space’ executes

the menu item. The menu systems for the table view, the map view, and the filter control window have different contents, containing only applicable menu items. Figure 4-26, Figure 4-27, and Figure 4-28 show the menu systems associated with the map view, the table view, and the filter control window accordingly.

All interface functions can be accessed both from the menu system and by directly pressing the hotkeys. Table 4-4 lists some hotkeys in the current iSonic implementation.

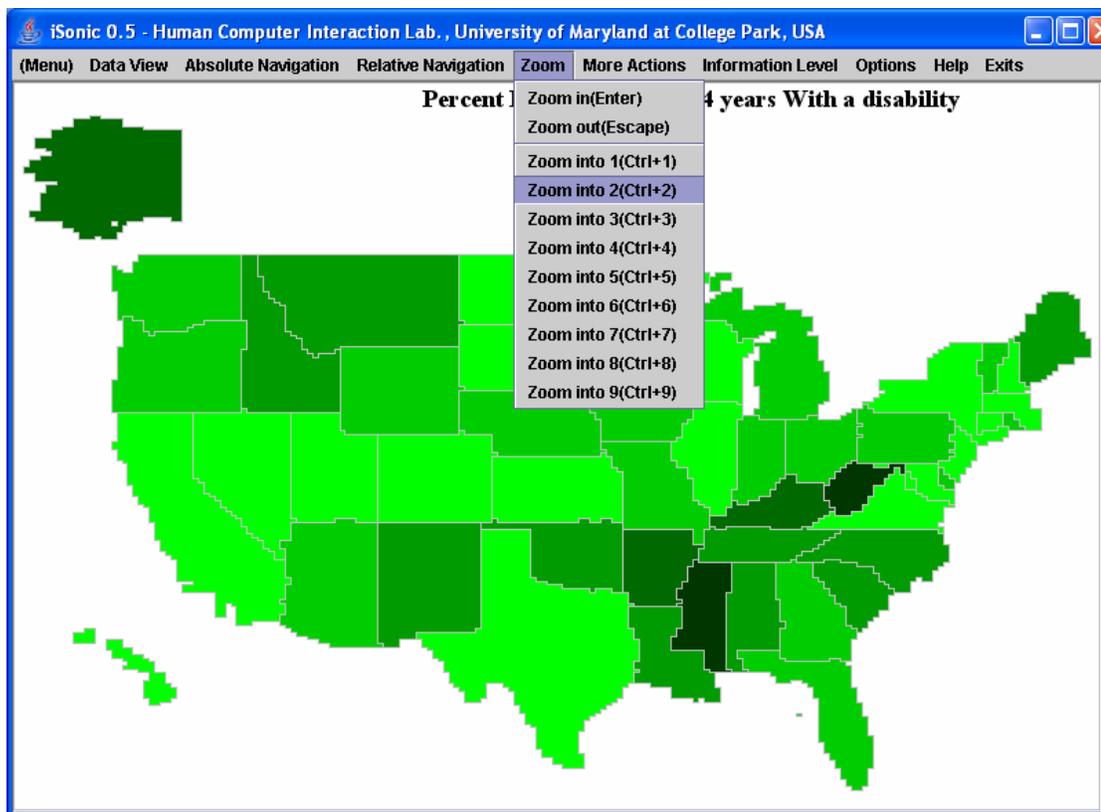


Figure 4-26: The menu system associated with the map view. The menu can be accessed by a keyboard and displayed both visually and auditorily.

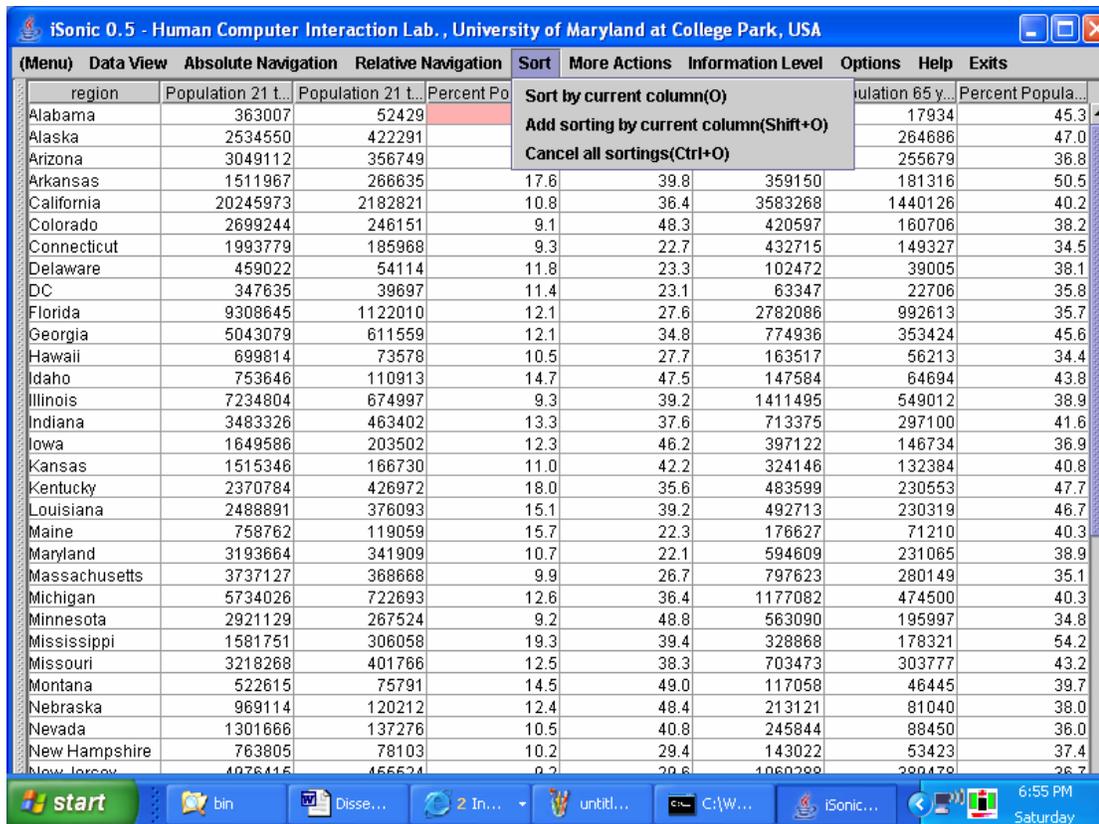


Figure 4-27: The menu system associated with the table view.

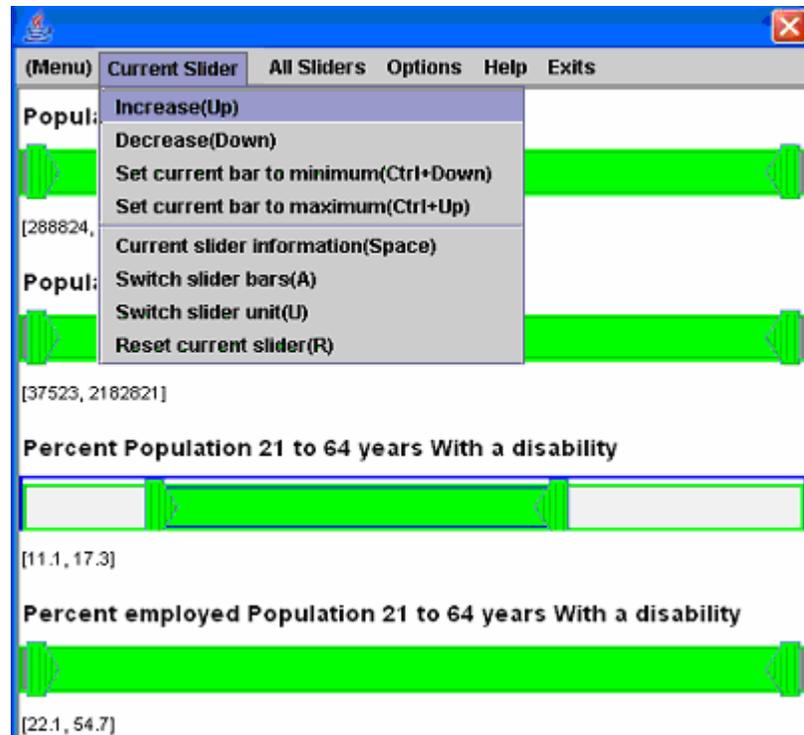


Figure 4-28: The menu system associated with the filter control window.

Hotkey	Function
	All
<l>	As in “Lock”, to toggle the select status of the current region
Ctrl+<l>	Unselect all
<f>	Launch filter control panel
<a>	Circle among region sets of different select and filter statuses
<i>	As in “Information”, to situate
<+>/<-> on num pad	Increase/decrease detail level
<space>	Details-on-demand
<F1>	Launch online help
<F10>	Go to menu
Alt+<F4>	Close current window
TAB	Switch between table and map views
	Map view
<0> on num pad	Play current range
<1-9> on num pad	Play range 1 - 9
<enter>	Zoom into neighborhood of current region
<escape>	Zoom out to previous level
Ctrl+<1-9 >	Zoom into range 1 - 9
<left>, <right>, <up>, <down>	Go to region on left, right, above, and below
<n>, <p>	Go to next/previous region along spatial sweep path
<*> on num pad	Switch among horizontal and vertical sweep orders
	Table view
<m>	Switch between cell and row/column navigation modes
<left>, <right>, <up>, <down>	Go to cell/row/column on left, right, above, and below
<home>, <end> Ctrl+<home>, Ctrl+<end>	Go to left, right, top, and bottom edges
<o>, Shift+<o>, Ctrl+<o>	Sort by current column only, further sort by current column, cancel all sorting
	Filter control
<left>, <right>, <home>, <end>	Go to next, previous, first, or last slider
<up>, <down>	Increase or decrease the active thumb(s)
Ctrl+<up>, Ctrl+<down>	Set the active thumb to minimum or maximum
<space>	Get current slider information
<a>	Switch active thumb
<u>	Switch between “by unit” and “by distribution”
<r> / Ctrl+<r>	Reset current slider / reset all sliders

Table 4-4: Some hotkeys in the iSonic implementation

4.4.2 Online Help

Pressing F1 activates the online help system. The online help consists of a collection of HTML documents. As users press the up/down arrow keys to move among the hyperlinks in the current document. The content of the hyperlink and the text following the hyperlink are spoken. The current hyperlink is highlighted in pink (Figure 4-29). Pressing the 'enter' key loads into the help window the new document linked from the current hyperlink. If the new document is a prerecorded wave sound file to demonstrate interface sounds, the sound file is played (Figure 4-30). Each time an HTML document is loaded into the help window, the first paragraph in the body of the document starts to be spoken. The first paragraph typically contains some instructions of how to navigate the documents.

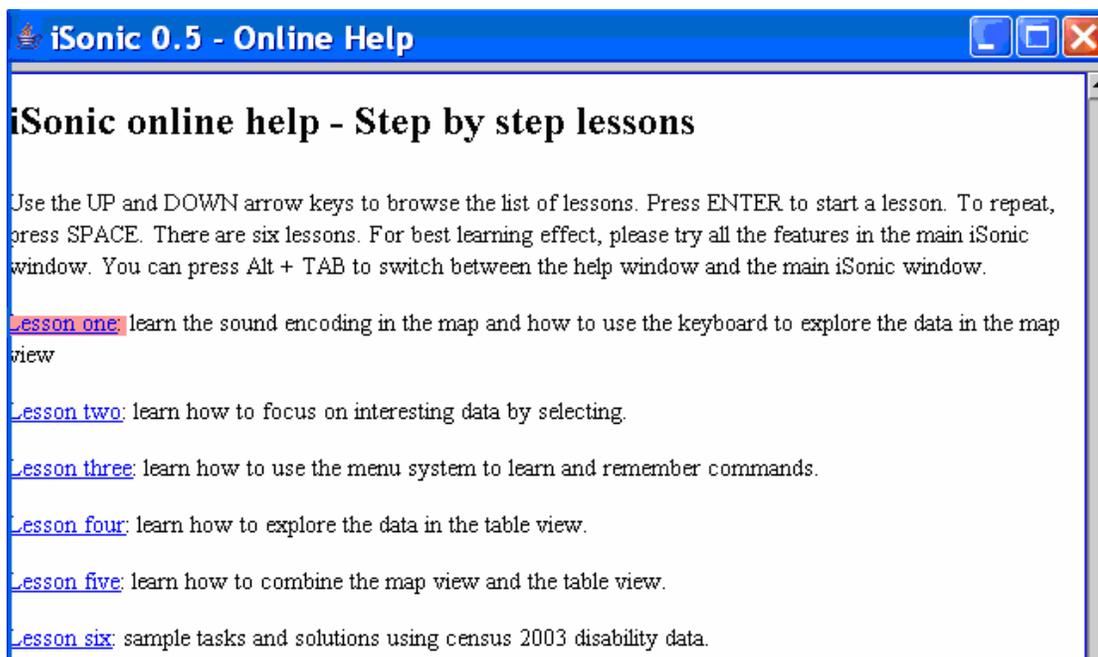


Figure 4-29: iSonic online help contains a set of HTML documents. Users use arrow keys to navigate the hyperlinks. The current hyperlink is highlighted in pink.

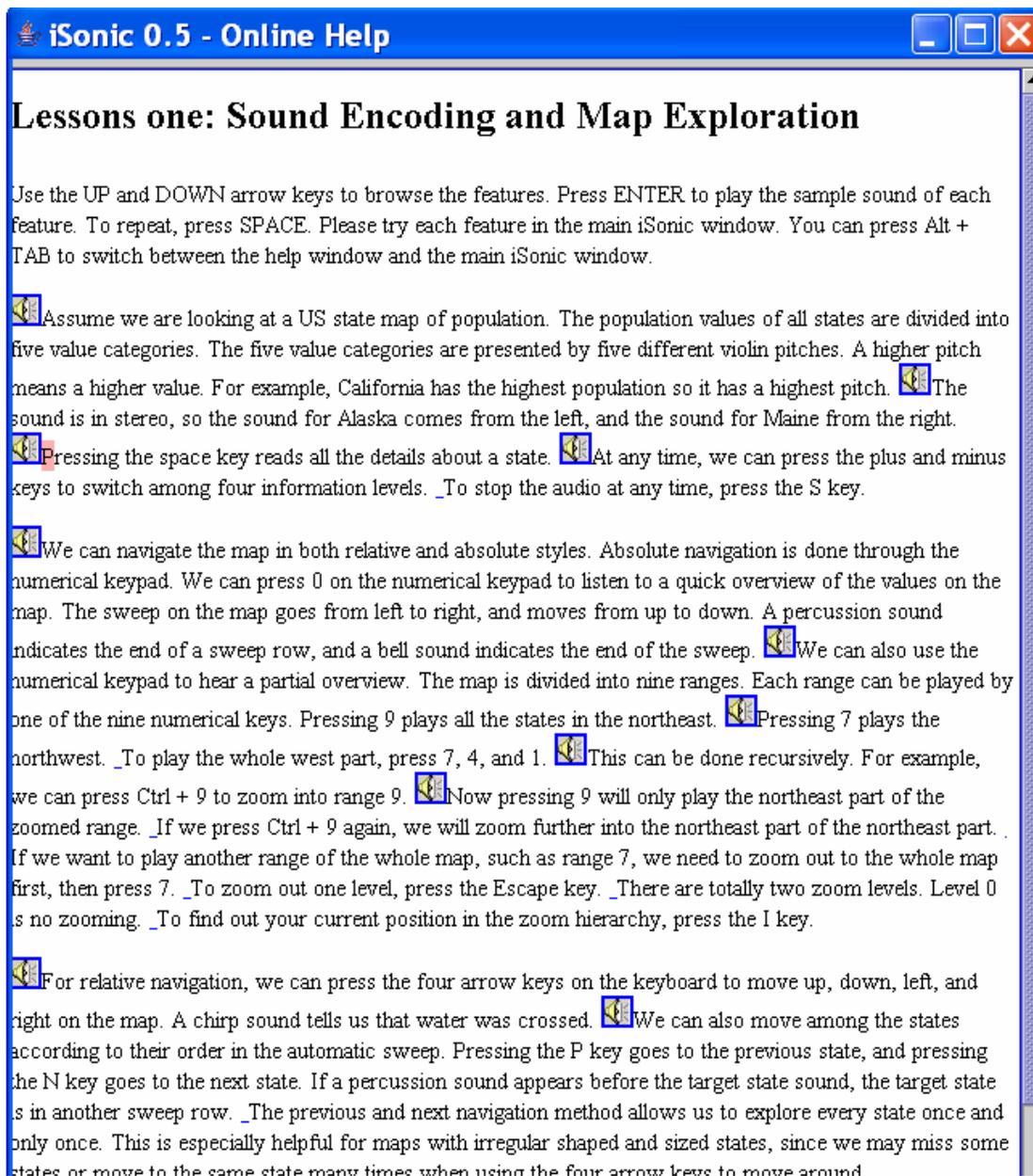


Figure 4-30: iSonic online help: if the current hyperlink points to a prerecorded wave sound file, pressing 'enter' plays the sound file.

4.5 A Sample Task Solution

Now I use a sample task to show how the auditory information seeking actions in iSonic

can be combined for problem solving.

Task: What and where are the top 5 states with the highest employment rate for people with a disability?

Step 1: <TAB> to switch to the table view. Use <+>/<-> to adjust the information level to 2 so that the column name and the region name are spoken during table navigation. The table navigation mode is cell mode by default. Use <left>/<right> to go to column “percent employed population 21 to 64 years with a disability” (Figure 4-31 (a)).

Step 2: Use <o> to sort the column descending. Ctrl + <home> to go to the first row (Figure 4-31 (b)). Then use <down> to move down one row at a time. The first 5 regions are the answers to the “what” part of the given task. To find out “where” the 5 regions are located on the map, continue to the next step.

Step 3: Use <l> to select the 5 states. The selected states are highlighted in blue (Figure 4-31 (c)).

Step 4: <TAB> to switch to the map view and the 5 selected states are also selected in the map view (Figure 4-31 (d)). Use <a> to switch to “selected regions only”.

Step 5: By pressing <1> - <9> to listen to the 9 map ranges according to the manual partition in Figure 4-16(a), it is clear that the top 5 states are in northwest part of the map (Figure 4-31 (e)). It is also clear that they are adjacent by using arrow key navigation.

region	Populati...	Population...	Percent P...	Percent employed Population 21 to 64 years With a disability	Po
Alabama	363007	52429	14.4		37.4
Alaska	2534550	422291	16.7		27.5
Arizona	3049112	356749	11.7		38.9
Arkansas	1511967	266635	17.6		39.8
California	20245973	2182821	10.8		36.4
Colorado	2699244	246151	9.1		48.3
Connecticut	1993779	185968	9.3		22.7
Delaware	459022	54114	11.8		23.3
DC	347635	39697	11.4		23.1
Florida	9308645	1122010	12.1		27.6
Georgia	5043079	611559	12.1		34.8
Hawaii	699814	73578	10.5		27.7
daho	753646	110913	14.7		47.5
Illinois	7234804	674997	9.3		39.2
Indiana	3483326	463402	13.3		37.6
owa	1649586	203502	12.3		46.2
Kansas	1515346	166730	11.0		42.2
Kentucky	2370784	426972	18.0		35.6
Louisiana	2488891	376093	15.1		39.2
Maine	758762	119059	15.7		22.3
Maryland	3193664	341909	10.7		22.1
Massachusetts	3737127	368668	9.9		26.7
Michigan	5734026	722693	12.6		36.4
Minnesota	2921129	267524	9.2		48.8
Mississippi	1581751	306058	19.3		39.4
Missouri	3218268	401766	12.5		38.3
Montana	522615	75791	14.5		49.0
Nebraska	969114	120212	12.4		48.4
Nevada	1301666	137276	10.5		40.8

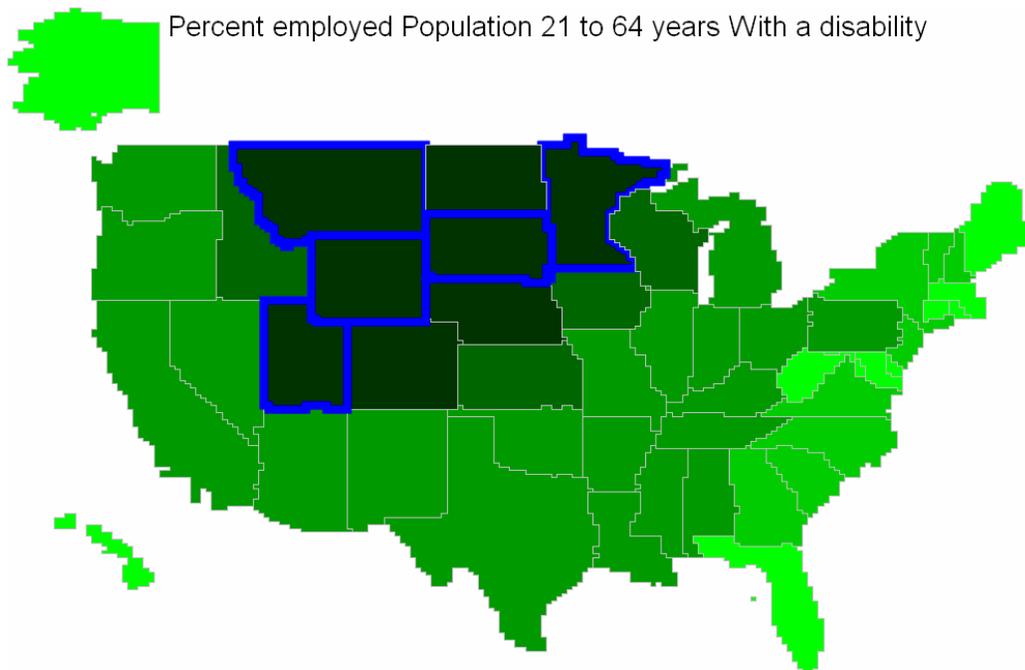
(a)

region	Populati...	Population...	Percent P...	Percent employed Population 21 to 64 years With a disability	Po
South Dakota	409763	37905	9.3		54.7
Wyoming	288824	37523	13.0		52.3
Utah	1249122	120419	9.6		49.3
Montana	522615	75791	14.5		49.0
Minnesota	2921129	267524	9.2		48.8
North Dakota	349000	38075	10.9		48.5
Nebraska	969114	120212	12.4		48.4
Colorado	2699244	246151	9.1		48.3
daho	753646	110913	14.7		47.5
owa	1649586	203502	12.3		46.2
Wisconsin	3123585	351938	11.3		43.9
Kansas	1515346	166730	11.0		42.2
Nevada	1301666	137276	10.5		40.8
Oklahoma	1933993	296682	15.3		40.4
Washington	3585533	457957	12.8		40.4
Arkansas	1511967	266635	17.6		39.8
Texas	12356358	1361880	11.0		39.7
Mississippi	1581751	306058	19.3		39.4
Illinois	7234804	674997	9.3		39.2
Louisiana	2488891	376093	15.1		39.2
Arizona	3049112	356749	11.7		38.9
Oregon	2082399	274957	13.2		38.7
Missouri	3218268	401766	12.5		38.3
Indiana	3483326	463402	13.3		37.6
New Mexico	1032807	149712	14.5		37.6
Ohio	6486863	862677	13.3		37.6
Alabama	363007	52429	14.4		37.4
California	20245973	2182821	10.8		36.4
Michigan	5734026	722693	12.6		36.4

(b)

region	Populati...	Population...	Percent P...	Percent employed Population 21 to 64 years With a disability	Po
South Dakota	409763	37905	9.3	54.7	
Wyoming	288824	37523	13.0	52.3	
Utah	1249122	120419	9.6	49.3	
Montana	522615	75791	14.5	49.0	
Minnesota	2921129	267524	9.2	48.8	
North Dakota	349000	38075	10.9	48.5	
Nebraska	969114	120212	12.4	48.4	
Colorado	2699244	246151	9.1	48.3	
Idaho	753646	110913	14.7	47.5	
Iowa	1649586	203502	12.3	46.2	
Wisconsin	3123585	351938	11.3	43.9	
Kansas	1515346	166730	11.0	42.2	
Nevada	1301666	137276	10.5	40.8	
Oklahoma	1933993	296682	15.3	40.4	
Washington	3585533	457957	12.8	40.4	
Arkansas	1511967	266635	17.6	39.8	
Texas	12356358	1361880	11.0	39.7	
Mississippi	1581751	306058	19.3	39.4	
Illinois	7234804	674997	9.3	39.2	
Louisiana	2488891	376093	15.1	39.2	
Arizona	3049112	356749	11.7	38.9	
Oregon	2082399	274957	13.2	38.7	
Missouri	3218268	401766	12.5	38.3	
Indiana	3483326	463402	13.3	37.6	
New Mexico	1032807	149712	14.5	37.6	
Ohio	6486863	862677	13.3	37.6	
Alabama	363007	52429	14.4	37.4	
California	20245973	2182821	10.8	36.4	
Michigan	5724828	722893	12.6	36.4	

(c)



(d)

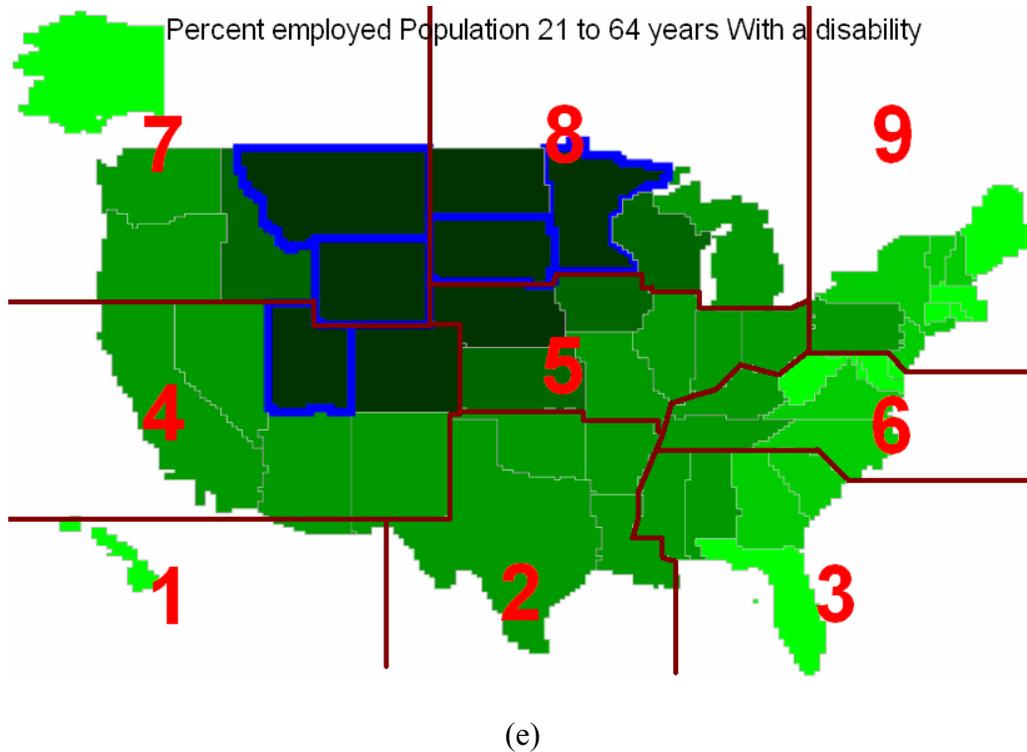


Figure 4-31: Steps to solve the task “what and where are the top 5 states with the highest employment rate for people with a disability”.

4.6 Discussion

Numerous options exist for designing an interactive auditory interface for geo-referenced data analysis. The ADC framework was used to guide the earlier explorative designs and evaluations to compare various design options along the dimensions of data views, auditory feedback encoding, and navigation structures. The evaluations also provided insights into human ability to perceive geographical data distribution patterns through interactive sounds, and partially confirmed the utilities of the framework. The current iSonic design is based on findings and lessons from those explorations as well as feedback from two blind users. However, many potential improvements could be done.

When the number of geographical regions increases, map spatial sweep and navigation become increasingly complex. For example, in a U.S. county map containing 3140 geographical regions, it is necessary to use data aggregation. Multiple aggregation methods exist. It could be done based on value clusters by grouping regions with similar values, or based on predefined region groups by showing a state level map and bringing up a county level map for each state upon zooming into that state.

While pitch is generally the best way to present quantitative data (see section 3.2.4.1), the range and number of easily discernable pitches are limited for most human. Musical sounds, such as those produced by a MIDI synthesizer, have a more complex spectrum than pure sine waves, making them easier to perceive, and more pleasant to listen to, than sine waves. While the range of available MIDI notes is from 0-127, it is suggested that on a standard PC soundcard only notes in the range 35-100 should be used for easy perception and differentiation [33]. To map a wide range of values to musical pitches, the values often need to be categorized and each category is mapped to a unique pitch. This hides the difference of values in the same category which can be revealed by other means such as speech. To utilize pitch more, the value to pitch mapping can be rescaled upon zooming, i.e., when users zoom to focus on a subset of data items, the new narrower value range can be remapped to the whole available pitch range instead of the original sub-range, possibly revealing more value differences. Of course, the mapping rescaling scheme can cause misinterpretation and disorientation, if no appropriate interface hints are provided upon rescaling.

Reversibility in arrow key navigation is an inherent problem for maps with irregularly shaped and sized regions. It could potentially be alleviated by grouping small

regions into a big virtual region and navigating the virtual regions.

iSonic is fully operable using a standard keyboard. A touchpad allows continuous exploration of the current map range, but it needs to be used together with a keyboard to issue other AISA commands. By designating a part of the touchpad surface to a virtual on-screen keyboard (see Figure 3-3) and some tactile prints on the touchpad to mark the zones, iSonic could be operated with a touchpad alone.

Chapter 5

iSonic Implementation

iSonic was implemented as a Java application that could also run from the Web through Java Web Start Technology [170]. The GUI part was written in Java JFC/Swing and the musical sounds were typically produced through the Java MIDI Sound technique. Speech was produced by sending command network datagram to an accompanying speech server built on Microsoft Speech SDK 5.1. iSonic can also produce virtual spatial sounds by sending command network datagram to a spatial sound server that simulates real world sounds using Head Related Transfer Functions (HRTF). The spatial sound server was developed at the Perceptive Interfaces and Reality Laboratory, University of Maryland at College Park. Additionally, iSonic can be configured to load and play prerecorded 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 usually less than 100 milliseconds. Readers can try iSonic or viewing the video demonstration by visiting www.cs.umd.edu/hcil/iSonic.

5.1 Architecture Overview

The architecture of iSonic implementation was based on the Model-View-Controller (MVC) paradigm. Originally introduced in the SmallTalk-80 interface from Xerox PARC [141], MVC is commonly used as a design framework for modern GUI interfaces, such as Macintosh, Microsoft Windows, and programming platforms such as Java. In the

MVC paradigm, the user input, the modeling of the external world, and the visual feedback to the user are explicitly separated and handled by three types of object [36]. The **view** manages the graphical and/or textual output to the portion of the bitmapped display that is allocated to its application. The **controller** interprets the mouse and keyboard inputs from the user, commanding the model and/or the view to change as appropriate. Finally, the **model** manages the behavior and data of the application domain, responds to requests for information about its state (usually from the view), and responds to instructions to change state (usually from the controller).

iSonic has both a GUI display and an auditory display. The implementations of the two displays are modularly separated into two parallel subviews (a visual view and an auditory view) residing inside each data relation view (e.g., map, table, and scatterplot). A simplified diagram in Figure 5-1 shows the core classes in iSonic and their interactions. The diagram only shows classes for two data relation views (map and table). An extension to the diagram to integrate the scatterplot view is given in Section 7.3. Dotted rectangles are not actual classes but to logically combine some classes in the diagram. Classes of the type *model* are gray and those of the type *controller* are green. The remaining are *views*, organizational classes, or miscellaneous classes for producing sounds.

DataSet and *InteractionParameters* are two model classes shared in the whole application. *DataSet* reads the original region-by-variable input data file and stores the data items in a 2-D data array. It also keeps the status of each geographical region, such as whether the region is filtered, selected, or in the current zoomed range.

InteractionParameters is a class that keeps track of interface operational status such as

which region and variable the user is currently focusing on, whether the user is examining all regions or selected regions only, and so on.

One Data Set - Multiple Views

For any given *DataSet*, there are three data relation views in iSonic – a table view, a map view, and a scatterplot view, each has its own model, controller, and subviews, wrapped in an organizational class. *MMMap* organizes all map specific MVC classes, and *MMTable* organizes all table specific MVC classes. Scatterplot classes are shown in the diagram extension in Figure 7-9. There can also be some non-traditional data relation views in iSonic, such as the filter view, organized by *FilterControls*. The filter view shows a two-thumb slider for each variable in the *DataSet*, with the thumb positions indicating some regions being filtered out.

Coordination and Synchronization

In iSonic, all the data relation views are highly coordinated by monitoring changes in *DataSet* and *InteractionParameters*, and adjusting their own graphical displays (through classes named ‘Viz*’) and auditory displays (through classes named ‘Aud*’).

The visual and auditory views are synchronized at a highly refined granularity. Classes in charge of auditory feedback (*AudMap*, *AudTable*, *AudFilterCtrls*) produce sounds by calculating *SoundingObjects* and sending *SoundMessages* to one or more *SoundServers* (*MidiServer*, *SpatialSoundServer*, *SpeechServer*, *AudioFileServer*). Controlled by the *Timer*, each *SoundMessage* is executed after a specified delay from the moment it was submitted. The execution of each *SoundMessage* usually generates a *SoundMessageExecutionEvent*. Classes in charge of graphical rendering (*VizMap*,

VizTable, *VizFilterCtrls*) listen to *SoundMessageExecutionEvents* and update the visual display upon qualified sound events.

Map

The map data relation view is a combination of a visual map view (*VizMap*) and an auditory map view (*AudMap*). *VizMap* is responsible for all map related graphical rendering. It renders animated map zooming by making use of Piccolo [19], a toolkit written in Java that supports the development of 2-D zoomable graphical applications. *AudMap* is in charge of producing auditory feedback. The synchronization among *VizMap* and *AudMap* is done via the firing and monitoring of *SoundMessageExecutionEvents*.

InputDeviceToActionMap is a controller class that listens to the keyboard and mouse input received by *VizMap*, and activates corresponding *MapActions* (note: *MapActions* is not a single class but stands for a collection of action classes for the map view). Each map action class performs some designated logical operations according to *DataSet* / *InteractionParameters* / *MapInfo*, making changes to *DataSet* / *InteractionParameters*, and requests *VizMap* and *AudMap* to generate visual and auditory feedback. The *MapInfo* class contains all map specific data, such as the vector representation (*PGeoRegion*) of the geographical shape of each region, map spatial sweep orders (*SweepLists*, either read from the input data file or automatically generated), and region adjacency information (*AdjacencyLists*).

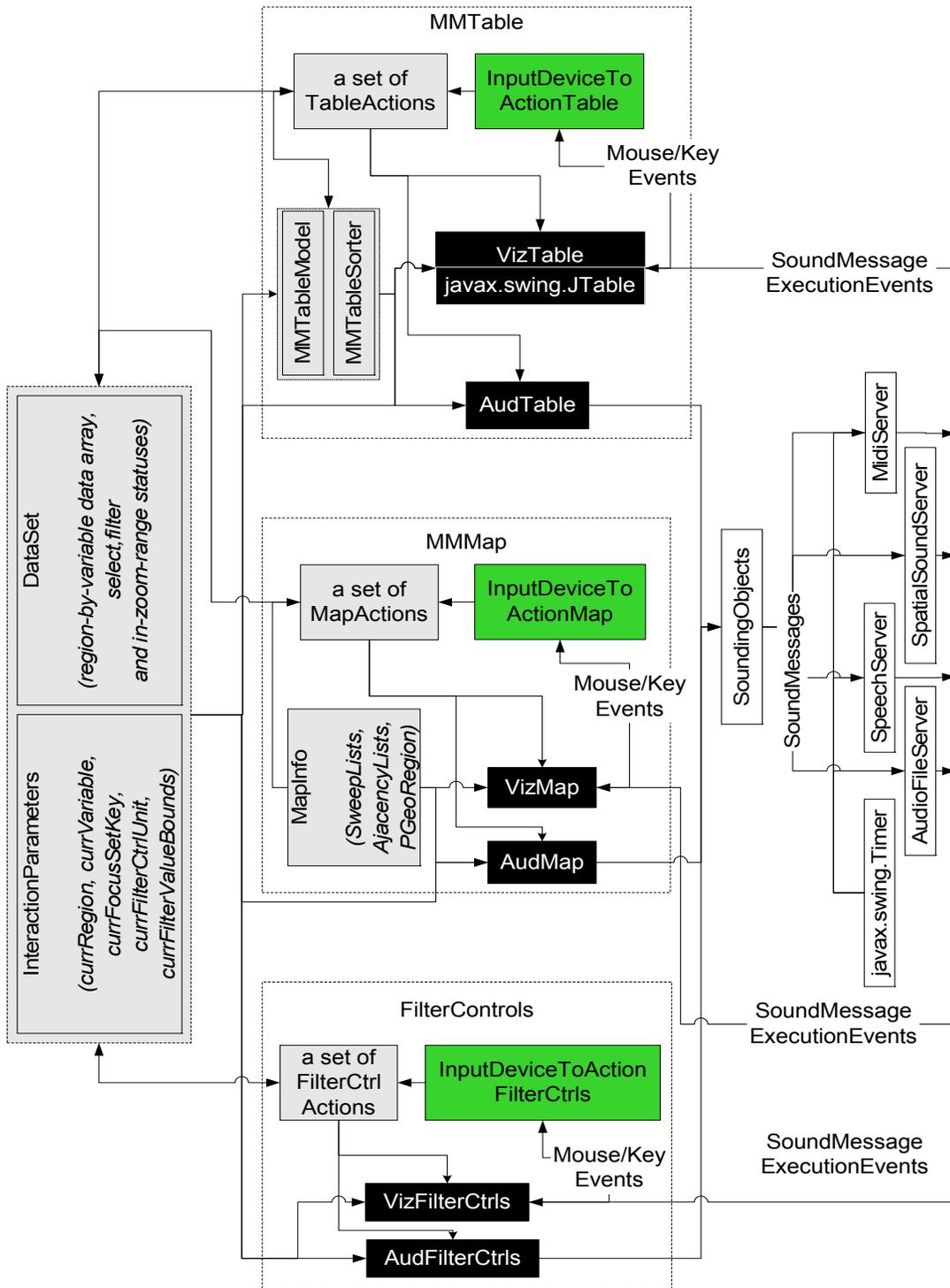


Figure 5-1: A simplified diagram showing the core classes in iSonic and their interactions.

Table

Similar to the map view, the table data relation view is a combination of a visual table view (*VizTable*) and an auditory table view (*AudTable*). *VizTable* is responsible for all table related graphical rendering. It renders a graphical table by using `javax.swing.JTable`. *AudTable* is in charge of producing auditory feedback. Similar to the map view, the synchronization among *VizTable* and *AudTable* is done via the firing and monitoring of *SoundMessageExecutionEvents*.

InputDeviceToActionTable is a controller class that listens to the keyboard and mouse input received by *VizTable*, and activates corresponding *TableActions* (note: *TableActions* is not a single class but stands for a collection of action classes for the table view). Each table action class performs some designated logical operations according to *DataSet / InteractionParameters / MMTTableModel / MMTTableSorter*, making changes to *DataSet / InteractionParameters / MMTTableSorter*, and requests *VizTable* and *AudTable* to generate visual and auditory feedback. *MMTableModel* and *MMTableSorter* are responsible for performing sorting transformations on the unsorted data from *DataSet* before it is rendered.

Filter Control

Users open a filter control panel (*FilterControls*) to perform filtering by manipulating a set of two-thumb sliders. The filter control panel is also a view of the data, usually working in parallel with either the map view or the table view. It is a combination of a visual filter control view (*VizFilterCtrls*) and an auditory filter control view (*AudFilterCtrls*). *VizFilterCtrls* is responsible for rendering a set of graphical two-thumb sliders. *AudFilterCtrls* is in charge of producing auditory feedback about slider

changes.

InputDeviceToActionFilterCtrls is a controller class that listens to the keyboard and mouse input received by *VizFilterCtrls*, and activates corresponding *FilterCtrlActions* (note: *FilterCtrlActions* is not a single class but stands for a collection of action classes for manipulating the sliders, such as navigating among sliders and adjusting slider thumbs). Each filter control action class performs some designated logical operations according to information stored in *DataSet / InteractionParameters*, making changes to *DataSet / InteractionParameters* (such as changing regions' filter status), and requesting *VizFilterCtrls* and *AudFilterCtrls* to generate visual and auditory feedback about the sliders. Since changes made to *DataSet / InteractionParameters* are immediately propagated to all views in iSonic, and the filter control view is usually shown together with either the map view or the table view, users also see and hear the filtering results in the context of either the map or the table.

5.2 Input Files

Upon startup, iSonic looks for a configuration file (typically called *appParameters.csv*) and initiates itself. Depending on the parameter values in the file, iSonic may load the following five types of input files during initiation and runtime: a region-by-variable data file, a map image file, information supporting map exploration, prerecorded wave sound files, and HTML files for the online help.

Configuration File

The configuration file (*appParameters.csv*) contains a list of parameter-value pairs. It tells iSonic where to find certain input files and information needed to connect to

external spatial sound servers, and defines the default visual and auditory behaviors during the runtime. The first trunk of configuration parameters specifies the directories and names of the region-by-variable data file, the map image, the information files supporting map exploration, and the entry HTML file. It also gives the directory of prerecorded wave sound files. The second trunk defines the default sound encoding schemes, such as musical instruments, sound volume, duration, inter-region pause, whether to encode region borders, whether to use end-of-sweep sounds, whether to use prerecorded sounds or spatial sounds, and so on. The third trunk tells iSonic how to communicate with the external spatial sound server, and the desired virtual auditory map size and shape, such as cylinder or sphere. The fourth trunk of information is specific to map exploration, such as the maximum zooming depth allowed, the coarseness of the mosaic map display, and the default spatial sweep order. The fifth trunk of information controls the interaction, such as to define the finger wobble buffer zone size, and to enable/disable each individual feature in iSonic. The sixth information block decides starting points, such as the information level, the region and the variable to start with. The remaining parameters specify miscellaneous configurations, such as whether to synchronize the visual and auditory displays, and whether to record logs.

Region-by-Variable Data File

The region-by-variable data file (.csv) contains the set of data to be explored (Figure 5-2). iSonic accepts three data types: String, Integer, and Float. The first and second columns in the file contain the FIPS (Federal Information Processing Standards Codes) and names of geographical regions. The remaining columns are numeric values of statistics variables for each region. The actual region names used by iSonic are in the

supportive information for map exploration (see below) instead of those in the name column.

	A	B	C	D	E	F	G
1	State Fips	State Name	Population 21	Population 21	Percent Popu	Percent emp	Percent empl
2	Integer	String	Integer	Integer	Float	Float	Float
3	1000	Alabama	363007	52429	14.4	53.4	76.5
4	2000	Alaska	2534550	422291	16.7	29.5	76.7
5	4000	Arizona	3049112	356749	11.7	38.9	75.4
6	5000	Arkansas	1511967	266635	17.6	34.8	79.6
7	6000	California	20245973	2182821	10.8	36.4	74.7
8	8000	Colorado	2699244	246151	9.1	48.3	79.6
9	9000	Connecticut	1993779	185968	9.3	42.7	79.2
10	10000	Delaware	459022	54114	11.8	43.3	80.4
11	11000	DC	347635	39697	11.4	38.1	74.9
12	12000	Florida	9308645	1122010	12.1	37.6	76.8
13	13000	Georgia	5043079	611559	12.1	34.8	78.4
14	15000	Hawaii	699814	73578	10.5	41.7	77.8
15	16000	Idaho	753646	110913	14.7	47.5	79
16	17000	Illinois	7234804	674997	9.3	39.2	76.3
17	18000	Indiana	3483306	463402	13.3	41.7	78.6

Figure 5-2: A sample region-by-variable data file for U.S. states (.csv).

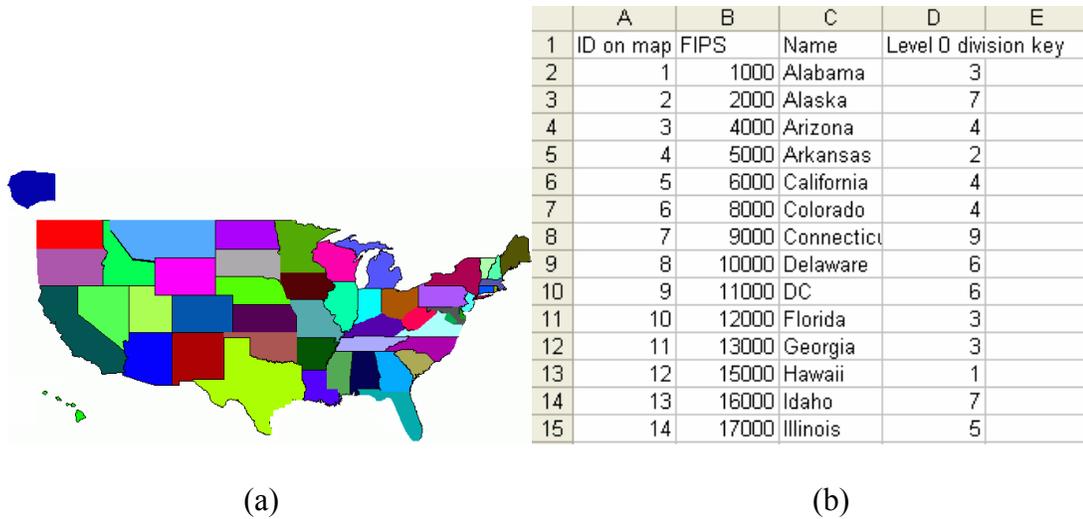
Map Data

A shape file [52] is a very commonly used data format for vector spatial data. The geometry for a spatial feature (such as point, line, polygon) is stored as a shape comprising a set of vector coordinates. A shape file is a nontopological data structure that does not explicitly store topological relationships, such as polygon adjacency, connectivity, and containment. Such topological relationships can be calculated by Geographical Information System (GIS) mapping software such as commercial software by ESRI [52], or a wide range of open source software listed at the Open Source GIS website (<http://opensourcegis.org/>).

Although adjacency information is not hard to calculate by intersecting each

polygon with other polygons and identifying the points of intersection of polygons that touch boundaries or overlap, I chose to use a color-coded pixel map image as the map shape input file for the current iSonic implementation, because of several reasons. First, such a map image (.gif) is often much smaller than the corresponding vector shape file. It allows faster Web download and improves user experience over a slow network connection [193]. Second, it was easier to digitize a raster map image to generate a mosaic version of the map for investigating cell-by-cell mosaic map navigations (see Chapters 4.3.2 and Chapter 4.1.3). It was also easier to calculate topological relationships from a grid representation of such a mosaic map.

iSonic accepts the map data in the format of an image-table pair. The decoding map image (.gif or .bmp) (Figure 5-3 (a)) contains region geometrical shapes, and the tabular map information file (.csv) (Figure 5-3 (b)) provides a list of valid regions with properties such as region names. The correspondence among the decoding map and the information file was established through ID color coding (Figure 5-3 (c)). The ID-color correspondence is further illustrated in an RGB color cube in Figure 5-4. A full explanation of how to generate and interpret such map data is available in [193]. Most GIS mapping software shall be able to generate such decoding map images given any shapefile. I have also developed a tool for this purpose. The tool is available by visiting www.cs.umd.edu/hcil/iSonic.



ID on map	State name	Color (RGB value) for encoding the ID (as shown on the decoding map in (a))
0	State border	0x000000
1	Alabama	0x000055
2	Alaska	0x0000AA
3	Arizona	0x0000FF
4	Arkansas	0x005500
5	California	0x0055AA
...
63	Background	0xFFFFFFFF

(c)

Figure 5-3: (a) A sample decoding map of U.S. states and (b) the corresponding information file (.csv) defining valid regions on the map and the top zoom level manual 3x3 recursive map partition; (c) the mapping between the decoding map and the information file is established by encoding IDs in colors. The encoding rule is illustrated in Figure 5-4.

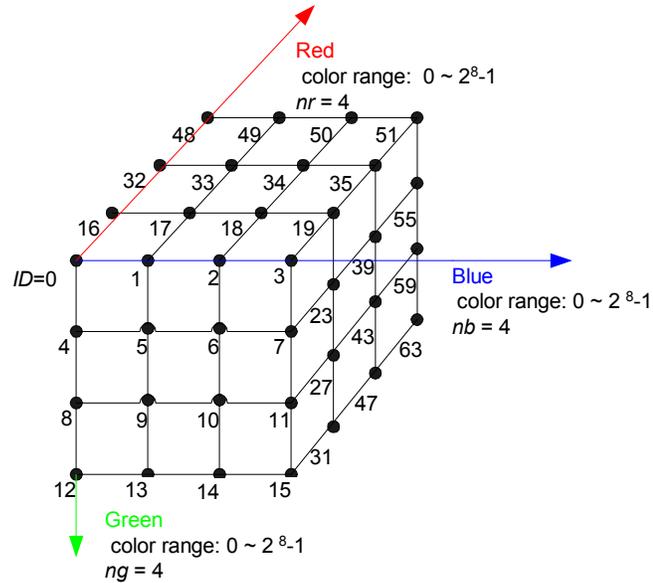


Figure 5-4: The RGB color space and the relationship between IDs and the colors to encode the IDs ($nr = ng = nb = 4$). The colors at the grid vertices are actually used.

The map information file (Figure 5-3 (b)) may also contain one or more columns defining manual 3x3 recursive map partitions that are used in the 3x3 map exploration. When such a definition does not exist for a zoom level, iSonic automatically generates it according to regions' mass center coordinates.

While iSonic can automatically generate map spatial sweep orders, users can manually define the spatial sweep orders through the use of one or more input files (.csv), suppressing the automatic generation. Figure 5-5 shows such an input file defining the horizontal sweep order for the 51 regions on a U.S. state map. The numbers are “ID on map” in the data file shown in Figure 5-3 (b).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2														
2		48			27	35	24								
3		38	13		51	42		50	23			33	46	30	20
4						28	16	14	15	36	39	31	22	7	40
5		5	29	45	6	17	26		18	49	47	9	21	8	
6				3	32	37	4	43	34	41					
7	12					44	19	25	1	11	10				

Figure 5-5: A sample file (.csv) defining a map spatial sweep order (horizontal sweep) for a U.S. state map (also see Figure 4-12 (a)).

5.3 Map Preprocessing

Upon startup, iSonic loads the decoding map (specified in the configuration file), and does a sequence of preprocessing (Figure 5-6) to generate a mosaic map (in a resolution requested by the configuration file) and related information, including a vector representation and an adjacency list for each region, and its mass center position.

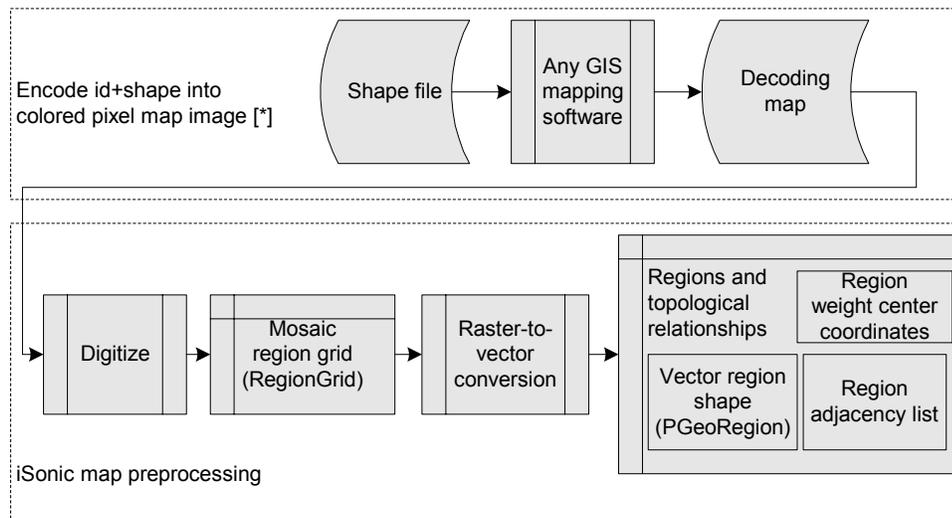


Figure 5-6: Read and preprocess map shape data. ([*]: see [193] in Bibliography for details)

To digitize the decoding map image, an even-sized grid of the requested resolution is laid on top of it. All the map regions are ordered increasingly by their bounding box sizes. Each region is then assigned one cell closest to the region's bounding box center, to ensure that even a small region gets one cell. After that, each unassigned cell is assigned to the region that has the maximum count of pixels in the cell. Figure 5-7 (a) shows a part of an actual decoding map (with many regions removed) and Figure 5-7 (b) shows the mosaic grid map after the digitization. The mosaic grid map was stored in a 2-D matrix *RegionGrid*.

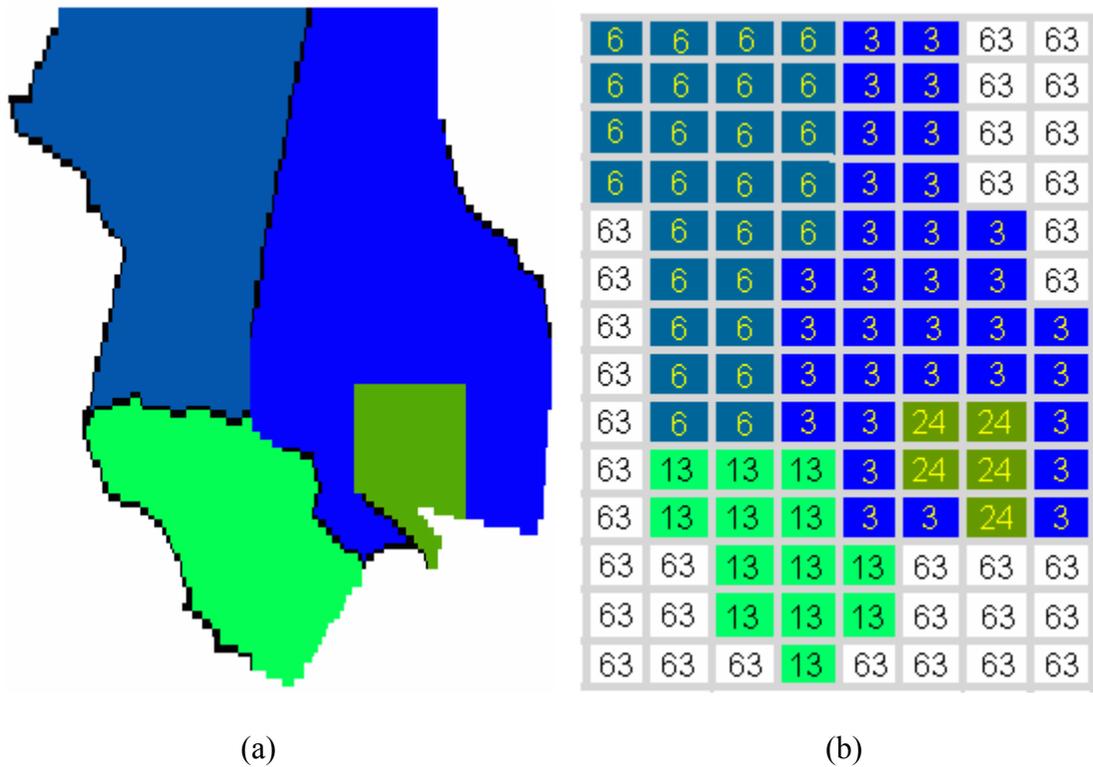


Figure 5-7: Map digitization: (a) a partial decoding map; (b) the mosaic region grid after the decoding map is digitized.

By tracing the block border of identical grid cells in *RegionGrid*, a vector representation of each region is generated, together with a list of adjacent regions. Figure 5-8 (a) illustrates the tracing route for region #3 and Figure 5-8 (b) shows the resulting vector representation consisting of a sequence of coordinates (not all coordinates are drawn in the figure): (4, 0) -> (6, 0) -> (6, 4) -> (7, 4) -> (7, 6) -> (8, 6) -> (8, 11) -> (7, 11) -> (7, 8) -> (5, 8) -> (5, 10) -> (6, 10) -> (6, 11) -> (4, 11) -> (4, 9) -> (3, 9) -> (3, 5) -> (4, 5) -> (4, 0). Region #3 is adjacent to regions #24, #13, and #6. Its mass center coordinates are (94/5, 214/11). Such border tracing algorithm in iSonic deals with regions with multiple detached parts, and both concave and convex shapes.

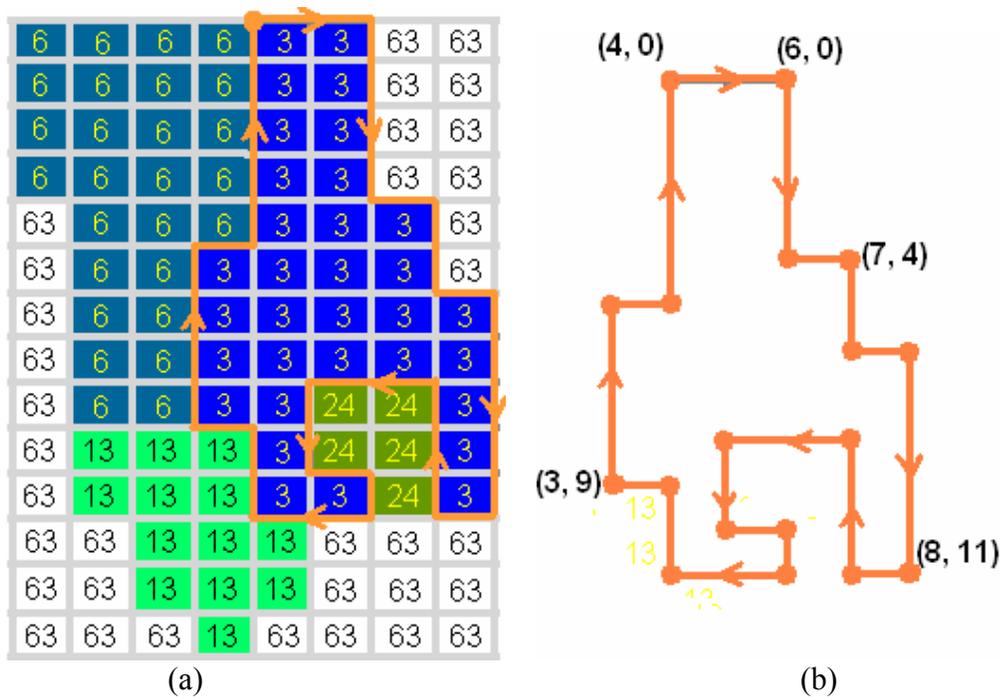


Figure 5-8: Raster to vector map conversion: (a) obtain a vector representation of region #3 by walking the mosaic region grid; (b) the vector representation of region #3 (not all coordinates are drawn).

5.4 Generate Map Spatial Sweep Orders

If no manual map spatial sweep order is defined in the form of a .csv input file, iSonic automatically generates the order(s) upon startup. To simplify the description, I focus on the discussion of generating a vertical sweep order. Horizontal sweep order generation can use a similar method. For an arbitrary map with a set of non-overlapping 2-D geometric polygons (regions), automatic vertical sweep order generation can be considered as an optimization problem to search for a collection of vertical traveling paths to connect the weight centers of these polygons. The polygons connected by each continuous vertical path belong to one subset. Intuitively speaking, the optimization goals are to minimize the X dimension interleaving among the resulting subsets, and to make the traveling paths as straight and vertical as possible.

There have been some classical optimization problems that deal with connecting points on a 2-D space. For example, the well-known traveling-salesman problem (TSP) [73] tries to find the shortest traveling path that starts from one point, visits each point once and only once, and returns to the starting point. The problem is known to be NP-complete. The optimization goal of TSP is different from the sweep order problem here because of two reasons. First, TSP only deals with points, not the shape and size of 2-D polygons. Although the shape and size of a polygon decide its weight center position, the weight center position does not sufficiently represent the polygon itself, thus cannot reflect the degree of polygon interleaving. Second, a traveling path that is the shortest is not necessary the straightest, nor does it minimize the degree of polygon interleaving. Figure 5-9 shows 8 regions. The whole traveling path in (a) is shorter than the whole path

(the sum of the two paths) in (b). However, it makes more sense to split the states as shown in (b) because the two resulting subsets clearly separate from each other visually, and the resulting path is perfectly straight and vertical.

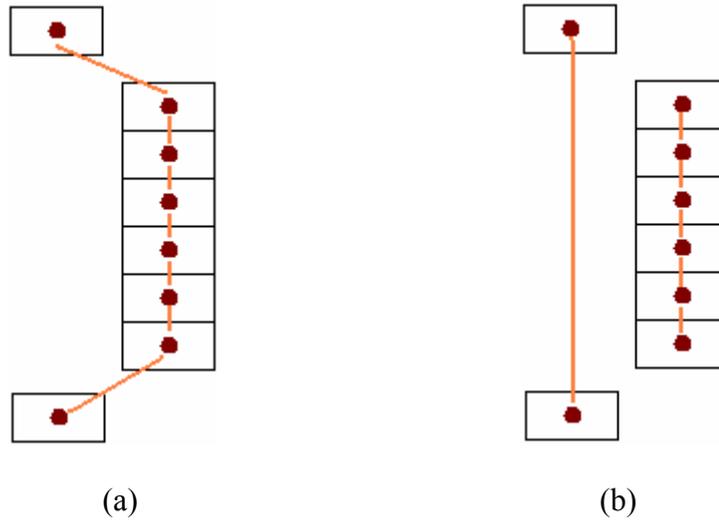


Figure 5-9: A shortest traveling path is not necessarily the straightest, nor does it minimize the degree of polygon interleaving. The whole traveling path in (a) is shorter than the whole path (the sum of the two paths) in (b). However, the path in (b) is straight and vertical, while the path in (a) is not.

In my investigation, I use the larger value of two area intruding ratios to quantitatively measure the degree of interleaving between two region subsets. The intruding ratio of subset #1 into subset #2 is a percentile. The numerator is the total size of all areas in subset #1 that fall into the bounding box of subset #2. The denominator is the total size of all regions in subset #1 that overlap with the bounding box of subset #2. The straightness and verticality of a traveling path can be measured by the sum of

deviation angles between the travel path segments and the vertical direction (Y axis). The second optimization goal could conflict with the first goal. For example, for the same set of regions in Figure 5-10, a split resulting in no intrusion (Figure 5-10 (a)) produces travel paths that are less straight and vertical than another split (Figure 5-10 (b)) that causes significant intrusion.

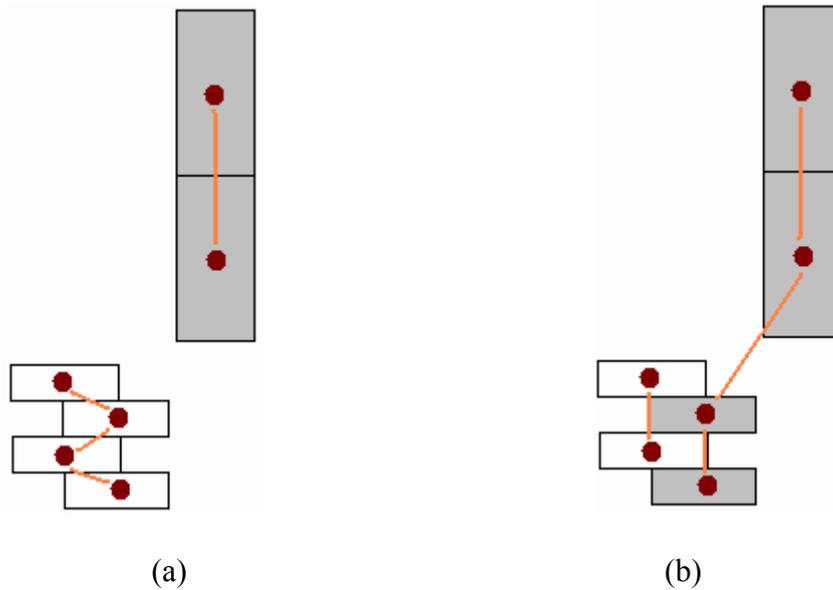


Figure 5-10 : Two competing optimization goals in map spatial sweep order generation. For the same set of regions, the split in (a) does not cause any intrusion between the two subsets but produces travel paths that are less straight and vertical than the ones in (b). However, the split in (b) causes significant X-dimension intrusion between the two subsets.

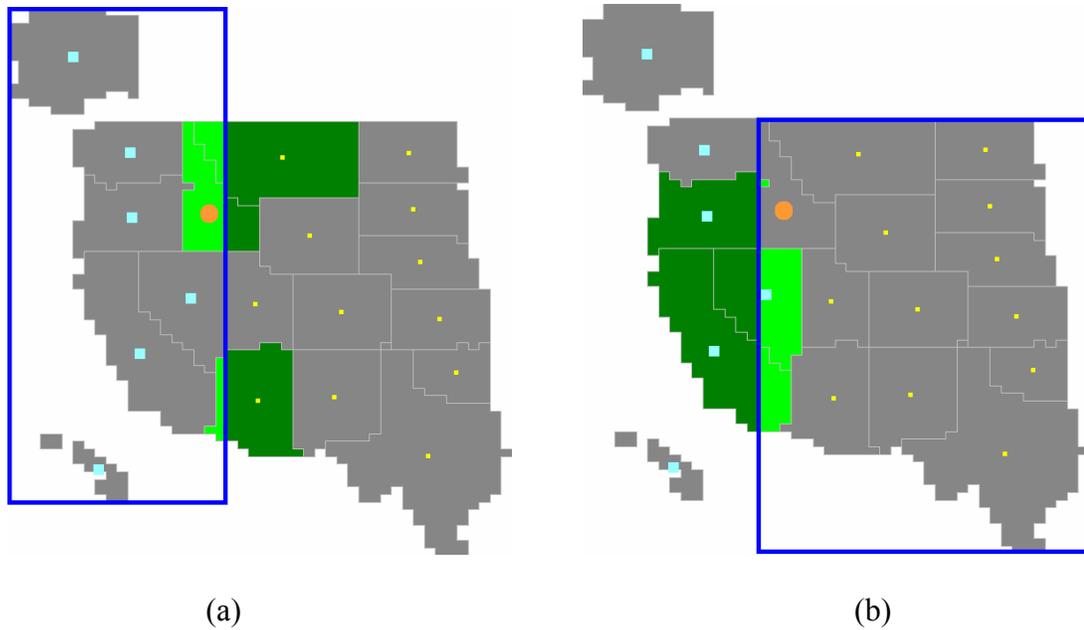


Figure 5-11: Calculating the overlap ratio among two region subsets when the regions are split vertically at the circular orange-colored mass center. All dots are mass centers. Large sky blue dots are for regions in the left subset, and small lemon yellow dots are for regions in the right subset. Light green marks the intruding areas and dark green marks the remaining areas of regions that intrude into the bounding box of the other subset.

In my preliminary investigation, I focus on minimizing the degree of subset interleaving in order to generate a vertical sweep order congruous to the visual scanning and grouping impression. How cleanly two subsets can be separated by a vertical split line makes more sense than producing straight paths in such context. To generate a vertical sweep order, a two-step procedure is used. First, regions are recursively split into several subsets. Then regions inside each subset are visited in sequence. The splitting takes a greedy approach without any backtracking, and does not guarantee a globally

optimized splitting. However, the procedure seems to have produced reasonable results on the maps tested, including a U.S. state map (Figure 5-12), a Maryland county map (Figure 5-13), and an Idaho county map.

To split regions into several subsets, a recursive procedure is used. Appendix B gives the algorithms used to check and split a region subset in each recursive pass. In each pass, iSonic first looks for a split position that causes the minimum area interleave between the resulting region subsets. All the mass center X coordinates are potential vertical split positions. The area interleave between two region subsets is measured by two intruding ratios. Figure 5-11 shows a potential split at the orange-colored mass center X coordinate. The mass centers of all regions in the left subset are sky blue, and those in the right subset are lemon yellow. The intruding ratio of the left subset into the right subset is a percentile (Figure 5-11 (b)). The numerator is the total size of all areas in the left subset that fall into the bounding box of the right subset, as marked in light green. The denominator is the total size of all regions in the left subset that overlap with the bounding box of the right subset, as marked in dark green. Similarly, the intruding ratio of the right subset into the left subset is illustrated in Figure 5-11 (a). After the two intruding ratios are obtained, each of them is compared against a predefined per region intrusion ratio (*intrusionRatioPerRegion*) multiplied by the number of intruding regions in the corresponding subset. Only when neither ratio is larger in the comparison, the larger value of the two ratios is compared with the *currMinIntrusion* value obtained about other possible splits checked so far. If the larger of the two ratios is smaller than *currMinIntrusion*, the current split is the best so far. After all potential split positions are checked, the best split is used to actually split the regions. If none of the split positions

produces a qualifying intruding ratio, the split stops for the current region subset.

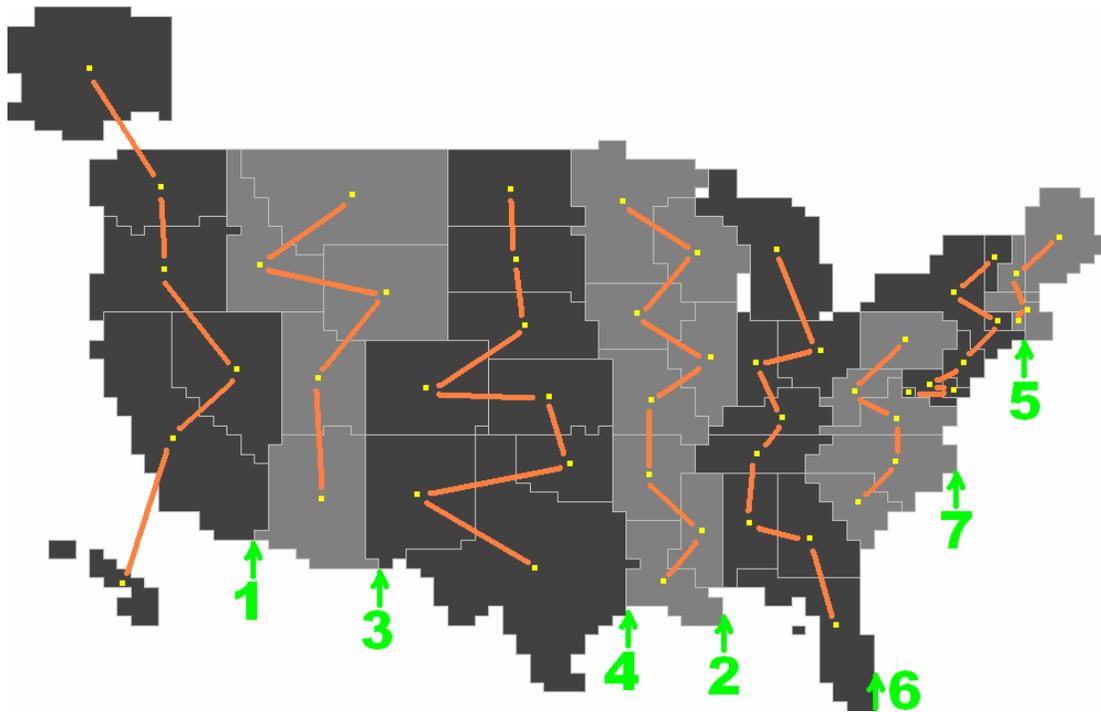


Figure 5-12: An automatically generated vertical spatial sweep order for a U.S. state map. The green numbers indicate the order of splits in the recursive splitting algorithm (per region intrusion ratio = 0.2). The orange lines indicate the sweep path. The yellow dots are region mass centers.

The per region intrusion ratio (*intrusionRatioPerRegion*), specified in the input configuration file, affects how fine regions are split. An empirical value of 0.2 produces reasonable results on a U.S. state map (Figure 5-12), a Maryland county map (Figure 5-13), and an Idaho county map.

After subsets are calculated through multiple recursive passes, regions inside each subset,

starting from the left most subset, are visited in sequence to form a sweep path. The order of regions in each subset is decided by the mass center Y coordinates. When two regions have the same Y coordinates, the one with the smaller Y coordinate is visited first. Figure 5-12 and Figure 5-13 show the resulting vertical spatial sweep orders on a U.S. state map and a Maryland county map. The green numbers indicate the split positions in the order of their occurrences.

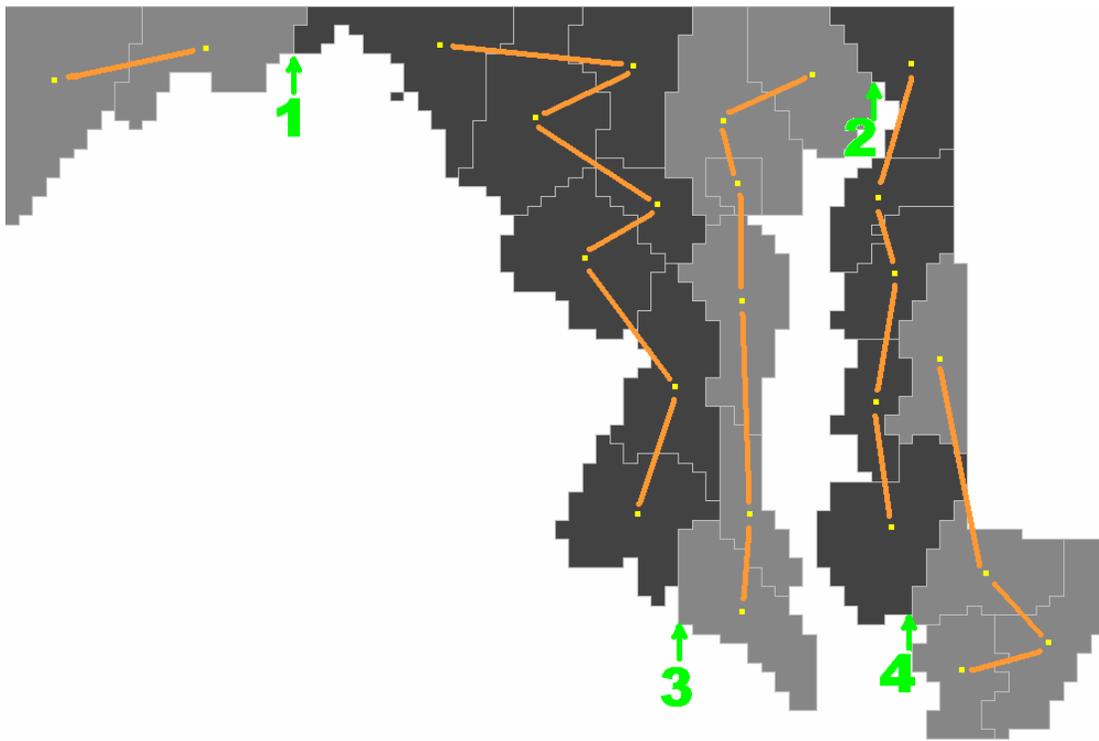


Figure 5-13: An automatically generated vertical spatial sweep order for a Maryland county map. The green numbers indicate the order of splits in the recursive splitting algorithm (per region intrusion ratio = 0.2). The orange lines indicate the sweep path. The yellow dots are region mass centers.

5.5 Generate Map Adjacency Matrices

When users press one of the four arrow keys in the map view, iSonic needs to decide which region users go to from the current region. The goal is to find a region that is as straight to the navigation direction as possible. Priority shall be given to regions that share borders with the source region, but then the algorithm also needs to consider other regions to ensure the connectivity of isolated regions. To simplify the description, I assume the users press the up arrow key to navigate upwards from a given source region

The target region is automatically calculated by a two-step algorithm. First, iSonic looks for an appropriate region in the source region's adjacency list (regions that share a common boundary with the source region). Second, if no suitable region is found in the first step, iSonic checks if there is a region that does not share a common border with the source region but is located along the navigation direction from the source region. Such region is usually separated from the source region by water. The second step considers the accessibility of individual isolated regions (such as islands) and the connectivity between region blocks separated by water. The algorithms are given in Appendix C.

Figure 5-14 and Figure 5-15 illustrate how the algorithms work. In Figure 5-14, the user moves upward from region #2. iSonic first check all regions in the adjacent list of region #2. The mass centers of those regions are marked with big orange dots). Region #1 has the smallest deviation angle with the upward direction among the three regions falling into the upper 90° radar range. Region #1 is the target. However, if the user moves to the right from region #2, none of the regions in the adjacent list is in the right side 90° radar range. iSonic first finds the right edge position of region #2 (marked as a big blue

In Figure 5-15, the user moves rightward from region #1. iSonic first checks all regions in the adjacent list of region #1 but none of them falls into the 90° radar range. iSonic then finds the right edge position of region #1 (marked as a big blue dot), then uses a scanner centered at the blue dot, starting from the right direction, gradually widening and elongating. The scanner first encounters region #4. Since region #4 is in the adjacent list of region #1, it is the target region.

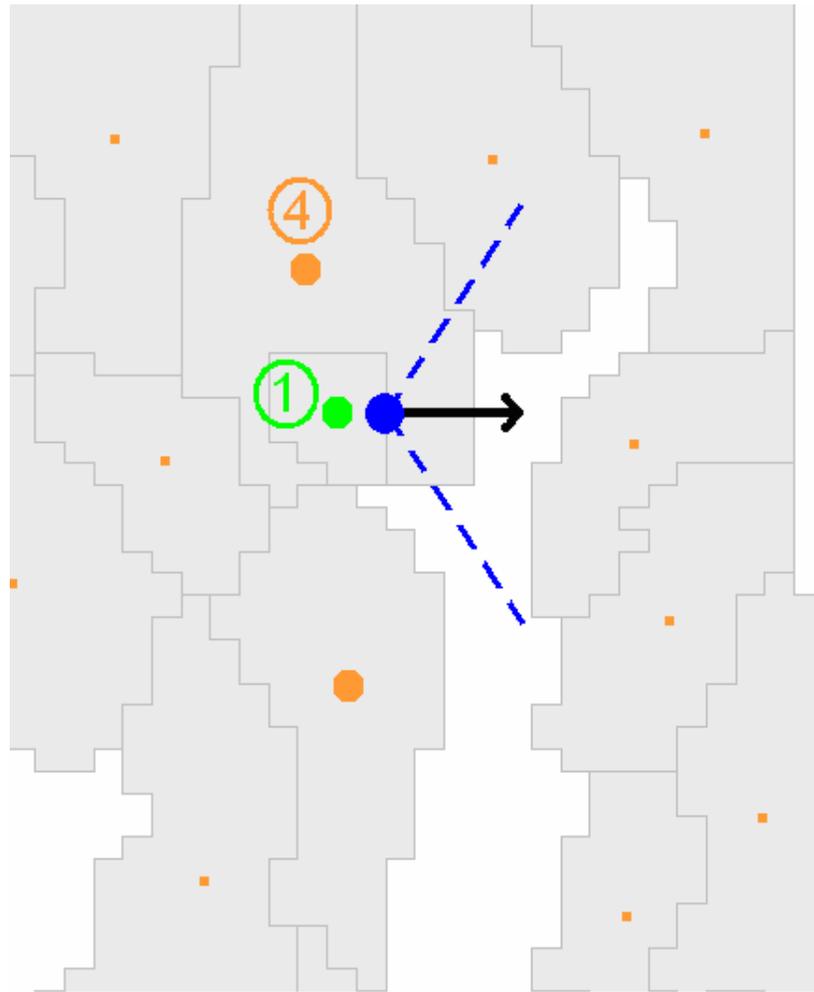


Figure 5-15: From region #1, moving to rightwards will land in region #4. All dots represent region mass centers. Big orange dots are for regions in the adjacent region list of the source region #1 (marked with a big green dot).

5.6 Discussion

The iSonic implementation is based on the Model-View-Controller (MVC) paradigm, resulting in a clear modular architecture that can easily be extended to include more data views. It also extended the use of MVC from GUI to Auditory User Interface (AUI) by treating the auditory rendering as a view that resides in parallel with the corresponding visual view.

The implementation allows many design features to be configured through the change of parameter values in the configuration file. This allows iSonic to be adapted for different users and to be used as a research tool to compare some design options. Some of the configuration parameters can be adjusted through an application control panel. A full implementation of the control panel that includes all parameters can make configuration even easier.

It is a very interesting and challenging problem to automatically produce map spatial sweep orders that are congruous to what people may define manually based on their visual impressions. Some preliminary algorithms were described. They use a greedy approach and do not guarantee globally optimized solutions, but produced reasonable results for the maps tested so far, including a U.S. state map, a Maryland County map, an Idaho county map, and a Montgomery county district map. The efficacy of these algorithms needs to be further tested on more maps. The problem could also be defined using optimization criteria that are different from the ones used in this chapter. For example, to generate a vertical sweep order, we can first morph all the polygons to fit into vertical stripes, and then travel the polygons in the order of the stripes. The optimization

goal is to minimize the degree of morphing which can be measured by the change in polygon size, weight center position, et al.

Chapter 6

Evaluations

By working closely with two blind users, I was able to receive continuous feedback that was used to improve iSonic design. In addition, multiple formal user evaluations were conducted, including both controlled experiments and case studies. Chapter 4 describes some earlier explorative designs and two controlled experiments with sighted users. The effort allowed me to identify effective design options for geo-referenced data that are used in the iSonic design. It also partially confirmed the utility of the ADC framework, and provided insights into human ability to perceive information from interactive sounds. In this chapter, I describe my continuing effort in understanding the effect of design choices and user difference in their geographical data pattern recognition ability. I also report 42 hours of case studies with blind users. The results allowed me to examine the effectiveness of iSonic in supporting complex data exploration tasks, and to further validate the utility of the ADC framework.

6.1 Geographical Data Pattern Recognition

6.1.1 Effect of Input Device and Blindness

Results of the two controlled experiments in Chapter 4 showed that geographical data distribution pattern recognition was possible, although hard, for sighted users when using keyboard relative navigation. In order to improve location awareness, pattern recognition performance, and map knowledge training, better sonification designs are needed. Design

lessons from the two studies were applied to improve iSonic design, and new navigation methods were designed, including the recursive 3x3 absolute keyboard navigation, and the absolute continuous touchpad navigation. Informal trials with three blind users showed that the absolute navigation methods improved location awareness and they were able to learn new maps.

To formally investigate the effect of input devices and user blindness on geographical knowledge learning and geographical data distribution pattern recognition, a third user study was designed [198]. The data collection is being carried out by researchers at the University of Rome “La Sapienza”, Italy [198], while this dissertation is being written. The study is a 2 x 3 between-subject experiment to compare the performance of the keyboard-only interface and the touchpad interface with 15 users with congenital blindness, 15 with acquired blindness and 15 sighted but blindfolded users. To test Italian subjects, an Italian configuration of iSonic was created by recording Italian speech wave files and configuring iSonic to play prerecorded wave files instead of producing speech through the speech server.

Our hypotheses are that touchpad user will have higher satisfaction and give more correct answers to questions about shape, size and density of areas; but they will also be slower at finding areas with specific values and may miss some areas entirely. We also hypothesize that sighted users will perform better than users with acquired blindness or congenital blindness and acquired blind users will provide a better performance than congenitally blind subjects, due to their earlier visual spatial experience. Blind subjects have more experience analyzing environmental sounds [69], but it is not clear whether this enhanced awareness extends to other auditory domain, such as listening to abstract

sounds, and they have less experience dealing with spatial information. One of the challenges with this experiment is to evaluate how well blind users perceive the pattern of data on the map, i.e. general trends, location of clusters and exceptions. We are testing what subjects have learned about the data pattern with a multiple choice test administered with printed tactile maps. We are also testing what subjects have learned about the map itself with a multiple choice test.

Preliminary data from 17 blind-folded sighted subjects showed that iSonic led to a fairly good representation of the map. For easier tasks, keyboard and touchpad exploration showed no differences in recognition performances. However, for the hardest task in the test, it seemed that the touchpad, with no direction constraints, allowed more accurate explorations, leading to higher scores of recognition. The data collection is still underway while this dissertation is being written. Observations and log analysis will allow us to compare user strategies, and inform future interface refinements.

6.2 Case Studies

During early iSonic design iterations, controlled experiments were conducted with sighted subjects to compare the effectiveness of design alternatives including the choice of data views, map navigation structures and auditory feedback encoding schemes [194][195] (also see Section 4.1). 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, a series of case studies were conducted with blind subjects. 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 blind users, producing 42 hours of observation and interview data, with an average of 6 hours per user. Using cross-case analysis, common user behaviors and feedback were extracted that allowed me (1) to evaluate the effectiveness of iSonic design decisions for blind users; (2) to identify features helpful to each data exploration task category and examine the utility of the ADC framework; (3) to identify task road blocks in order to target training and modifications to the interface and the framework.

6.2.1 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 at the time of the study. 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, and a master degree in rehabilitation. 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.

6.2.2 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. All the features were available except for the slider-based filtering. Filtering is complex even for sighted novice users and was not evaluated in the studies.

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 U.S. 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 U.S. states and Maryland counties ranged from excellent to very poor. This allowed me 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 for iSonic are in Table 6-1.

Id	(Type)/Task
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?

Table 6-1: The seven test tasks for iSonic in the case studies.

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 an intention to compete with Excel. Rather, Excel was considered as the standard tabular data viewer, and the comparison was used 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. No tactile maps were provided when subjects used Excel, because many blind users do not have access to tactile maps. Only P5 owns one for Maryland counties.

6.2.3 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. Subjects were asked to do the Excel task first because I 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.

6.2.4 Description of Seven Cases

Table 6-2 lists the task completion of each subject during the training and the actual tests.

Table 6-3 summarizes the task success rates. Readers can see Appendix A – 1 through Appendix A – 7 for a complete description of each of the seven cases. The seven cases are summarized and analyzed in Section 6.2.5 to derive common user behaviors and identify individual differences. The cross-case analysis of the seven cases reveals strengths and weaknesses of iSonic design and provides insights into the validity of the ADC framework.

Tasks		P1	P2	P3	P4	P5	P6	P7
T1	iSonic training	■	■	■	■	■	■	■
	iSonic testing	■	■	■	□	□	■	□
	Strategy carry-over	—	—	—	—	—	—	—
	Excel w/ JAWS	■	■	■	□	■	■	■
T2	iSonic training	■	■	■	■	■	■	■
	iSonic testing	■	■	■	■	■	■	■
	Strategy carry-over	—	—	—	—	—	—	—
	Excel w/ JAWS	■	■	■	■	■	■	■
T3	iSonic training	□	□	□	□	□	□	■
	iSonic testing	■	■	■	□	■	■	■
	Strategy carry-over	√	√	√	×	√	√	—
	Excel w/ JAWS	□	■	■	■	■	■	■
T4	iSonic training	■	■	■	■	■	■	■
	iSonic testing	■	■	■	■	■	■	■
	Strategy carry-over	—	—	—	—	—	—	—
	Excel w/ JAWS	■	■	■	■	□	□	■
T5	iSonic training	■	■	■	×	■	□	□
	iSonic testing	■	■	■	□	■	■	■
	Strategy carry-over	√	√	√	√	√	√	×
	Excel w/ JAWS	×	×	×	■	■	—	—
T6	iSonic training	□	■	■	■	□	□	○
	iSonic testing	■	■	■	■	■	■	■
	Strategy carry-over	√	—	—	—	×	√	—
	Excel w/ JAWS	—	—	■	—	■	—	—
(Continued on the next page)								

(Continue the previous page)								
Tasks		P1	P2	P3	P4	P5	P6	P7
T7	iSonic training	■	■	■	■	■	■	×
	iSonic testing	■	■	■	■	■	■	■
	Strategy carry-over	√	—	—	—	—	√	—
	Excel w/ JAWS	■	■	■	—	×	—	—

Table 6-2 : Task completion of all the seven subjects in the case studies.

‘■’ means the task was successful without viewing the sample solution (in iSonic training) or any other help. ‘■’ means correct answer was obtained by following a sample solution (in iSonic training) and/or with some help from the experimenter. Some subjects directly followed the sample solution w/o first trying on their own. ‘○’ means the subject did not use any help but gave the wrong answer because of careless mistakes. ‘×’ means wrong answer was given even after viewing the sample solution (in iSonic training) and/or with help from the experimenter. ‘×’ (in ‘strategy carry-over’) means subjects chose their own strategies over the sample solution although they remembered the sample solution. ‘—’ means the subject did not try because of the lack of geographical knowledge (in Excel w/ JAWS), or the situation was not applicable because the subject did not view the sample solution (in ‘strategy carry-over’).

	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%

Table 6-3 : Task success rates in the case studies.

6.2.5 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 self-paced auditory lessons, subjects successfully completed 67% of the training tasks without referring to the sample solution or any other help, as shown in Table 6-2. 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. (Note: iSonic's one-key in-place sorting feature was much preferred than Excel's sorting dialog window. However, reader shall bear in mind that Excel is a general application. Therefore, research is needed to investigate how such one-key in-place solution would scale when the variety of interface functions increases.)

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 geo-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.

6.2.6 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. The map was useful not only for blind users to solve geo-related tasks but, according to user comments, was considered every helpful for them to communicate with sighted people.

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 most 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.

Pitch

Using pitch 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.

Panning

Some subjects reported to have noticed the stereo sound panning, but P2 was the only one that intentionally used it as a position cue in tasks.

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. 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 method 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, and some instantly connected the keys to the concepts of west, middle, northeast, and so on. 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. Some explanations included that 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 method. 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 one subject mentioned the irreversibility problem, she thought it was a natural fact about maps and there was no good solution. 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 the “situate” feature 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, many subjects explicated said that this function was essential because they would 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.

6.2.7 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.

Reflecting on the study methodology, case studies turned out to be very productive in revealing design strengths and opportunities for further improvements even with a small number of subjects. An exploratory data analysis task is a complex process that involves multiple interface components and can be affected by many inherent human subject variations. By spending intensive quality time with each subject, observing their

behaviors, hearing their thoughts and comments throughout the tasks and interviews, I was able to not only understand their mental strategies that sometimes evolved over extended interface usage, but also identify the causal relationship of design factors and user performance. Case studies not only allowed capturing common user behaviors, but also revealed inherent individual differences. Identifying individual differences can help making design compromise that work for a wide range of users.

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. The studies only tested highly educated users without any residual vision. Further studies with partially sighted users or users with lower education levels may reveal different usage patterns and visual- auditory interactions that may modify the results and framework.

However, the studies showed that the Action-by-Design-Component (ADC) 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. I 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. The 4 information 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 number pad 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 number pad 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.

6.3 Summary and Discussion

Starting from the very early stage of explorative designs and user evaluations, the ADC

framework was used to guide the systematic exploration of various design options for geo-referenced data, along the dimensions of data views, auditory feedback encoding, and navigation structures. The framework continued to be used to guide the design space exploration, such as the choice of input devices. The series of user evaluations provided invaluable insights into human ability to perceive complex information from interactive sounds, and revealed some general design implications

Particularly, the series of evaluations showed that it was important to choose appropriate data views for different tasks, and sonification helped presenting tables and choropleth maps to both sighted and blind users. Blind users also had needs for maps but the lack of appropriate map tools limited them from performing geo-related tasks.

The evaluations showed that dramatic benefits accrued from rich combinations of task-oriented interactions. The AISA interactions in the ADC framework, applied to multiple highly coordinated data views, were impressively effective for even blind users to accomplish complex exploratory tasks. The ADC framework, which is analogous to many interactions in visualizations such as the widely adopted information seeking mantra, works for auditory interfaces when applied with appropriate adaptation.

The case studies only tested motivated advanced computer users because they were the current target users. The studies were done under supervised experimental setups. It would be interesting to see the actual use of iSonic in a natural daily work environment, and whether naïve blind users would be able to learn iSonic with the self-paced auditory tutorial, and to use iSonic to solve complex tasks. There were also some aspects of iSonic design that was not evaluated, such as the slider-based filter action. Slider-based dynamic and continuous filtering was found to be complex even for sighted

novice users. It would be interesting to find out whether blind users are able to handle its adapted version in iSonic, and if not, what the alternative designs are.

Chapter 7

Interactive Sonification of Abstract Data Graphs

User evaluations showed that blind users were able to effectively handle highly coordinated data views, and dramatically benefited from a rich combination of task-oriented user interactions. In this chapter, I extend the ADC framework to other graphical data views, including bivariate scatterplots and line graphs. For scatterplots, I compare two interactive sonification designs, and describe how one design has been implemented and integrated into iSonic. For line graphs, I analyze how some of the AISAs were supported in previous work and how the remaining AISAs could be designed for keyboard interaction.

7.1 Scatterplot

A visual bivariate scatterplot is a graph commonly used for revealing the linear correlation relation in a set of bivariate data (two variables). Each data item is a dot on the graph, with the (X, Y) coordinates determined by its values for the two variables. A scatterplot has some similarities with a univariate choropleth map. An essential difference is that many data items (dots) in a scatterplot may overlap with each other while data items (regions) in a choropleth map will not.

There are several possible methods for sonifying a scatterplot. Below is the discussion of two methods. The dot scanning method treats individual data item (dot) separately, and the dot aggregation method clusters and aggregates dots spatially.

7.1.1 Dot Scanning

The dot scanning method extends an earlier work by Flowers and colleagues [60] which examined the equivalence of visual and auditory bivariate scatterplots with sighted users. In their study, subjects listened to a quick sonified overview of a scatterplot containing 50 randomly generated bivariate data samples. The sonified overview consisted of representing one variable (Y variable) by pitch and the other variable (X variable) by time. It was based on dot scanning in which a vertical scan line moved along the X axis at a constant speed and the dots were played when encountered by the scan line. The pitch of each dot represented its Y value. Such a sonified overview gives the impression of a "cloud" of musical notes, and was found to be as quite efficient as the visual equivalence in conveying sign and magnitude of correlation. The effect of outliers on judged magnitude of correlation was also similar for both the auditory and visual overviews.

Flowers' work gave a sonified overview (gist), but did not support any user interactions with the scatterplot. Extending this basic dot scanning method, the design below supports Auditory Information Seeking Actions (AISAs) for exploring a scatterplot.

Gist

A gist of the scatterplot can be produced by moving a vertical scan line along the X axis and playing the dots encountered by the scan line. Multiple dots in a scatterplot may have the same X value, and some dots may completely overlap each other. In Flowers' auditory scatterplot, those dots were played in parallel. This caused multiple overlapped data items to potentially appear as one single data item. Another way is to

play those dots sequentially and use a special click sound to indicate the end of such a dot group. The scan line pauses during the playback of the group and resumes movement upon the special click sound. The sequential method also makes it possible to play multiple dots at a highly detailed information level. A gist could also be done by moving a horizontal scan line along the Y axis, with the X value mapped to pitch.

Navigate

Users can navigate the scatterplot in a relative or absolute style, similar to map navigation.

Relative navigation involves movements in the left/right/up/down directions. In a scatterplot that has real number (non-integer) values for the X axis or Y axis, it is not practical to use fixed-distance incremental movements because there may be an infinity number of dot distance values. Instead, each movement shall go to the next valid X or Y value. Relative navigation can combine two modes. In the scan line mode, pressing the left/right arrow key moves a vertical scan line among valid X values along the X axis, each time causing a group of dots with the same X value to be played. Pressing the up/down arrow key moves a horizontal scan line among valid Y values along the Y axis, each time causing a group of dots with the same Y value to be played. In the dot mode, pressing the left/right arrow keys move from one dot to the adjacent dot with the same Y value and the next valid X value. Pressing the up/down arrow keys move from one dot to the adjacent dot with the same X value and the next valid Y value. If two dots completely overlap each other, the navigation may enter a loop which needs to be broken. The dot mode supplements the scan line mode and allows the examination of individual dot. The dot mode alone may not cover the whole scatterplot. Relative navigation reveals relative

dot positions but does not convey their distances. As a result, multiple different scatterplots may produce the same auditory display. For example, the two scatterplots in Figure 7-1 are different but will produce the same navigation and sound effects. Sometimes, the problem can be remedied by using absolute navigation described below. A more complete solution is offered by the dot aggregation method.

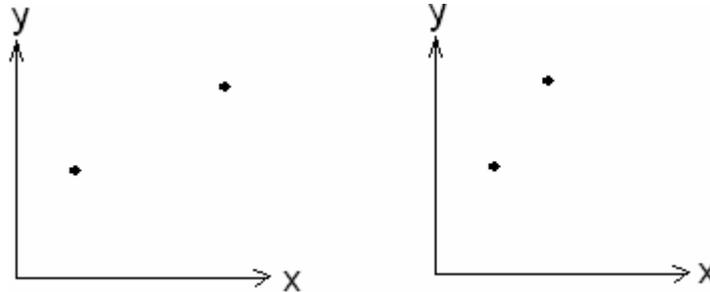


Figure 7-1: Different scatterplots may produce the same auditory display in the dot scanning relative navigation methods.

Absolute navigation refers to dividing the 2-D space of a scatterplot into 3x3 ranges and using the number pad to activate the gist of each range. Recursive 3x3 space partition provides coarse spatial localization that helps realizing the distribution of dots in the 2-D scatterplot.

Zooming is important for detailed exploration of a scatterplot with overlapping dots. Users could zoom into each of the predefined 3x3 ranges, or zoom into the neighbor of a dot. The relative navigation and the 3x3 exploration methods could further be used within the zoomed range.

Filter

Filtering could be done by launching a filter control (same as the one in Chapter

4). Data items that do not meet the query criteria are marked filtered in the data collection (such as the shared *DataSet* in Chapter 5 and reflected in the scatterplot. Users could switch focus among all data items and unfiltered data items.

Information level and Details-on-demand

An auditory scatterplot could also provide multiple information levels. For example, the feedback of each dot could include just the Y value pitch, or pitch plus the ID of the data item (if it is applicable, such as region names in the geo-referenced data). A coefficient value in speech could also be given with the gist.

Select, situate and brush

Users could press any key to stop at the dot currently being played. The current dot can be selected/unselected by a keystroke. Users could switch focus to only selected data items. The message for the situate action could include the two variable names, the number of dots, the current navigation mode and position, the number of dots selected, and so on. Brushing among the scatterplot and other data views can be done similarly as the brushing between the table view and the map view.

7.1.2 Dot aggregation

While information of individual dot can be interesting in a scatterplot, sonifying every single dot could become overwhelming for scatterplots containing a large number of data items or many overlapping data items. In a scatterplot, the spatial distribution of dot densities decides the actual correlation strength. The dot aggregation method spatially clusters nearby dots and sonify the scatterplot like a 2-D heat map. Based on the dot aggregation method, I implemented and integrated into iSonic a scatterplot view with

highly synchronized visual and auditory displays. The scatterplot is coordinated with the table and map views, and allows keyboard only interactions.

Figure 7-2 is a bivariate scatterplot showing the correlation of “average household income” and “employment rate” for 24 Maryland counties. This simple scatterplot clearly illustrates the spatial clustering. 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 (Figure 7-3). This spatial clustering and binning method is also fundamental in calculating “entropy”, a measurement strongly advocated by MacEachren et al. [108] for analysis of 2-D scatterplots as opposed to correlational coefficient.



Figure 7-2: A bivariate scatterplot showing the correlation of “average household income” and “employment rate” for 24 Maryland counties.

The heat map is sonified and explored in a similar way as a choropleth map, except that now it is a regular grid which makes the design easier. The value of each grid cell is mapped to pitch, with a higher pitch presenting a higher value. The current implementation uses 10 pitches from an increasing musical scale of CEGBCEGBCE starting from the middle C. For a given pair of variables, the lowest and highest heat values are mapped to the lowest and highest pitches. The mapping stays the same for all partial scatterplots during the exploration such as zooming or viewing selected regions only. Cell positions are mapped to sound positions.

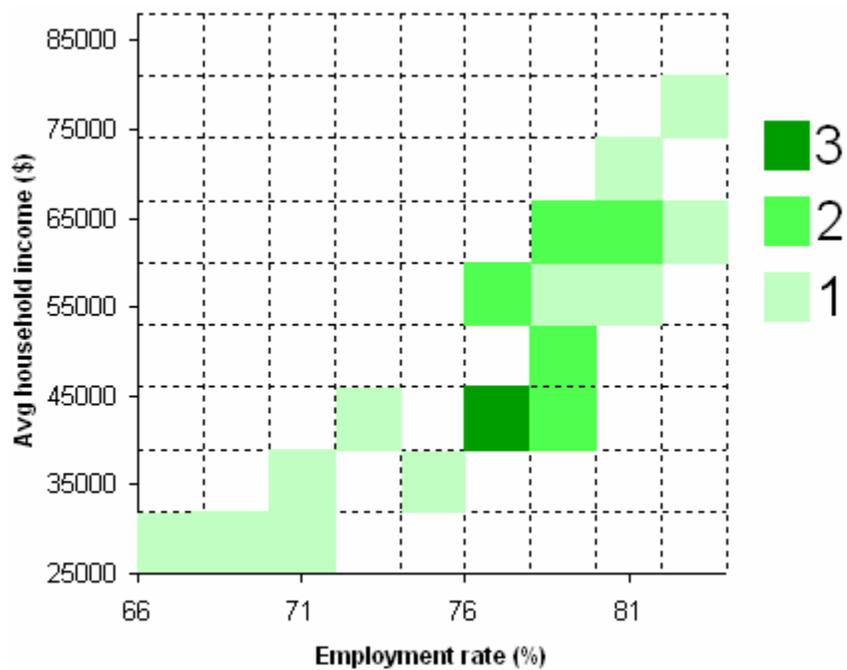


Figure 7-3: A heat map at the resolution 9x9 grid for the scatterplot in Figure 7-2. The color indicates the number of dots in each grid cell. A darker green indicates a higher value.

Gist

A gist of the scatterplot heat map is produced by spatially sweeping the grid cells. There is a tick mark sound at the end of each sweep row/column, and an end-of-sweep sound upon completion. 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 either add a pause or a background sound to the gist. Users can press the ‘*’ key on the number pad to switch among horizontal sweep and vertical sweep. Figure 7-4 (a) illustrates a horizontal sweep including empty cells for the heat map in Figure 7-3.

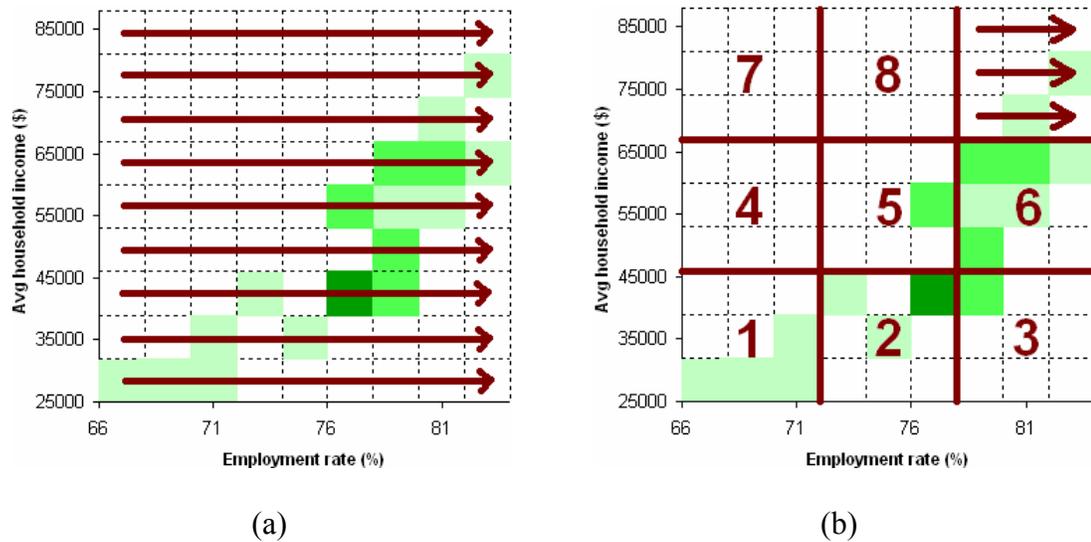


Figure 7-4: A gist of a scatterplot heat map and 3x3 exploration: (a) a horizontal sweep of a heat map, with each empty cell contributing a pause of a fixed duration to the gist; (b) divide the same heat map into 3x3 ranges and press 9 to sweep one range.

Navigate

A scatterplot heat map can be navigated in a relative or an absolute style. In the relative style, users press the four arrow keys to move among adjacent grid cells, activating the corresponding sound of the target cell. It is important to provide a sound feedback for an empty target cell, as according to the lessons learnt from earlier user evaluations (see Section 4.1.3.3) Users can also navigate the heat map in an absolute style by using the number pad to activate a sub-gist of each of the 9 ranges (Figure 7-4 (b)).

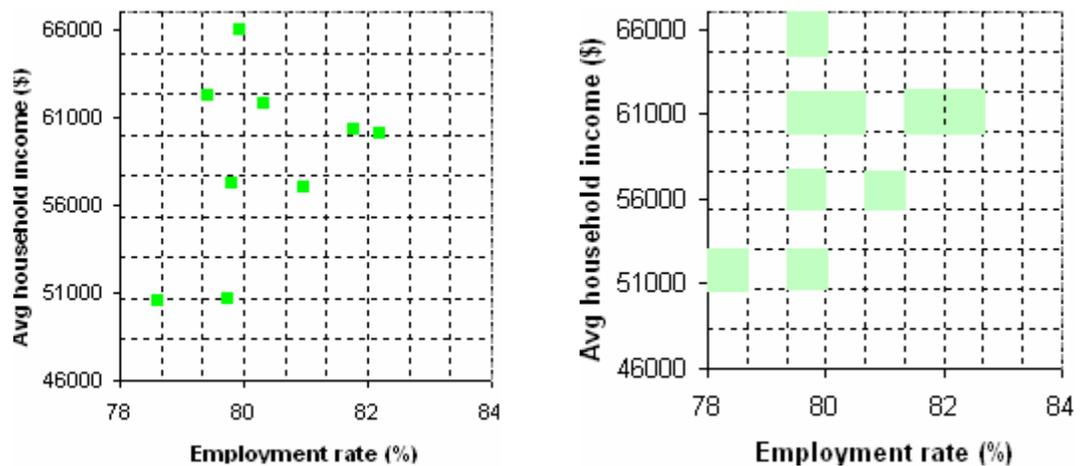


Figure 7-5: 3x3 zooming: (a) zoom into range 6 of the scatterplot in Figure 7-2; (b) the 9x9 heat map for the partial scatterplot in (a).

Zooming allows in-depth focused examination of specific parts of the scatterplot. Upon zooming, the heat map disaggregates. For example, Figure 7-5 (a) shows zooming into range 6 of the scatterplot in Figure 7-2. The disaggregated heat map is in Figure 7-5

(b). Figure 7-6 (a) shows zooming into the grid cell containing 3 dots in Figure 7-2. The disaggregated heat map is in Figure 7-6 (b). The relative and absolute navigation methods described above can be recursively applied to the zoomed range.

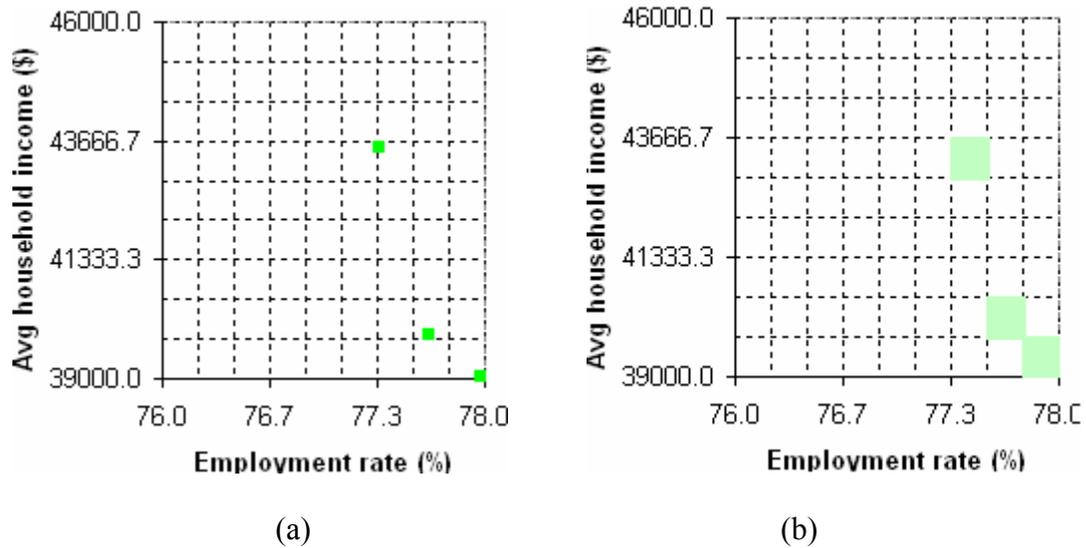


Figure 7-6: Zoom into a scatterplot heat map cell: (a) zoom into the 3-dot cell of the scatterplot in Figure 7-2; (b) the corresponding 9x9 heat map.

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 'space' to request details, the X and Y value ranges of the current cell are also given besides the information given at level 3.

Select

Selection is applied to individual grid cells in the heat map. When a cell is selected/unselected, all dots (geographical regions) inside that cell are affected. Users can press ‘A’ to switch among viewing “all regions” or “selected regions only”. When users view “selected regions only”, unselected dots fade into gray (e.g., **Figure 7-7**), and a new heat map is calculated to count only selected dots (e.g., **Figure 7-7**). This is useful especially when users select some regions in another data view and want to examine only those regions in the scatterplot view (see *Brush and coordination with other data views* below). It is also usually possible to select a single dot in a multi-dot cell, by zooming into that cell to get a more scattered view. The dot aggregation sonification method can also be combined with the dot scanning method as two sonification modes.

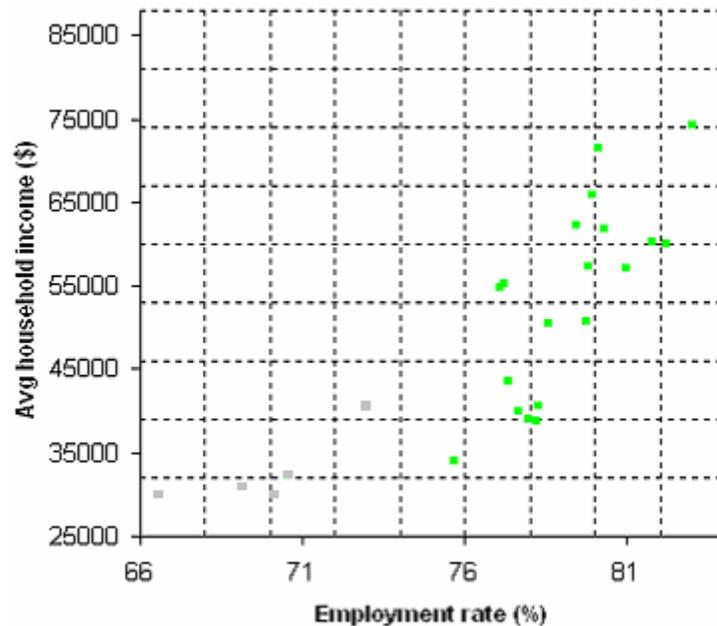


Figure 7-7: Unselected dots fade into gray when users choose to view “selected regions only”

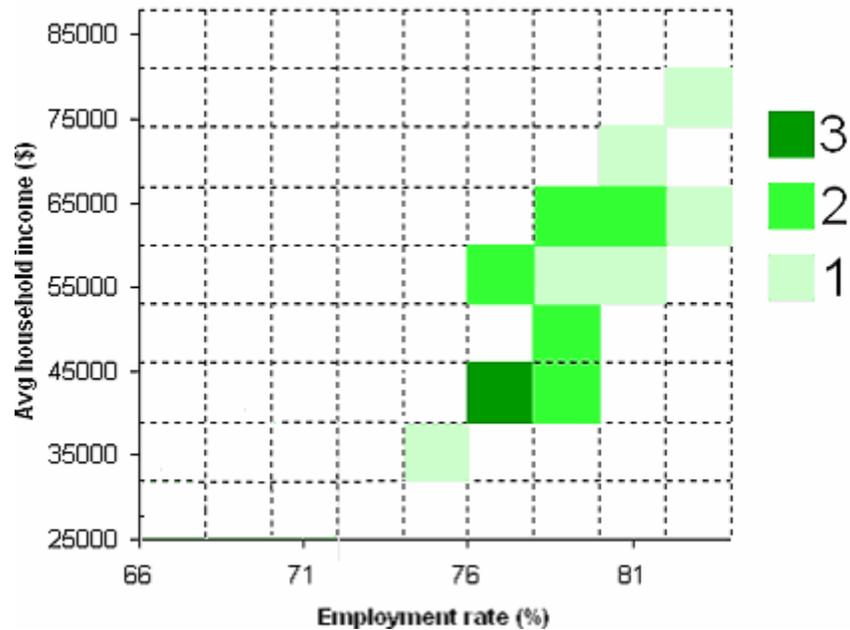


Figure 7-8: Heat map for the scatterplot in Figure 7-7.

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 yet to be implemented.

Brush and coordination with other data views

The scatterplot view is coordinated with the table and map views. Users can change the X and Y variables of the scatterplot view by pressing Shift+L to select 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. When users choose to view “selected regions only”, the scatterplot view uses a heat map that only counts the selected

regions. Unselected regions are grayed out. This allows users to focus on the bivariate correlational relation of the selected data subset.

7.2 Line graph

A visual line graph is often used to show the value change of one (or more) cases for one variable 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. Brown and colleagues did a series of studies investigating the design and human recognition ability of line graphs containing one or two data series [31][32][33]. In their work, each Y value was mapped to a musical pitch, with a higher pitch representing a higher value. Two data series were fed separately to the two ears, one series to the left and another to the right. Users could use a keyboard to browse the line graph in a parallel mode and a sequential mode. They were asked to locate the position where the two data series intersected, and to locate the positions with the maximum and minimum Y values. It was found that the parallel and sequential modes were useful for different types of tasks, with the parallel mode making the intersection task faster.

Analysis of Brown's line graph sonification shows that the keyboard interactions supported in the design were to allow users to obtain line graph "gists" and to "navigate" among data points along the X axis. After describing how "gist" and "navigate" were done in Brown's design, I argue that users could largely benefit from additional support of other Auditory Information Seeking Actions.

Gist

In Brown's line graph, users could hold down the 2 key (used as the down arrow key) on the number pad to rapidly move along the X axis to hear a quick overview of the graph (gist). In the parallel browsing mode, each data series generated a subgist. The whole graph gist was a combination of the subgists of the two data series, one heard through the left ear and the other through the right ear. In the sequential mode, users pressed the 4 and 6 keys (used as left/right arrow keys) to switch among the two data series. Only the subgist of the current data series was heard when the 2 key was held down.

Navigate

In Brown's line graph, users could navigate along the X axis and navigate between the two data series. In the parallel mode, users pressed the 8 and 2 keys (used as up/down arrow keys) to move along the X axis and play the two Y values simultaneously. In the sequential mode, users pressed the 8 and 2 keys to move along the X axis to only play the Y value of the current data series. By pressing the 4 and 6 keys (used as left/right arrow keys), users could switch between the two data series.

Information level and Details-on-demand

In Brown's line graph, users could press the 5 key on the number pad to repeat the musical pitch(es) corresponding to the Y value(s) of the current data item(s). It was in a sense a details-on-demand action. However, there was only one information level (pitch only) available so such an action did not provide any extra information about the current data item. Users could benefit from having multiple information levels. For example, there could be another information level giving the spoken value of the current data item.

Compared to pitch, speech can give more accurate information about the values. To find the maximum or minimum value, users can use the pitch only information level to quickly navigate through the data items to narrow down to a few candidates, then use the information level with spoken values to accurately judge if the value is actually the maximum or minimum.

Select

Brown's line graph did not support selection. However, from the user evaluations, Brown and colleagues realized that providing a position saving feature could improve the accuracy of the maximum/minimum position finding task. By allowing users to save positions, they can return to a few interesting locations at any time without moving through the rest of the graph. This essentially can be described as "selecting" to isolate some interesting data items from the rest for focused examination.

Situate

In Brown's line graph, users did not get any auditory feedback about their navigation position along the X axis. Two data series were fed separately to the two ears, so the navigation position could not be conveyed through the sound stereo panning effect. Except for counting the key strokes, it was hard for users to know how far they had moved along the X axis. They could have benefited from a "situate" message that speaks information such as the X value of the current navigation position.

Filter and brush

Filter can easily be added to line graphs in a similar way as to the table and map views. A line graph can also be integrated with other data views. Change of line graph variables can be done through selecting variables in the table view, just like changing the

scatterplot variables.

7.3 Implementation

As shown in Section 7.1 and Section 7.2, it was easy to systematically draw the design blue print, guided by the ADC framework. Many interface designs for choropleth maps can be directly applied to scatterplots after dot aggregation, such as the sound-encoding schemes, spatial sweep, 3x3 recursive 2-D space exploration, and arrow key navigation, filtering, and so on. Of course, the effectiveness will need to be examined through user evaluations.

In terms of the actual implementation of the scatterplot view, it was quick to setup the modules and functional call relationship by duplicating Java classes from the choropleth map implementation. Figure 7-9 shows the extension to the diagram in Figure 5-1 to incorporate the scatterplot view. Due to the high similarity of choropleth sonification and scatterplot heat map sonification, much code from the map view can be reused in the scatterplot view without much modification. One part of the main effort lies in the design and implementation of data structures to generate, store, and maintain dot coordinates and semantically zoomable heat map representations of the scatterplot (class *ScatterplotInfo*). The other part involves modifying the code in the duplicated classes to visually render the scatterplot using the Piccolo toolkit [19], and to produce auditory feedback by calculating *SoundingObjects* according to the information in *ScatterplotInfo*.

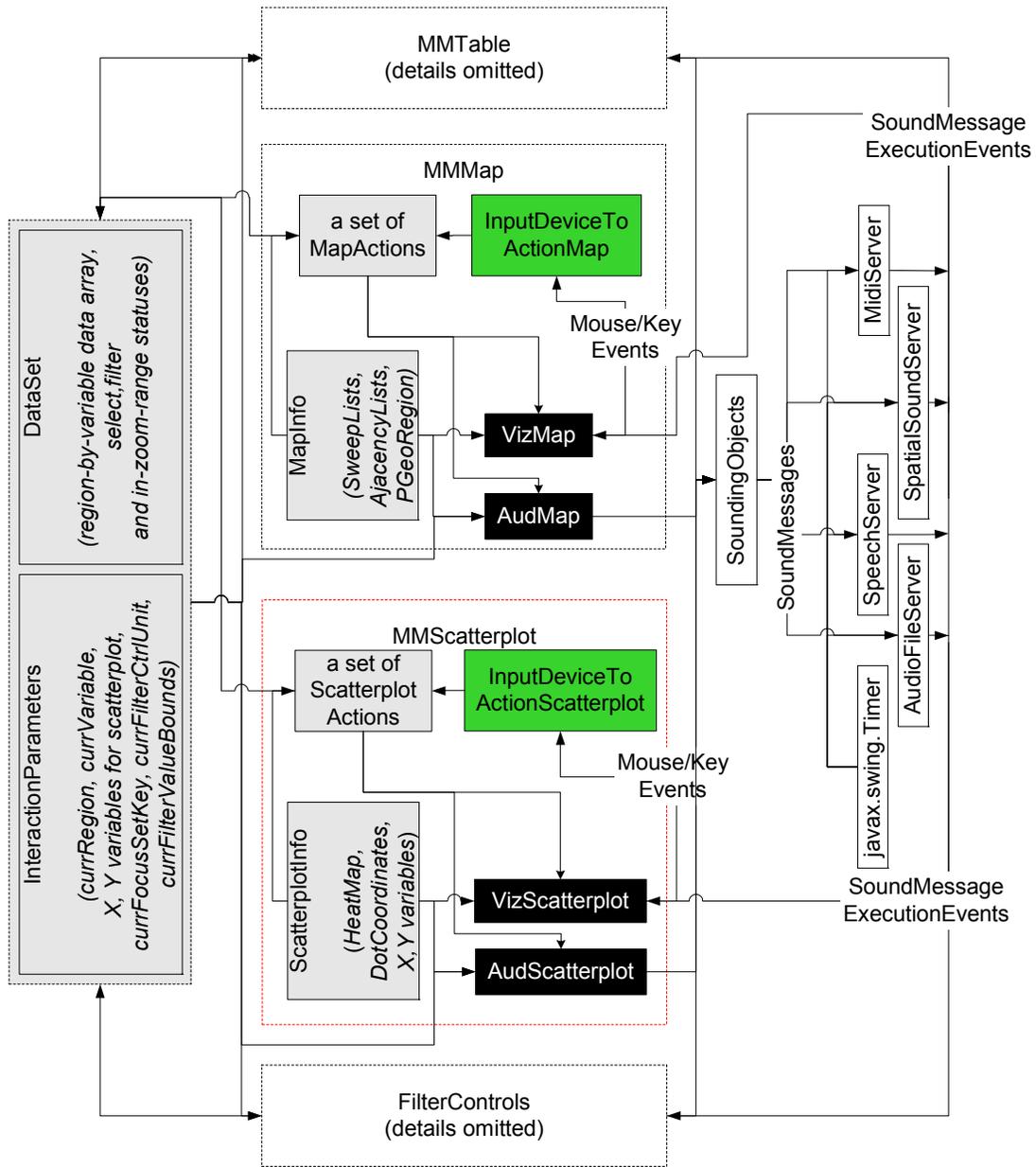


Figure 7-9 : Extend the diagram in Figure 5-1 to incorporate the scatterplot view. Much code from the map view can be used in the scatterplot view with minor modification due to the high similarity of their sonification designs.

Line graphs are yet to be implemented in iSonic. For a line graph containing one data series, the sonification design in Section 7.2 generates an auditory display that is essentially the same as the auditory display generated in the table view when users sort the column containing the X variable and play the column containing the Y variable. Therefore, a good portion of the code from the table view implementation could be used with minor modification to implement a line graph containing only one data series. One major effort would be to visually render the line graph and to synchronize the visual display with the auditory display. Another major effort is to extend it to show multiple data series.

7.4 Discussion

Previous work has shown that sonified scatterplots and line graphs were quite effective in conveying trends and correlation relations. Only a very limited set of AISAs was provided in previous work. I showed that those existing sonification designs of scatterplots and line graphs could be extended to incorporate other AISAs. Many design options exist for each type of graphs. One scatterplot sonification design was implemented and integrated into iSonic. User evaluations will be necessary to examine the effectiveness of the designs. By allowing a set of unified interactions in each auditory graph and among multiple graphs, those graphs can be integrated as multiple data views into a unified auditory work space for analytical data exploration.

As with any scientific studies, the generalizability and validity of the results in this research need to be rigorously tested by evaluation with a broad range of users and applications. Chapter 6 of this dissertation shows the validity of the ADC framework with

highly educated users with no residual vision. The effort in this chapter demonstrates the generality and flexibility of the ADC framework and the utilities of the AISAs. Tables, choropleth maps, scatterplots, and line graphs are four familiar spatial displays, and there are others that might also be accommodated with modest effort, such as matrices, tree diagrams, networks, and maps with point landmarks like cities. Similar interactive sonifications could broaden the capabilities and job prospects for many blind workers.

Chapter 8

Contributions and Future Work

8.1 Contributions

While information visualization has empowered users with normal vision to think, the ability of visually impaired users to discover information in complex data collections is limited by the nature of the text-based speech accommodation and the lack of appropriate tools. In spite of the long existing and increasing interest and effort in using sonification for data presentation, the support for task-oriented user interactions that are essential for exploratory data analysis, is typically missing.

This dissertation offers an auditory interface design framework and a set of interactive sonification designs in a geo-referenced data exploration tool called iSonic.

Major contributions include:

- (1) An Action-by-Design-Component (ADC) framework to guide auditory interface designs and evaluations for exploratory data analysis. The framework characterizes task-oriented data interaction in the auditory mode as a set of Auditory Information Seeking Actions (AISA): gist, navigate, situate, search, filter, details-on-demand and adjust information level, select, and brush. The framework also contrasts AISAs with visualization actions, and highlights general design principles of a set of Design Components to support AISAs.
- (2) Validate the ADC framework by applying it to geo-referenced data. Analogous to many standard techniques used in visualization, the set of Auditory information

Seeking Actions (AISA) were impressively effective for even blind users to accomplish complex exploratory tasks on geo-referenced data, using highly coordinated table and choropleth map views. By showing how AISAs can be applied to other graphs (scatterplot and line graph), the research encourages others to support unified user actions (AISA actions) in their auditory graphs, and to tightly couple the graphs as multiple data views in an auditory workspace.

- (3) Interactive sonification designs and guidelines for tables, choropleth maps, and scatterplots. Guided by the ADC framework, the design space of geo-referenced data was investigated, in the dimensions of data view, navigation structure, auditory feedback encoding, and input device. Some effective design options were identified and insights were obtained regarding human ability to perceive complex information, such as those with spatial structures, from interactive sounds.
- (4) A tool (iSonic) for blind users to effectively explore geo-referenced data in highly coordinated table view, choropleth map view, and scatterplot view, without special interaction devices. Allowing flexible configuration of a rich set of design options, iSonic is also used as a research tool to investigate various design options. This dissertation also describes preliminary algorithms to address the implementation challenges of automatically generating geographical region spatial sweep orders that are congruous to visual impressions for arbitrary maps.

Readers can try iSonic or viewing the video demonstration by visiting

www.cs.umd.edu/hcil/iSonic

8.2 Future work

However, this research is just the beginning. It opens new possibilities for designing interactive interfaces for visually impaired users to better understand complex data collections.

Integrate More Data Views

This research showed that blind users benefited from a rich set of task-oriented Auditory Information Seeking Actions (AISA) with multiple highly coordinated data views. I investigated in-depth how to support AISAs in the sonification of simple tables and choropleth maps, and integrated them as two data views in the iSonic data exploration tool. I also analyzed previous work on scatterplots and line graphs, discussed how AISAs can be supported in such graphs, and demonstrated some design options through the implementation of scatterplot and its integration into iSonic. There are many other researchers who are interested in graph sonifications. The goal was to encourage other researchers to provide unified user actions (AISA actions) in their auditory graphs, using the ADC framework as a design guide and in turn to validate and refine the framework. The future steps towards the goal of building a multi-view auditory workspace include investigating and implementing more auditory graphs, generalizing iSonic into an environment with plug-and-play APIs that can easily integrate multiple auditory graphs as well as other types of data views that support AISAs.

Another approach to build a multi-view auditory workspace is to automate the generation of auditory displays. Mackinlay [109] showed that static graphical displays can be automatically generated for any relational data. His work is not about interaction.

To automate an auditory workspace with multiple interactive auditory graphs, we need a thorough understanding of effective data to sound mapping designs. Previous psychoacoustic and sonification research have yielded many valuable insights into effective sound designs for single or multiple variables. Future work needs to investigate how those results would apply in various graphs, especially in the context of user interaction, such as the AISA actions in the ADC framework. My work on map and scatterplot show that some graphs share common effective auditory display and interaction designs. With a thorough understanding of a wide range of interactive auditory graph designs, we may be able to extract and defines rules that can be used to automate generation of interactive auditory graphs.

Refine iSonic Implementation

Upon startup, iSonic checks the parameter values in the configuration input file that decide its behavior. Users can modify some of the parameter values during the runtime by launching a configuration panel. Currently the panel allows adjustment of speech rate and volume, the mapping scheme such as what musical instruments to use, sound duration, loudness, and so on. In the future, the panel shall include the adjustment of more parameters from the configuration file, and let users save new configurations into files.

iSonic currently produces speech by sending command network datagram to a speech server that is built on Microsoft Speech API (SAPI) 5.1 and runs as a separate Windows applications. As a result, iSonic relies on Windows to run, although the rest of iSonic is written in Java. A full Java implementation is possible with the merge and maturation of open source speech synthesizers written entirely in Java, such as the

javax.speech.synthesis package.

It is difficult to define a spatial sweep order and region adjacency navigation matrix satisfactory to everyone. It is even more challenging to automatically generate them for any map. I described some preliminary algorithms used in iSonic for these problems (Section 5.4 and Section 5.5). They generated reasonable results for maps I have tested. Better algorithms may exist and the problems by themselves are interesting research challenges.

There are also some commonly requested interaction features that need to be implemented in iSonic. Examples include a “search” function and the ability to “select” a group of items.

Scale Up the Design

Auditory perception is sequential and has much coarser granularity than visual perception. The size of the data collection people encounter, without size reduction, often exceeds the amount of data items that can be effectively recognized through audio. Auditory interface designs need to be able to scale up to large data collections. Data aggregation, combined with zooming, is often necessary in auditory display, just like data aggregation is a common solution for occlusions and clutter in visualization. The choice of aggregation methods will depend on the nature of the data and the task. iSonic currently supports spatial zooming, but does not do data aggregation.

Another dimension that needs to scale up is the range of values. As discussed in Section 4.6, the range and number of easily discernable pitches are limited for most human. This is also true for other sound properties, such as duration and loudness. To map a wide range of values to musical pitches, categorization is necessary.

Categorization hides the difference of values in the same category which can only be revealed by other means such as speech. iSonic currently divides numeric values into 5 categories by default, as this is a common practice for choropleth maps. But the number of categories can easily be changed through the configuration file. To fully utilize the discernable pitch range, the value to pitch mapping can be rescaled upon zooming, i.e., when users zoom to focus on a subset of data items, the new narrower value range can be remapped to the whole available pitch range instead of the original sub-range, possibly revealing more value differences. Appropriate interface hints must be provided upon rescaling to prevent misinterpretation and disorientation.

More User Evaluations

Besides replicating existing user studies, many other user evaluations would be interesting.

In the case studies, I only tested users without any residual vision who were highly educated. Further study with partially sighted users or users with lower education levels may reveal different usage patterns and visual-auditory interactions that may modify the results and framework. An average of 6 hours' use in an administered experimental setup was not enough to go beyond the novice usage stage. Investigation of iSonic's long-term use in real work circumstances will provide further understanding. I already provided iSonic on the Web with the real data and maps that one of the blind users requested for her work. Follow-up examination of her usage experience will provide further understanding. The self-paced auditory step-by-step lessons were overall very effective, but many minor clarifications and adjustments to the content could make the online help more pleasant to use, such as to add more explanation so users do not get

confused about the name of certain keys due to the unsatisfactory speech quality commonly present in many speech synthesizers.

Although the result of the first controlled experiment (section 4.1.2) did not indicate that the use of non-individual HRTF spatial sound helped with subjects' pattern recognition performance, I expect that spatial sounds, especially individual HRTF spatial sounds, may enhance listening realism and improve users' subjective satisfaction. Several administrative concerns exist in measuring the individual HRTF for a large number of subjects, especially for blind subjects. For example, current HRTF measuring techniques involve inserting earphones or speakers into the subject's ear canal. People with visual impairments may be much more concerned than sighted people about having any foreign object inserted into their ears by non-medical professionals. It is also difficult for them to keep their heads steady enough for an extended period of time in order for the measuring process to continue. Fortunately, new techniques are coming out that would significantly reduce the length of the measurement time [205]. Once individual HRTFs are obtained for a large number of subjects, it would be very interesting to quantitatively compare the use of spatial sound to stereo sound.

To minimize hardware dependence, this research focuses on supporting effective interaction with a standard keyboard, although some other common pointing devices, such as touchpads, are also investigated. Observations of the two blind consultants and partial results from the controlled experiment on the effect of input device and user blindness (see Section 6.1.1) showed that both keyboard and touchpad interactions in iSonic are effective. However, users seemed to enjoy the touchpad interaction more, and stated that they would invest in a \$150 touchpad if such a device can prove to be effective

for other common purposes besides exploring maps. The decision on whether to use a special piece of hardware shall be based on cost/benefit studies. In the short term, systems that use widely available hardware, such as a standard keyboard, are more likely to bring immediate benefit to a large number of users. In the long term, researchers shall expect more opportunities brought by advances in hardware.

It will also be interesting to compare iSonic to systems that use a tactile approach. For example, VTPlayer is a product by VirTouch that makes use of a special mouse to deliver a tactile sense of the screen to blind and low vision users (Figure 2-2). The device costs hundreds of dollars and there has not been any established evidence on its effectiveness. It will be interesting to investigate the benefits and limitations in using such a device compared to iSonic which is a pure audio-based solution and does not rely on special devices.

Appendix A - 1: Description of Case One (P1)

Profile

P1 was a male junior college student in his late 30s who was also working as a health insurance specialist within a U.S. federal government agency. He was diagnosed with Retinitis Pigmentosa at 13, gradually lost his vision over the years, and only had some light perception at the study time but with no other disabilities. With limited Braille skill, he mainly relied on JAWS for Windows for a lot of Windows applications daily at work and home. Keyboard was the only input device he uses, and the number pad was mostly used as JAWS navigation keys. He had very limited experience with tactile maps, but reported to have good knowledge about the U.S. state map and Maryland county map. He often needed to read documents in tables, but did not like tables because they were often not accessible. He used Excel on a limited basis but was very capable of basic Excel operations such as navigating cells and sorting columns.

Task completion and error causes

P1 had no problem getting started with iSonic by launching the online help to follow the self-paced auditory lessons. The training was very effective, allowing P1 to achieve 100% task success rate in the testing. One road block encountered in the training was that regions selected in a previous task remained selected and became mixed with regions selected in the next task. The training lessons did not give explicit instructions to clear previous selections before the start of a new task, so P1 was not aware of the persistency effect among tasks. With the experimenter's hint, P1 immediately developed a habit of

using the 'i' key to query the interface status and clearing all previous actions at the beginning of each task.

Except for the correlation task that took P1 quite a few minutes to understand the concept and the sample solution in the training, it was not difficult for P1 to come up with efficient task strategies using the appropriate interface features. While reporting the key mapping to be intuitive and easy to remember (8 on a 10-point scale with 10 being easy), P1 reported spending 40% of the effort recalling the hotkeys and believed he could do a lot faster once becoming more familiar with the mapping.

Task strategies and iSonic usage patterns

Map vs. table:

P1 used the table for most value comparisons, and used the map when he needed the region location and adjacency information. "I enjoyed using the table a lot. Selecting and ordering in the table are very handy." Table sorting was not only used to find minimum, maximum, or top 10 in a sequence of values, but also to compare value pairs with the same pitch. Although P1 reported to have fairly good knowledge about the maps used in the study, he said the map view made things easier. During the study, right after accomplishing task T6 with the map view, he was asked to do the same task but with another variable "population" and without the map view. Although he just reported the three adjacent states in the previous task, he was not able to remember them and had to use the map view again. This indicated the usefulness of the map view even for users who were already given chances to acquire map knowledge. "(With the table), I would need a column to tell me where each state is. Even when such data is available, the tasks will still be hard without the map". P1 also enjoyed exploring the map and spent

extensive time during the training to find and listen to the facts of some individual states.

Brush:

P1 was very comfortable selecting items in each view and switch between the two views during the task. While P1 had a big doubt about the usefulness of iSonic in the beginning, he became very excited after he tried the brush feature with the highly coordinated views. “Now I see the value of this tool. I can use both the table and the map, select stuff in the table and show them on the map, or select stuff on the map and look at them in the table.” “It is cool to use the table to control the map, showing things on the map when I need to show what I found to other (sighted) people”.

Pitch:

P1 reported that the use of pitch made it a lot easier and quicker to compare values. During tasks, P1 frequently adjusted the information level. In default, he used the level with the pitch and the region name. To focus on value comparison, especially for the correlation (T3) and geographical pattern (T7) tasks, P1 turned off the name. When the pitches of the values to be compared were the same, he increased the level to include the value in speech as well. P1 was also very comfortable using only the pitch to tell if a value was in the highest value range. P1 almost never used the details-on-demand command but relied on adjusting the information level instead, and never used the “name only” information level.

Panning:

P1 reported not having paid attention to the sound panning in either the map or the table view.

Gist:

P1 was able to handle the two table navigation modes. Most of the time, he used the cell mode. But for the correlation task, he switched to the row/column mode to activate the gist of each column involved in the task. In the map view, gist was used a lot to read the regions automatically in order to find a specific region on the map. P1 considered it hard to make sense of a spatial sweep of the whole map with only the pitch when he was not familiar with what the map looked like.

Navigate:

Table cell navigation was very intuitive for P1, but extra explanation was needed during the training on the behavior of the row/column mode. P1 became comfortable with both modes but preferred to use the cell mode.

For the map, P1 frequently alternated between the 3x3 method and the arrow key relative navigation. The previous/next navigation method was never used. The 3x3 exploration was used to narrow down to a map range to find a specific region, as well as to find out what regions were in each part of the map. P1 understood the spatial layout meaning of the 3x3 method but did not automatically connect the concept of “west” to the three numeric keys (1, 4 and 7).

P1 was able to zoom in/out on the map and be aware of his zoom situation. He was also very comfortable with the arrow key navigation in a zoomed range. One zooming-related problem was encountered when P1 tried to find the adjacent states of Kansas but was not able to reach two of them after zooming because they were in a different range partition from Kansas.

Select:

P1 liked very much the ability to select and focus on specific regions without having “other things getting in my way”. When using Excel, P1 tried to copy a few regions out of the data sheet and paste them to a new sheet but did not succeed.

Situate:

P1 reported never to feel lost. He used the “situate” feature a lot more in the table than in the map “to get the current region name, table sorting status, and what elements are viewable”.

Menu:

P1 spent a lot of time going through each menu item during the training, and only used the menu once to recall the hotkey for a command.

Potential benefits

P1 was very enthusiastic in seeing iSonic appearing on government web sites and becoming a distributable application. “Although I’m quite familiar with many counties and states, geographical tasks are much easier in iSonic...iSonic lets me do trend analysis very easily. It’s not impossible in Excel with JAWS, but it’s very difficult. For finding specific information, although iSonic currently does not have search functionality, it is still a lot better because it lets me easily select and isolate items, and it’s overall easy to use.” “This tool will be good for research and data analysis, such as to analyze trends in diseases in the health care arena. It can be useful for all kinds of table data, such as Medicare data, disability data, and so on. For my current professional work, I avoid tables as much as possible because they are often not accessible or hard to make sense. My current information source is text information fed by other people. This tool allows me to drill down into the data and find interesting facts, and provide to other people.”

Appendix A - 2: Description of Case Two (P2)

Profile

P2 was a 23-year old male with a bachelor's degree in Criminology/Criminal Justice. He was born totally blind with no other disabilities. While fluent at Braille, P2 used computers extensively every day for both school work and leisure time entertainment, with JAWS for Windows as the screen reader. P2 was experienced in a wide variety of software from Microsoft Office Suite and media editing, to web browsing and file sharing programs. Keyboard was the only input device he used, and the number pad was rarely used. P2 was capable in reading tactile maps but rarely used them. Growing up in the state of Maryland, P2 only had some knowledge about the Maryland county map and fairly good knowledge about the U.S. state map. He was familiar with table browsing and Excel spreadsheets. As a design partner, P2 was very familiar with the value to sound mapping scheme and had extensively used the map spatial sweep and navigation, but had not actually used other AISA actions or operated the table view, the menu and the online help system before the study.

Task completion and error causes

Instructed to go through the training as everyone else, P2 launched the self-paced auditory lessons and quickly went through all iSonic features, reviewing old features and learning new ones. He was able to accomplish all but one training tasks without any hint, and achieved 100% task success rate in the testing. Unlike some subjects who needed a hint, P2 was clearly aware of the persistent effect of actions on the interface among tasks

and always reset all previous actions before beginning a new task.

Except for referring to the sample solution for the correlation task during the training, P2 combined the interface features to come up with efficient task strategies with ease (9 on a 10-point scale with 10 being very easy). P2 attributed most of his thinking time to recalling the hotkeys for features not frequently used, but still reported the key mapping to be intuitive and easy to remember (8 on a 10-point scale with 10 being very easy).

Task strategies and iSonic usage patterns

Map vs. table:

P2 was very comfortable (10 on a 10-point scale with 10 being very comfortable) with the concepts of map and table. He used the table for value comparison, used the map whenever the task involved geographical locations, and was very comfortable comparing items in each view. Although P2 grew up in the state of Maryland, he only knew the locations of a few counties at the beginning of the study. From using the map view of iSonic, P2 reported to have acquired more location knowledge that allowed him to make educated guesses for geo-related tasks in Excel. However, for each geo-related task, his reaction was “I need a map”.

Brush:

P2 was very comfortable selecting items in each view and switching between the two views during the task. During the iterative design process, P2 had always been enjoying playing with the map view but indicated that the practical help from the map component was limited without a musically-enhanced table view because he would not need to know region locations for many tasks. The map and the table views combined

made each other more useful.

Pitch:

“(Using musical pitches to present values) is a new thing and I needed time to adjust, just like other new things, but it is very helpful. It makes it a lot easier and faster to compare values”. While P2 mostly used the information level 3 by default, he said that he was often using the pitch to narrow down the answer, and ignoring the value speech. This was also revealed by the observation that he frequently moved from one data item to another before the value speech of the item even started. “When I need to confirm the exact value, I wait for the speech to finish”. To play a gist such as the map spatial sweep, the information level 2 was used instead to speed up the playback. The details-on-demand command was never used since it was the same as the information level 3. The information level 0 (region name only) was also never used. P2 reported not to have experienced any interference problem from the simultaneously presented speech and pitch.

Panning:

P2 reported to have noticed the stereo panning of sounds in both the map and the table views. “It is very helpful in map. I used the sound location to judge the region location. It may not be necessary in the table. I didn’t use it”. In one task where P2 needed to report all the selected counties that were in the west, he activated a map spatial sweep to play the five selected counties one by one (at information level 2), and immediately reported the counties in the west by judging from the sound location. Combining the sound location and counting steps in the arrow key navigation, P2 was able to report the approximate layout of regions on a new map with good accuracy.

Gist:

P2's experience with whole map spatial sweep showed that gist on maps could be of tremendous value, especially with an appropriate initial training and practice. When P2 first encountered the spatial sweep of U.S. state map in earlier design iterations, he was able to accurately tell 5-category value trends related to left-right (corresponding to west-east on maps), but not up-down (corresponding to north-south on maps). After using a tactile map to go through the spatial layout of U.S. states, he was able to tell 5-category value trends related to up-down with an amazing accuracy. The experience seemed to have given P2 a lot of confidence in using gist to quickly tell the value trend on maps, even for new maps he never encountered before. In the post-test free exploration of an unknown map (Idaho county map), he first used the arrow key navigation combined with sound location to sense the overall region layout, then listened to a whole map gist several times at the pitch only information level, and report the overall distribution pattern of the highly populated counties. Understanding the gist of a whole map was considered at the easy end at the time of the study, but P2 said some adjustment time was necessary to quickly react to the meaning of different sounds, such as the value pitch, the row-switching sound, and the end of sweep sound, in such a rapid playback.

Navigate:

P2 showed no problem with either of the table navigation modes. He reported to understand the two modes very well, and only used the row/column mode in the tasks.

During the training, P2 played for quite a while with the 3x3 map navigation, recursive zooming, and arrow key navigation within a zoomed range. He showed no problem with any of those navigation methods, and reported them to be very easy to use

(10 on a 10-point scale with 10 being very easy). During the test, P2 only used the arrow key navigation and reported to have a good sense of his wherebeing on the map. “I don’t need to use the others today because the sound location already gives me plenty of information”. The previous/next navigation method was never used after the training.

Select:

P2 said it was very easy to toggle region selection status and circulate interest among sets of region with different selection status. “Selecting lets me view only a certain part of the data, it makes comparison faster”.

Situate:

P2 tried the “situate” feature while playing with the map zooming during the training, but never used it in the test. He had a habit of resetting the interface, such as unselecting all regions and canceling all sorting, before starting a new task. “I usually reset everything then start a new task. I remember everything so I didn’t use ‘i’ (to get the status information). I never felt lost”. In the contrary, P2 had trouble keeping track of and inquiring status in Excel, such as how he had sorted the table or whether the table was successfully sorted by what he specified in the sorting dialog box. Excel did not have a “situate” function for centralized status inquiry. It also lacked hotkeys to directly clear sorting. While in-place hotkey sorting was easy in iSonic, P2 needed several trials to cancel sorting or achieve multiple sorting using the Excel sorting dialog window.

Menu:

P2 went through the menu system during the training and never used it later because he already remembered all the hotkeys needed for the tasks.

Potential benefits

P2 was looking for a job at the study time. He said the tool could help him do research on statistics, most likely data related to law and criminology. P2 never thought it could be so easy to related data of employment, disabilities, education, and so on and look at them this way. He also said that it could help him personally if he was moving to a new location and wanted to compare the facts of different places. P2 wished he had such a tool to look at map-based data when he took his geography class.

Appendix A - 3: Description of Case Three (P3)

Profile

P3 was a male in his late 30s who had a Jurors Doctorate degree and worked as a director at the National Federation of the Blind. He was completely blind since birth and had no other disabilities. While fluent in Braille, P3 used JAWS for Windows daily for computer based activities such as browsing the Web, reading and editing Microsoft Word documents, organizing and processing information for management duties with Microsoft Outlook. P3 was comfortable navigating simple tables and Excel spreadsheets. When it came to complex tables, he either printed them in Braille or had his sighted assistant read to him. But neither was satisfactory. “Large tables are often too big to print in Braille. Sometimes it is hard for others to explain a table in words”. P3 used a standard keyboard and frequently used the number pad as JAWS navigation and command keys. Although “it is often difficult to capture all the information shown on most (tactile) maps”, P3 enjoyed maps and had very good knowledge about the U.S. state map and the Maryland county map.

Task completion and error causes

Except for a hint to switch between the iSonic main window and the help window, P1 had no problem learning iSonic by following the self-paced auditory lessons. He was able to accomplish all but one training tasks without any hint, and achieved 100% task success rate in the testing. P3 was very explorative. He often went beyond the content of the training lessons, coming up with his own tasks and making new combinations of interface

features to see if they work. The results were always satisfying.

Except for the correlation task that P3 failed to accomplish during the training without viewing the sample solution, it was not difficult for him to come up with efficient task strategies using the appropriate interface features. P3 reported the key mapping to be intuitive and easy to remember (9 on a 10-point scale with 10 being very easy). When he paused during a task, it was mainly to decide the task strategy and the order of the steps, instead of recalling keys.

Task strategies and iSonic usage patterns

Map vs. table:

P3 had no problem understanding the concepts of map and table, and was very comfortable using both and switching between them according to task needs. He had excellent knowledge about U.S. state and Maryland county maps, but thought map was very helpful and necessary. “The map often tells me things that I don’t know or things that may not be that on top of my head”. When he looked for Dorchester on the Maryland county map, he recalled from the water sound that Howard and Dorchester were across the bay. For location related tasks, he much more preferred iSonic because “I can use both table and map here. The information is best displayed in map. There’s no way to explain what the map is like in Excel”. Although he was successful with the geographical value distribution trend task using Excel, he said “in Excel, there’s continuity problem of not having the information displayed as map”.

Brush:

P3 was very excited about the ability to select things in one view then look at them in another view. “The ability to change view quickly is valuable”. “If it’s just for

the table, Excel is a quite good software. The biggest advantage of iSonic is to switch back and forth between table and map.”

Pitch:

Using pitches to present values was a totally new method to P3. But he soon got used to it, found it helpful, and made increasing use of it in tasks. In the beginning, P3 mostly used the information level 3, but stayed with the information level 2 more and more after a few tasks. To find minimum/maximum values, he used pitches to check if the sorting was successful. “Pitch helps me to quickly skim through the values as a confidence builder”. He used pitches for the correlation tasks exclusively. P3 was able to tell if a value was in the highest value category. For pair-wise comparison tasks, he used pitch for comparison, and only turned on the numbers when the pitches were the same. To increase the benefit of pitches, P3 suggested that instead of dividing the values into 5 categories and mapping to only 5 pitches, he would like to have more pitches by expanding the upper and low bounds of the pitch ranges.

Overall, P3 mostly used the information level 2 (region name + value pitch) for tasks. “Different information levels are good for different tasks. For this task (correlation), I used pitch only”. P3 also used information level 1 (pitch only) to sweep the map for the geographical value distribution pattern tasks. “It’s faster without the region name”. “It works and it’s fun”. The “name only” information level and the details-on-demand command were never used.

Panning:

P3 reported that panning did not help much. “I didn’t pay attention to the sound position. Maybe it’s because the speech takes away the stereo panning”.

Gist:

P3 did not have problem understanding how the map sweep was done. “It’s easy to understand though a simple sweep does not give a full picture of the map because map is not a perfect grid”. “It will be my first step to get acquainted with a new map to get the big picture. Then I’ll need to explore with the arrow keys and the number pad”. In post-test free exploration of an unknown map, he used the table to display the population values on the map. After sweeping the map once with pitch only, he said that the high values were probably in the east part. He then used the 3x3 method and correctly reported that the west and the east were more populated than the middle and the very high values were in the west. Occasionally, map sweep was also used at the information level 2 with the arrow keys to find a state given the name.

Navigate:

P3 was capable of exploring the table in both navigation modes, but mostly used the cell mode, especially in the test. He used the row/column mode for the correlation and pair-wise comparison tasks in the training, but said he preferred the cell mode. “The row/column mode seems to be a quite complicated feature. The information is overloaded”.

P3 said it was really intuitive to navigate the map. He frequently alternated between the 3x3 method and the arrow key relative navigation. The previous/next navigation method was occasionally used. P3 often used the 3x3 method to listen to what counties were in each map range. He thought it was particularly fun to combine it with selection to find out where a selected county belonged to. To find a county on the map, P3 often used his geographical knowledge to pick a range that was likely to contain that

target county, listened to all counties in that range to confirm, and then used arrows keys to move to that county. P3 understood the spatial layout meaning of the 3x3 method and automatically connected the concepts such as west, middle, east, north, and south to the numeric keys.

P3 reported that the zooming was very easy to use but he did not see how to expand its value.

Select:

P3 had no problem using selection or combining it with other actions such as sorting, 3x3 map exploration, or circulating focus among sets of regions with different selection status.

Situate:

P3 had a habit of resetting the interface before each new task. He never felt lost. “I feel in control. I can always find out my situation”. During the training, P3 carefully tried the “situate” feature under different interface status to make sure that he understood every piece of information in the message. For example, he played with the map zooming and used the ‘I’ key to listen to region position changes at each zooming scale. During the test, P3 frequently used the “situate” feature to request status such as “how many regions were actually selected”, “what variables were sorted”, and so on. The “situate” feature was also the first thing he resorted to for trouble shooting.

Menu:

During the training, P3 thoroughly explored the whole menu system to make sure that no commands were left out. The menu was never used in the test since P3 remembered all the hotkeys.

Potential benefits

P3 was interested in using iSonic to look at information such as “disability information of states”. As a director at the National Federation of the Blind, he had a sighted personal assistant who would read and interpret such information to him, but some information was hard for others to explain in plain words. “If it’s just for the table, Excel is a quite good software. Being able to switch between table and map is good. The biggest advantage of iSonic is to switch back and forth between table and map”.

P3 also enjoyed using the map view. “I like maps. Maps are not possible in the usual software I use”. “Information is often too crowded on maps. This tool (iSonic) can be extended and useful to select information categories in table then see them on the map. Blind person can look at different theme maps such as water map, elevation map, city map, highway map, bus schedule map, etc.”

Appendix A - 4: Description of Case Four (P4)

Profile

P4 was a female in her middle 50s with a BA in English and French and also an MA in ESOL (English as a second language). P4 worked as a teleservice representative at the Social Security Administration. She was also a member of NACED (National Advisory Counsel for Employees w/ Disabilities). P4 was totally blind since birth and had no other disabilities. While fluent in Braille, P4 used JAWS for Windows and Windows-Eye daily for computer based activities such as browsing the Web, reading and editing Microsoft Word documents, processing emails with Microsoft Outlook, and inquiring information in Microsoft Access Databases. P4 was comfortable with tables and had limited experience with Excel spreadsheets. At the beginning of the test session, p4 received a quick review of spreadsheet navigation and sorting, and practiced until she reported to be comfortable. Her experience with tactile materials other than Braille was mainly tactile pictures in children's books, a talking globe, and occasionally some simple raised line maps. Without any formal training of tactile map reading, P4 said she was not comfortable with them. P4 used a standard keyboard and frequently used the number pad for screen reader commands. P4 had very limited knowledge about the U.S. state map and Maryland county map.

Task completion and error causes

P4 needed a hint to switch between the iSonic main window and the help window, but was able to learn iSonic by following the self-paced auditory lessons, occasionally asked

for clarifications. While P4 was able to clearly describe some correct and detailed strategies for all tasks except for the correlation task, she sometimes failed to follow through due to some repeated mistakes, and became frustrated because she thought her solutions would not work. The actual test task success rate was only 4 out of 7. “I have several conflicting ideas and cannot decide which one is the best. I tried one and it didn’t work out, so I switch to another one which didn’t work out either”.

There were two error-causing mistakes that P4 made repeatedly. The first was sorting the wrong column. Two factors possibly have contributed to the error. One was that some variable names were very similar. The other was that there was no feedback about which variable was actually sorted as the result of the sorting command. (Such feedback was added to iSonic for studies with later subjects.) The second mistake was skipping the 1st table row. Unlike in Excel, the table column header in iSonic was not a row that users could navigate to. Pressing the down arrow key after the “already top edge” hint would take the user to the 2nd data row. While all the other subjects had the habit of going back up to get the 1st data row, P4 kept skipping it even after the mistake was pointed out.

During most tasks, P4 kept asking for reassurance from the experimenter of what she was planning to do next. This showed that she was very afraid of making mistakes. “I think I know but I’m not sure. I don’t trust my memory” “Everyday at work, I have too much information to deal with. My supervisor is critical so I always double check to make sure I get it right”.

Task strategies and iSonic usage patterns

Map vs. table:

P4 had no problem understanding the concepts of map and table, and was very comfortable using both and switching between them according to task needs. With limited knowledge about Maryland counties, her first response to each geo-related task was always “I need a map”. The map was used whenever she needed the location information of a region. It was also used a lot to find a region given the name.

Brush:

P4 said that she enjoyed the feature of switching between the map and the table. By doing so, “they help each other”.

Pitch:

P4 mostly used the information level 2 (region name + pitch) during tasks and used pitch extensively to compare values. Pitch was considered to be extremely helpful when there were many numbers to be compare, such as in the case of geographical data distribution pattern or correlation tasks. “When a lot of numbers need to be compared, I won’t remember the numbers. Pitch forms clear patterns, and I don’t need to remember numbers”. Pitch was also used for comparing a few values such as in the pair-wise comparison task. “It’s faster. It gives me a clear cut and makes things shorter”. The detail-on-demand command was never used because she adjusted the information level instead.

Panning:

P4 reported to have noticed the stereo panning of sounds. However, it was not something that she had thought about or used for tasks.

Gist:

The whole map sweep and the 3x3 map sweep were often used by P4 to narrow down her search for a region or to get an idea of what each map range contained. Such sweep was typically done at the information level 2. When encouraged, P4 also swept the map with pitch only and accomplished the geographical data distribution pattern tasks. “It’s much faster. I don’t need the names when I only want the general pattern”.

Navigate:

P4 had great difficulty mastering the row/column table navigation mode and was instructed to skip it due to study time limitation. She was very comfortable with the cell navigation mode.

For the map, P4 alternated between the 3x3 method and the arrow key navigation methods. She said the 3x3 method was really helpful, especially for unfamiliar maps. “If I’m familiar with the map, arrow keys won’t be problematic. I’m familiar with U.S. states, so I used the arrows a lot in the beginning and forgot the 9 keys. But I’m not familiar with Maryland, so arrow keys took me all over the map and it’s not efficient. Later I found the 9 keys are easier to use so I used them a lot. The 3x3 method tells information about each range”. P4 often used the map to locate a region given the region name. To do so, she first used the 3x3 method at the information level 2 to find out which range contained the region, and then used arrow keys to go to the target region.

P4 understood the spatial layout meaning of the 3x3 method and automatically connected the keys on the number pad to geographical concepts such as west, middle, east, northeast, and so on. However, she sometimes got a little confused with the assignment of numbers because “the order on the PC number pad conflicts with the

phone”. To avoid such conflict, she usually only used the number pad as arrow keys and used the number keys on the top of the keyboard to enter numbers.

P4 tried zooming during the training and did not seem to have any problem using it and keeping oriented. Zooming was not used in the test because “there’s no need”. The previous/next navigation was never used in the test.

Select:

P4 was able to select/unselect regions and to combine it with other actions such as sorting, 3x3 map exploration, or circulating focus among sets of regions with different selection status. “Selecting let me isolate the things I want to focus on”.

Situate:

P4 had a habit of resetting the interface before starting a new task. She tried the “situate” feature in the training but did not use it in the test. When she kept finding that the table was not sorted as she was expecting (which happened often), she became frustrated and did not think of using the “situate” feature to check the status. Instead, she just resetting the interface and restarted with the task. P4 reported to feel lost sometimes.

Menu:

P4 did not spent much time exploring the menu in the training, but spent 15 minutes going through the menu at the beginning of the test session to refresh her memory of the available commands. During the test, she also used the menu system twice to recall two hotkeys. The menu system was reported to be easy to use.

Potential benefits

P4 was very enthusiastic about iSonic. “All the other applications are boring. iSonic has its personality. The map is really helpful, and I really enjoyed it. Blind users do not deal

with maps much, at least not in a successful way. The tones are entertaining, fun, and I enjoyed it. The tones won't replace data but it's a good reinforcement, it's a backup and complements the data. The map can have more sounds and can be used for teaching and learning for disable kids". P4 said she would like to use iSonic occasionally at work. "We get data once a year from the office of civil rights and employment support. I need to look at them. The map can be very helpful".

Appendix A - 5: Description of Case Five (P5)

Profile

P5 was a 40-year old female with a bachelor's degree in Sociology. She worked as an Information and Referral Specialist for a Maryland county government. She had progressive myopia since birth and completely lost her vision at age 15, but with no other disabilities. While fluent at Braille, P5 used computers extensively every day at work, with JAWS for Windows as the screen reader. P5 was experienced in Microsoft Office Suite including Word, Excel, Outlook, Internet Explorer, and Access. Keyboard was the only computer input device she used, and the number pad was frequently used. P5 was impressively good at spatial orientation even in new buildings and was very capable of traveling by herself via metros and buses. She owned a Maryland county tactile map and she had excellent knowledge about the U.S. state and the Maryland county map. P5 was very comfortable with tables. She also received a quick review of spreadsheet navigation and sorting and spent sometime practicing. As a design partner, P5 was very familiar with the value to sound mapping scheme. She also extensively used the map view, and defined the behavioral design of the menu system. However, she had not operated on the table view, and the online help system before the study.

Task completion and error causes

Instructed to go through the training as everyone else, P5 launched the self-paced auditory lessons and quickly went through all iSonic features, reviewing old features and learning new ones. Except for the correlation task that took P5 quite a while during the

training to master, overall it was reported to be very easy to combine the interface features to come up with efficient task strategies (10 on a 10-point scale with 10 being very easy). P5 made a careless mistake in the test by giving the highest five while the question asked for the lowest five, resulting in a test success rate of 6 out of 7.

P5 reported the key mapping to be very intuitive and easy to remember (10 on a 10-point scale with 10 being very easy). She attributed her thinking to deciding on the best strategy to use.

Task strategies and iSonic usage patterns

Map vs. table:

P5 was very comfortable with the concepts of map and table and switching between them during a task. Although P5 was very familiar with the locations of Maryland counties, she still thought the map view was helpful to tasks involving geographical locations. “I use the map as an extra information source to confirm, to get more accurate information”. When asked to perform geo-related tasks in Excel, she said she would have taken out her Maryland tactile map right away if she was carrying it.

Brush:

P5 usually relied on her memory instead of brushing to examine the same set of regions in both views. For example, for T6, she first used the map to find the adjacent counties, remembered all their names, then switched to the table, sorted the table by the desired variable, started to check from the first row and reported the first region name she encountered that was in the adjacent set she remembered. P5 gave several reasons for not having used selecting and brushing although she knew she could. “I just did a similar task in Excel. I’m already in the Excel mood, so I just used the same strategy in iSonic instead

of a different strategy. But I used the map as extra information source”. “There’re only 5 counties to remember and I’m familiar with all the names”. “It’ll be nice to select many things at once instead of one by one”.

However, in the post-test free exploration of an unknown map, P5 used the “brush” feature extensively. For example, she selected a few regions with interesting names in the table, and then switched to the map view, using the 3x3 method to locate them. She also sorted the table to find which selected region had the highest value, then used the 3x3 method again to locate it on the map. P5 was also excited that she could manipulate the table data then display her findings on the map to sighted people using iSonic.

Pitch:

P5 reported that using pitch to present values was very helpful (10 on a 10-point scale with 10 being very helpful). “Tones are easier to compare than numbers. If I am going to write a report, I’ll need the numbers. I’ll use tones to skim through the information when investigating, and then get the numbers when writing the report”. “iSonic and Excel are equally easy in finding specific information. But tones can tell trends and make iSonic easier to find data trends”. “Tones are helpful to confirm sorting”.

P5 frequently adjusted the information level and mostly used level 1 (pitch only) and 2 (region name + pitch). She used information level 1 for comparing sequences of values, such as in the correlation tasks and geographical data distribution trend tasks. The information level 3 (with values in speech) was only used occasionally for the pairwise comparison task. The details-on-demand command and the information level 0 (region name only) were never used.

Panning:

P5 reported to have noticed the stereo panning of sound in both the map and the table views, but did not make use of it intentionally.

Gist:

P5 understood how map spatial sweep was done but rarely used the whole map sweep except to automatically play all selected regions on the map. However, P5 extensively used partial map sweep at information level 1 and 2 by using the 3x3 exploration method.

P5 spent significant amount of time listening to gists of table rows/columns at information level 1 and 2 by using the row/column mode. Although P5 did not use gists much in the test session, she said it was helpful when she wanted to skim through a lot of information using tones.

Navigate:

P5 understood both table navigation modes and was very comfortable switching between them. P5 used the row/column mode a lot in the training at information level 1 and 2, but did not use it much in the testing. She preferred the cell mode in most cases, because “it gives more refined control of the interface”.

For the map, P5 mostly used the 3x3 method. “I didn’t use arrow keys much, although I had to use it for adjacency. Otherwise, I also want to find what are in west/east/north/south to get the bigger picture. The number pad works for that purpose”. Using the recursive 3x3 method, P5 was able to quickly recognize the overall map shape. For example, she correctly reported that the map in the post-test free exploration was ‘h’ shaped, and the northeast part of the U.S. state map was “going like a triangle”.

To find a region on the map, P5 usually first used the 3x3 method at information level 2 to listen to what each range contained then used the arrow keys to get to that region. P5 never used the previous/next navigation because “there’s no need to use it”.

P5 reported both the 3x3 method and the zooming to be very easy to use (10 on a 10-point scale with 10 being very easy). Although P5 did not use zooming in the test, she said it could be very helpful for focusing on an area on a big map. “24 counties in Maryland isn’t a lot. If it’s California or Virginia that has a lot more counties, I could use zooming”. One zooming-related problem was encountered when P5 tried to find the adjacent states of Kansas but was not able to reach two of them after zooming because they were in a different range partition from Kansas.

Select:

P5 had no problem selecting/unselecting regions and circulating focus among sets of region with different selection status. Selection was considered to be very helpful. “If selection is not available, I would’ve used my (Braille writing) sled to take notes or type in Word”. She also wanted to be able to select a range of table rows, and to select variables to put them next to each other for comparison.

Situate:

P5 had the habit of resetting the interface before beginning a new task, although she forgot to do so once during the training. P5 did not use the “situate” feature in the test because she never felt lost. “I know what I’m doing”. However, P5 said the “situate” feature was very necessary. “If I’m half way through a task, I’ll need it after a bathroom visit”.

Menu:

P5 spent quite some time exploring the menu. It was not used in the test because she remembered the commands already. “But if the product is released a year later, I’ll already forget the commands. The menu will be a way to learn the system”.

Potential benefits

P5 was very enthusiastic about using iSonic to examine and show her data. “For professional work, it can help me to move upwards in my career to be a planning specialist. It can help me to understand census data better. For example, I can look at the census data to find what and where more services are needed and where we don’t need more. I can look at all the states, then all the counties in Maryland, and clusters in the Montgomery county”. P5 also said iSonic could help her communicate with sighted colleagues. “I just got a laptop. I can put information on the laptop to take to a meeting, show the data on the iSonic map to other people. Many people don’t want to read a lot of data. I know what the data and the tones mean and other people can see it on the map with colors”.

P5 also suggested that tones could be used to help blind persons in other application areas. “I once got a call from a blind person who wants to be a radiologist. But nobody wanted to sign him up for the radiology class because he cannot see. We can detect the color difference in the x-ray image, detect anomalies and present them as special tones”.

Appendix A - 6: Description of Case Six (P6)

Profile

P6 was a female in her middle 40s with a master degree in Social Work. She worked remotely for the research and rehabilitation and training center at a private university, provided training on public disability programs and the impact on work to people with disabilities receiving benefits. P6 was born with congenital glaucoma and gradually lost her vision until she became legally blind around age 15. She was totally blind at the study time. P6 also had repetitive motion damage to both arms from computer use, and used a special split keyboard. P6 had very limited Braille skill, and relied on screen readers to access electronic information everyday, both at work and at home. JAWS for Windows was the most frequently used screen reader but she had also used other screen readers for DOS-based systems. P6 was experienced in Microsoft Office Suite including Word, Excel, Outlook, PowerPoint, and Internet Explorer. She also used a few other specialized accessibility software for portable gadgets. P6 used the number pad for screen reader navigation and entered numbers from the top row. P6 used to love maps when she could read them. The level of exactness and details in tactile or Braille maps was not possible in the usual size of Braille maps. She had excellent knowledge about U.S. states but only knew a few Maryland counties. P6 reported to have extensive experience with tables and spreadsheets.

Task completion and error causes

P6 had no problem learning iSonic features by launching and following the self-paced

auditory lessons. After the training, P6 was able to achieve a 100% task success rate in the test. Although P6 was able to quickly combine the interface features and came up with efficient task strategies, she reported the tasks to be on the difficult side (5 on a 10-point scale with 10 being very easy). Several reasons potentially attributed to such a subjective perception. First, unlike other subjects, P6 strongly disliked the quality of the speech synthesized by the Microsoft speech synthesizer. “The speech quality is awkward. It is too mechanized. The words are clipped. It takes a lot of processing effort. It makes me very tired”. During the study, the experimenter frequently clarified the meaning of individual words, and towards the end of the training, had to read the text in the training lessons because P6 became too frustrated to listen to long paragraphs of mechanized speech. Second, the musical tones interfered and worsened the speech quality, making it harder to focus on the speech when she wanted to. Third, some key assignments in iSonic were different from JAWS. P6 reported to spend half of her effort thinking of what keys to press. For example, P6 kept trying to use the number pad as arrow keys, as she had always been doing with JAWS. She also kept using the Control key instead of the ‘s’ key to stop the sound and was frustrated when it failed. P6 asked for the ability to let users reassign keys according to JAWS behaviors. “Key mapping has become a physical habit. It’s less cognitive than physical pattern. I have muscle memory of keys and expect certain functions”.

P6 reported the correlation task to be hard and the geographical value distribution task to be even harder, because she had to remember many spatial locations and values. However, P6 was able to accomplish both tasks fairly quickly. For correlation, P6 said she would normally use a scatter plot or a 3-line line graph to plot the data, and ask her

sighted human reader to tell her which correlates the best. For the geographical value distribution pattern task, she suggested it might be easier for her to group the values then judge the location of each value group. (Note: it can be supported by providing a sequence of separate spatial sweeps based on value clusters.)

Task strategies and iSonic usage patterns

Map vs. table:

It was very easy for P6 to understand the concepts of map and table and switch between them according to task needs. With very limited knowledge of Maryland counties, P6 extensively relied on the map for all geo-related tasks.

Brush:

P6 was very comfortable selecting and brushing among the map and table views. While P6 said that the geo-related tasks would not be hard given a tactile map, she thought such tasks would still be easier in iSonic because she could select the counties and easily switch between the map and the table. “I really like the integrated table and map. The biggest advantage is to be able to switch between the map and table. It’s valuable especially for big maps”. In the post-test free exploration, P6 used brushing to find a region in the table and locate it in the map.

Pitch:

P6 enjoyed the different musical sounds in the interface and kept sweeping the map at information level 1 (pitch only). “The tones are like in video games, very entertaining”. “It is music. I kept playing it like a piano”. However, at the default volume setup, P6 reported the pitches and the numbers in speech to have significantly interfered with each other. “There are multiple audio stimuli. If the TV is on and there is another

audio stream at about the same volume level, I can process neither”. “I can either listen to the tones or listen to the speech, but if the tones and speech are together, I cannot listen to any of them”. When offered to completely remove the pitch, she declined because it did have its value. “I use tones to get trend and the numbers to confirm”. “The tones as a soft background music is actually quite amusing”. Instead, she requested the musical sound volume to be tuned down significantly. “When the tone volume is lower, it’s a lot better”.

P6 frequently adjusted the information level and mostly used level 2 (region name + pitch). She also requested the ability to turn on and off the musical sound easily, to have an information level with speech only, just like she could have a level with tones only. The details-on-demand command and information level 0 (region name only) were never used.

Panning:

Although it was described in the training material, P6 reported not to have noticed or paid attention to the stereo panning of the sounds.

Gist:

P6 tried the whole map sweep at information level 1 and 2 during the training, and said that they were hard to follow, although she knew how the U.S. map looked like. “I like the ‘ding’ sound at the end, but the ‘click’ sound doesn’t help in the sweep”. P6 said that automatic sweep was helpful when the tones were helpful (for the trend tasks), but not helpful when the tones were not helpful (for non-trend tasks).

Navigate:

P6 spent quite some time in the training trying the table row/column mode. She also tried to use the mode for the correlation task in the test, but ended up switching to the cell

mode. “(The row/column mode) didn’t do much for me, because I won’t be able to tell where I am”.

For the map, P6 kept alternating between the 3x3 method and arrow key relative navigation method. Although she was able to report her location on the map with good approximation using the arrow key navigation, she said it was hard to know the location. She also reported the reversibility problem in the arrow key navigation and commented that there was not any easy solution to it. The 3x3 method was considered to be very easy to use and helpful in telling geographical locations. In the post-test free exploration, P6 used it with brushing to find out the geographical location of a region she had heard of the name but did not know the location. She was also able to recognize the overall map shape and compare relative region densities from the 3x3 method. P6 was able to use the zooming but thought it to be complicated because “I need to keep track of the zoom level”. P6 did not use the previous/next navigation at all and said that it did not make sense on maps.

Select:

It was very helpful to P6 to be able to easily select/unselect regions and focus on selected regions only by a simple keystroke. When P6 tried to compare two regions during the test (T4) in Excel, she first opened Microsoft Word and tried to pull the data of the two regions into a Word table. She had trouble operating the table, and received a suggestion from the experimenter to try a new spreadsheet. Although P6 was fluent in operating spreadsheets, it took her 20 minutes to finally pull out the data needed. P6 said that its still worthy because she might need the numbers later on. This suggested the need for iSonic to support easy export of selected data.

Situate:

P6 typically reset the interface before a new task. She occasionally used the “situate” feature in the test, e.g., to find how many regions were selected. She also wanted to skip information she did not need in the “situate” message.

Menu:

P6 spent quite some time exploring the menu but did not use it in the test. “It’s helpful as part of the learning. I like the organization.”

Potential benefits

Every year, P6 gets data from CMS (Center for Medicare and Medicaid Services) and needs to investigate the data for both geographies and numbers. She was interested in using iSonic for tasks that require map context which she could not complete with Excel. She also thought iSonic could be used as a teaching tool.

To make iSonic actual beneficial, P6 suggested that it should be made into a Web application accessible on public Websites. It shall also be compatible with JAWS and automatically take the JAWS setting, such as the speech and key assignments.

Appendix A - 7: Description of Case Seven (P7)

Profile

P7 was a male in his early 40s. Before suddenly going completely blind 13 years ago, he had no vision problem. With a bachelor degree in Commerce, he worked as an internal auditor and business analyst for 12 years. After becoming blind, he kept working as an internal auditor for another 5 years, then obtained a master degree in Rehabilitation Science and Technology, and was working as a Web and software accessibility specialist for a company that did Web accessibility evaluation and repaired tools and provided related services. With no Braille skill, P6 completely relied on screen readers (mainly JAWS for Windows, also IBM Home Page Reader and Windows Eye) for most social and work related communication needs. He spent at least 50 hours a week using computers. P7 was experienced in Microsoft Office Suite including Word, Outlook, Internet Explorer, and Excel. He also evaluated the accessibility of a variety of Web content and software applications. P7 used the number pad frequently. P7 was not able to use maps after becoming blind, but he often tried to trace fingers over normal map prints with sighted guidance, and to create a mental map of roads that he traveled often. He knew some U.S. states but had almost no knowledge about Maryland counties. As a formal accountant, P7 was an expert in tables and spreadsheets.

Task completion and error causes

P7 encountered no problem learning iSonic by launching and following the auditory lessons. It was easy for him to combine the interface features to come up with efficient

task strategies (9 on a 10-point scale with 10 being very easy). Unlike in Excel, the table column header in iSonic was not a row that users could navigate to. As a result, P7 made a mistake and skipped the 1st data row in T1, resulting a test success rate of 6 out of 7. Unlike all the other subjects, the correlation task was very easy for him. However, he reported the geographical data distribution task to be very difficult, although he was able to accomplish it successfully.

P7 reported the key mapping to be very intuitive and easy to remember (10 on a 10-point scale with 10 being very easy). He attributed 70% of his thinking to deciding on the best strategy to use.

Task strategies and iSonic usage patterns

Map vs. table:

P7 was very comfortable with both the map and the table and switching between them during a task. He had almost no knowledge of Maryland county locations, so the map was used for all geo-related tasks. “I use table for what table is best for, like sorting. I use map for what map is best for, like geography related information”. “Sorting isn’t the best use of the map”. “We could do something similar (dividing the map into 3x3 ranges) for Excel but it’s easier to use the map than presenting the geographical information as columns in the table. The table will be too crowded with the information. It’s also not easy to convert geographical information to text”.

Brush:

P7 really liked the ability to select and brush among the map and the table. “The number one distinguished feature of iSonic is the combined map and table. I can use map for what map can do best and use table for what table can do best, and toggle between

them”. Sometimes instead of selecting and brushing to view a few items in both views, he just identified and remembered them in one view, and relocated them in the other view, because there were only 2 or 5 items to remember. “Using the tightly-coupled selection in map and table will work for a larger number of items. For a few items, using memory is ok”. The dependence on the memory has caused some errors during the tasks.

Pitch:

P7 reported the use of pitch to be amusing, enjoyable, and very helpful, especially for discovering patterns in a series of values. When playing a whole table column in pitch only, he said that it was like sitting in the opera. P7 used information level 1 (pitch only) exclusively for correlation tasks. “When there are only 2 numbers to compare, numbers give more precise information. It’s not easy to look at the patterns of more numbers. Music makes a lot of sense in the case. If it’s difficult to tell from the music, I can use advanced features to calculate the coefficients. But this is not what an ordinary and nonexpert user can do”. “Tones are helpful in some extent for me to do things I used to do with graphs”. The use of pitch was also considered to be helpful for non-trend tasks. “The music is a good 1st cut”. P7 had no problem processing speech and music presented in parallel. “When the pitch isn’t needed (such as in the task to find a specific value), it doesn’t bother me, rather it works like separator between states.”

P7 frequently adjusted the information level, using only information necessary for the current goal and turned off all other information. All levels were used, including the name only level 0 which was ignored by all other subjects. The details-on-demand command was never used because adjusting the information level was enough.

Panning:

P7 reported to not have used the stereo panning of sounds in tasks.

Gist:

P7 extensively used the 3x3 method for partial map sweep, typically at information level 1 (pitch only) and 2 (region name + pitch). He also used the row/column table mode to sweep table columns at information level 1. “Automatically sweep the whole map isn’t difficult to understand but it took some practice. I’m comfortable with it but it’s not easy to lay it out as north to south, west to east. It’s only helpful for some tasks such as to find what counties are in a partition or to find where a county is.”

Navigate:

P7 understood both table navigation modes and was very comfortable switching between them. He used the cell mode most of the time, but the row/column mode for the correlation task and pair-wise comparison task.

Navigating the map was considered to be very easy overall. P7 automatically connected the number pad layout to geographical location concepts, such as west, east, or middle. The 3x3 method was considered very easy to use and very helpful (10 on a 10-point scale with 10 being very helpful). To find a region on the map, he usually first used the 3x3 method at information level 0 (region name only) to narrow down the range then used arrow keys to move to that region. By using the recursive 3x3 method, P7 was able to report in a fine detail the geographical layout of counties in the northeast and northmiddle parts of Maryland. However, P7 was concerned of the mental workload. “When zooming is involved, it’s intensive, requiring memory and mental processing. If

the scale of the map is large, it becomes more difficult. It's useful but it depends on the scale of the map". The previous/next navigation method was considered useless and not used at all.

Select:

P7 had no problem selecting/unselecting regions and circulating focus among sets of regions with different selection status. When performing pair-wise comparison tasks in Excel, he used the "hide rows/hide columns" command to temporarily remove regions and table columns not needed for the task. P7 said selecting in iSonic was easier than hiding items in Excel. He wanted to be able to select variables in iSonic as well.

Situate:

P7 had the habit of resetting the interface before beginning a new task, and was never lost. He used the "situate" feature often to confirm the sorting status and sometimes requested the table headings. P7 took a restroom break during the test and used the "situate" feature to request the situation upon return.

Menu:

P7 spent quite some time exploring the menu and said that the structure made sense. It was not used in the test because he remembered the commands already.

Potential benefits

P7 was very interested in using iSonic to look at government statistics, and other data, such as company product sale information by states, like for Wal-Mart. "This is an excellent application. And you brought it up to a good level". As an accessibility specialist, P7 also thought it could be used to improve the accessibility of content for government agencies, such as the social security agency and U.S. administration for

whom he was working to make the content accessible. “According to section 508, when information is converted into colors (on maps), the information should be converted into other means as well. Music tones can be used here... The number one strength of this tool is the combination of the map and the table, also the ability to communicate geographical information using the map. It can be used to improve the accessibility of maps on the Web”. P7 also suggested that the tool could be used for education purpose, to show other types of maps, such as maps showing waters and lands with basic physical attributes.

Appendix B: Algorithms for Region Set Split

The algorithms below were used recursively to split a set of regions in iSonic so that a vertical spatial sweep order can be generated by connecting regions in each subset.

```
int sortedMassCenterXCoords [] = mass center X coordinates sorted
    ascendingly for all regions in the current subset to be split;
//each X coordinate can be a potential split position. Test each one to find out the split
position causing minimum interleave between the resulting two subsets.
float currMinIntrusion = 1.1; //the intrusion ratio will not exceed 1.0
Set currBestSplitLeftSet = null; //no good split yet
Set currBestSplitRightSet = null; //no good split yet
for (int index = 0; index < sortedMassCenterXCoords.length(); index++) {
    //try splitting all regions into two subsets
    int testSplitPos = sortedMassCenterXCoords [index];
    Set leftSet = all regions whose mass center X coordinate is less than
        testSplitPos;
    Set rightSet = all regions whose mass center Y coordinate does not exceed
        testSplitPos;

    //calculate the intruding ratios of the two subsets, measured by the percentile
//area in the leftSet that overlaps with the bounding box of rightSet
//(intrusionLeftInRight) and the percentile region area in the rightSet that overlaps
//with the bounding box of leftSet (intrusionRightInLeft)
    float intrusionLeftInRight = calcIntrusionRatio(leftSet, rightSet);
    float intrusionRightInLeft = calcIntrusionRatio(rightSet, leftSet);
    if (intrusionLeftInRight <= intrusionRatioPerRegion * #region in leftSet
        && intrusionRightInLeft <= intrusionRatioPerRegion * #region in rightSet
        && max (intrusionLeftInRight, intrusionRightInLeft) < currMinIntrusion) {
```

```
//the split at the current X position is the best so far
    currBestSplitLeftSet = leftSet;
    currBestSplitRightSet = rightSet;
}
}

float calcIntrusionRatio (Set intruder, Set intrudee) {
    Rectangle intrudeeBB = intrudee.boundingBox;
    int overlapSize = sum of all intruders' areas overlapped with intrudeeBB;
    int intruderSize = sum of region sizes for all intruders who overlap intrudeeBB;
    return overlapSize / intruderSize;
}
```

Appendix C: Algorithms for Finding Target Region in Map

Adjacency Navigation

The algorithms below find the target region when users press the up arrow key to navigate upwards from the given source region during map adjacency navigation.

```
Region findAdjRegionUpwards(Region source){
    //first check the adjacent region list to see if there is any region whose mass center falls
    in the 90° radar range pointing upwards from the source region. If there is any, the one
    with the smallest deviation angle is the target region.
    Region target = null;
    float currMinDevAngle = 46°; //cannot deviate more than 45° from upwards
    for (testRegion = each Region in the adjacent region list of source) {
        if (testRegion.massCenter falls into the 90° radar range upwards from source) {
            float testDevAngle = the angle between the upward direction and the line from
            source.massCenter to testRegion.massCenter;
            if (testDevAngle < currMinDevAngle) {
                target = testRegion;
                currMinDevAngle = testDevAngle;
            }
        }
    }
    if (target != null) return target; //target region found in the first step

    //if nothing is found in the first step, scan the 120° upward radar range centered at the
    upper edge of the source region, with the scanner starting from the upward direction and
    gradually widening and elongating, looking for the first region encountered.
    Position sourceUpEdgePos = the top most position in the source region that has the
    same X coordinate as the source region's mass center;
```

```

    Region firstRegionScanned = the first region that intersects the radar scanner centered
    at sourceUpEdgePos;

    //return firstRegionScanned as the target region if any of these two criteria is met:
    //(1) firstRegionScanned is in the source region's adjacent list;
    //(2) the angle between the upward direction and the line from the source region mass
    center to the firstRegionScanned mass center is no greater than 45°.
    if (firstRegionScanned is in the adjacent region list of the source Region) {
        return firstRegionScanned;
    }
    float devAngle = the angle between the upward direction and the line from the
    source.massCenter to firstRegionScanned.massCenter;
    if (devAngle <= 45°) {
        return firstRegionScanned;
    }

    return null; //no adjacent region is found after the two steps.
}

```

Bibliography

- [1] ABC Evening News, "Opticon for blind", June 15, 1972, available from Television News Archive at Vanderbilt University, <http://openweb.tvnews.vanderbilt.edu/1972-6/1972-06-15-ABC-19.html> (accessed on 2/23/2006)
- [2] Abel, S.M., Figueiredo, J.C., Consoli, A., Birt, C.M., and Papsin, B.C., "The effect of blindness on horizontal plane sound source identification," *International Journal of Audio*, 41(5), pp. 285-92, July 2002.
- [3] ACM Conference on Universal Usability, <http://sigchi.org/cuu/>, 2000.
- [4] Aleman A, van Lee L, Mantione MH, Verkoijen IG, and de Haan EH., "Visual imagery without visual experience: evidence from congenitally totally blind people," *Neuroreport*, 12(11), pp. 2601-2604, Lippincott Williams & Wilkins, August 2002.
- [5] Algazi, V. R., Duda, R. O., Thompson, D. P., and Avendano. C., "The CIPIC HRTF database," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 99-102, Washington, D.C.: IEEE Computer Society, 2001.
- [6] Alty, J.L. and Rigas, D., "Communicating graphical information to blind users using music: the role of context," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 574-581, New York: Association for Computing Machinery, 1998.

- [7] American National Standards Institute, “American national psychoacoustic terminology (No. S3.20)”, *American National Standards Institute*, New York, 1973.
- [8] American Standards Association, “Acoustical terminology (No. S1.1)”, *American Standards Association*, New York, 1960.
- [9] Anderson, J.R., “Arguments concerning representations for mental imagery”, *Psychological Review*, 85, pp. 249-277, District of Columbia: American Psychological Association, 1978.
- [10] Andrienko, G. and Andrienko, N., “Interactive maps for visual data exploration”, *International Journal of Geographical Information Science*, 13 (4), pp.355-374, Abingdon, UK: Taylor & Francis, June 1999.
- [11] Arno, P., De Volder, A.G., Vanlierde, A., Wanet-Defalque, M.C., Streel, E., Robert, A., Sanabria-Bohorquez, S., and Veraart, C., “Occipital activation by pattern recognition in the early blind using auditory substitution for vision,” *Neuroimage*, 13(4), pp. 632-45, Elsevier, April 2001.
- [12] Arons, B., “A review of the cocktail party effect”, *Journal of the American Voice I/O Society*, 12(7), pp. 35-50, 1992.
- [13] Arthur, E. J., Hancock, P. A., and Chrysler, S. T., “The perception of spatial layout in real and virtual worlds,” *Ergonomics*, 40(1), pp. 69-77, England: Taylor & Francis, 1997.
- [14] Ashmead, D.H., Wall, R.S., Ebinger, K.A., Eaton, S.B., Snook-Hill, M.M., and Yang, X., “Spatial hearing in children with visual disabilities,” *Perception*, 27(1), pp. 105-22, London: Pion, 1998.

- [15] Atkinson, R.C. and Shiffrin, R.M., "The control of short-term memory," *Scientific American*, pp. 82-90, August 1971.
- [16] Axen, U., Choi, I., "Using additive sound synthesis to analyze simplicial complexes," *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 31-43, Sante Fe, NM, USA, 1994.
- [17] Axen, U. & Choi, I., "Investigating geometric data with sound," *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 25-28, Palo Alto, CA, USA, International Community for Auditory Display, Nov 4-6, 1996.
- [18] Barrass, S., "TaDa! demonstrations of auditory information design," *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 1996.
- [19] Bederson, B.B., Grosjean, J., and Meyer, J., "Toolkit design for interactive structured graphics", *IEEE Transactions on Software Engineering*, 30 (8), pp. 535-546, Washington, D.C.: IEEE Computer Society, 2004.
- [20] Bennett, D.J., Edwards, and A.D.N., "Exploration of non-seen diagrams", *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 1998.
- [21] Ben-Tal O., Berger J., Cook B., Daniels M., Scavone G., and Cook P., "SONART: the sonification application research toolbox," *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2002.
- [22] Blattner, M., Sumikawa, D., and Greenberg, R., "Earcons and icons: their structure and common design principles", *Human Computer Interaction*, 4(1), pp.

11-44, 1989.

- [23] Blattner, M., Greenberg, R.M., and Kamegai, M., "Listening to turbulence: an example of scientific audiolization," *Proceedings of the ACM SIGCHI Workshop on Multimedia and Multimodal Interface Design*, pp. 1-8, New York: Association for Computing Machinery, 1990.
- [24] Blauert, J., *Spatial Hearing: the Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1997.
- [25] Bonebright, T.L., Nees, M. A., Connerley, T.T., and McCain, G. R., "Testing the effectiveness of sonified graphs for education: a programmatic research project," *Proceedings of the International Conference on Auditory Display (ICAD)*, Espoo, Finland, pp. 62-66, International Community for Auditory Display, 2001.
- [26] Bregman, A.S., *Auditory Scene Analysis*, Cambridge, MA: MIT Press, 1990.
- [27] Brewster, S.A., *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*, PhD Thesis, UK: University of York, 1994.
- [28] Brewster, S.A., "Using nonspeech sounds to provide navigation cues," *ACM Transactions on Computer-Human Interaction*, 5(3), pp. 224-259, 1998.
- [29] Brewster, S.A., Capriotti, A., and Hall, C., "Using compound earcons to represent hierararchies," *HCI Letters*, 1(1), pp. 6-8, London: Springer-Verlag, 1998.
- [30] Brewster, S.A., Wright, P.C., and Edwards, A.D.N., "The design and evaluation of an auditory-enhanced scroll," *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 173-179, New York: Association for Computing Machinery, 1994
- [31] Brown, L., Brewster, S.A., Ramloll, R., Yu, W. & Reidel, B., "Browsing Modes

- for Exploring Sonified Line Graphs,” *Proceedings of the British HCI Group Annual Conference (IHM-HCI)*, pp. 6-9, 2002.
- [32] Brown, L., Brewster, S.A., “Drawing by ear: interpreting sonified line graphs,” *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 152-156, International Community for Auditory Display, 2003.
- [33] Brown, L., Brewster, S.A., Ramloll, R., Burton, M., and Riedel, B., “Design guidelines for audio presentation of graphs and tables”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 284-287, Boston, International Community for Auditory Display, July 6-9, 2003.
- [34] Brown, C. and Duda, R., “A structured model for binaural sound synthesis”, *IEEE Transactions on Speech and Audio Processing*, 6, pp. 476-488, Washington, D.C.: IEEE Computer Society, September, 1998.
- [35] Brown, M.H., and Hershberger, J., “Color and sound in algorithm animation”, *IEEE Computer*, 25(12), pp. 52-63, Washington, D.C.: IEEE Computer Society, 1992.
- [36] Burbeck, S., “Applications programming in smalltalk-80(TM): how to use model-view-controller (MVC)”, available at <http://st-www.cs.uiuc.edu/users/smarch/st-docs/mvc.html> (accessed on 01/12/2006)
- [37] Bussemakers, M.P. and Haan, A. de., “When it sounds like a duck and it looks like a dog... auditory icons vs. earcons in multimedia environments”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 184-189, International Community for Auditory Display, 2000.
- [38] Buxton, W., “Introduction to this special issue on nonspeech audio,” *Human*

- Computer Interaction*, 4(1), pp. 1-9, 1989.
- [39] Card, S., Mackinlay, J.D., and Robertson, G., “The design space of input devices,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 117-124, New York: Association for Computing Machinery, 1990.
- [40] Card, S., Mackinlay, J.D., Shneiderman, B., *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, 1999.
- [41] Carlile, S., *Virtual Auditory Space: Generation and Applications*. New York: RG Landes, 1996.
- [42] Cherry, E.C., “Some experiments on the recognition of speech”, *Journal of Acoustic Society of America*, 25, pp. 975-979, 1953.
- [43] Choi, I. and Bargar, R., “Interfacing sound synthesis to movement for exploring high-dimensional system in a virtual environment,” *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3, pp. 2772-2777, Washington, D.C.: IEEE Computer Society, 1995.
- [44] Christou, C., and Bühlhoff, H. H., “View dependence in scene recognition after active learning,” *Memory and Cognition*, 27, pp. 996-1007, 1999.
- [45] Cohen, B.H., *Explaining Psychological Statistics*, 2nd Edition, John Wiley & Sons, 2001
- [46] Corda Technologies Inc., *Corda Products Optimap 5*, available at <http://www.corda.com/products/optimap/>, (accessed on 23/10/2003).
- [47] Cronly-Dillon, J., Persaud, K., and Gregory, R.P.F., “The perception of visual images encoded in musical form: a study in cross modality information transfer,”

- Proceedings of the Royal Society London*, B266, pp. 2427-2433, 1999.
- [48] Dang, G., North, C., and Shneiderman, B., “Dynamic queries and brushing on choropleth maps”, *Proceedings of the IEEE International Conference on Information Visualization*, pp. 757-764, Washington, D.C.: IEEE Computer Society, 2001.
- [49] Donker, H., Klante, P., and Gorny, P., “The design of auditory user interfaces for blind users,” *Proceedings of the Nordic Conference on Human-Computer Interaction*, pp. 149-156, 2002.
- [50] Duda, R., “Modeling head related transfer function”, *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, pp. 457-461, 1993.
- [51] eEurope, *eEurope Standards*, <http://www.e-europestandards.org/>, (accessed on 10/09/2005).
- [52] Environmental Systems Research Institute, Inc. (ESRI), *ESRI Shapefile Technical Description*, available at <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, July 1998, (accessed on 01/12/2006),
- [53] European Blind Union, *Statistics*, available at <http://www.euroblind.org/fichiersGB/STAT.htm>, (accessed on 10/19/2005).
- [54] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., *Advances in Knowledge Discovery and Data Mining*, MA: MIT Press, 1996.
- [55] Fernstrom, M. and Brazil, E., “Sonic browser: an auditory tool for multimedia asset management,” *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 132-135, International Community for Auditory Display,

2001.

- [56] Filepp, F., Challenger, J., and Rosu, D., "Improving the accessibility of aurally rendered HTML tables," *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 9-16, New York: Association for Computing Machinery, 2002
- [57] Fisher, P., "Hearing the reliability in classified remotely sensed images", *Cartography and Geographic Information Systems*, 21(1), pp. 31-36, 1994.
- [58] Fitch, T. and Kramer, G., "Sonifying the body electric: superiority of an Auditory over a visual display in a complex, multi-variate system," In G. Kramer (Ed.) *Auditory Display: Sonification, Audification and Auditory Interfaces*, Proceedings Volume 18, Santa Fe Institute, Studies in the Sciences of Complexity, Addison-Wesley Publishing Company, 1994.
- [59] Flowers, J.H. and Hauer, T. A., "Musical versus visual graphs: cross-modal equivalence in perception of time series data," *Human Factors*, 37(3), pp. 553-569, Santa Monica, CA: Human Factors and Ergonomics Society (HFES), 1995.
- [60] Flowers, J.H., Buhman, D.C., and Turnage, K.D., "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples," *Human Factors*, 39(3), pp. 340-350, Santa Monica, CA: Human Factors and Ergonomics Society (HFES), 1997.
- [61] Franklin, K.M. and Roberts, J.C., "Pie chart sonification", *Proceedings of the IEEE International Conference on Information Visualization*, pp. 4-9, Washington, D.C.: IEEE Computer Society, July 2003.
- [62] Franklin, K.M. and Roberts, J.C., "A path based model for sonification",

- Proceedings of the IEEE International Conference on Information Visualization*, pp. 865-870, Washington, D.C.: IEEE Computer Society, 2004.
- [63] Freedom Scientific, *JAWS® for Windows*, available at http://www.freedomscientific.com/fs_products/software_jaws.asp (accessed on 10/09/2005)
- [64] Gardner, W.G., "Transaural 3-D audio," *MIT Media Laboratory Perceptual Computing Section*, Technical Report No. 342, July 20, 1995.
- [65] Gaver, W., "Using sound in computer interfaces," *Human Computer Interaction*, 2(2), pp.167-177, 1986
- [66] Gaver, W., "The SonicFinder: an interface that uses auditory icons," *Journal of Human Computer Interaction*, 4, Hillsdale: Lawrence Erlbaum Associates, pp. 67-93, 1989.
- [67] Golledge, R. G., "Tactual strip maps as navigational aids", *Journal of Blindness and Vision Impairment*, 85(7), pp. 296-301, American Foundation for the Blind, 1991.
- [68] Golledge, R. G., Loomis, J. M., Klatzky, R. L., Flury, A., and Yang, X. L. "Designing a personal guidance system to aid navigation without sight: progress on the GIS component", *International Journal of Geographical Information Science*, 5(4), pp. 373-395, Abingdon, UK: Taylor & Francis, 1991.
- [69] Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R.J., and Belin, P. "Pitch discrimination in the early blind", *Nature*, 430, pp. 309-310, London: Nature Publishing Group, 15th July 2004.
- [70] Griffin, A., "Feeling it out: the use of haptic visualization for exploratory

- geographic analysis”, *Cartographic Perspectives*, 39, pp. 12-29, North American Cartographic Information Society, 2002.
- [71] Gröhn, M. and Takala, T., “MagicMikes – methods for spatial sonification,” *Proceedings of the IS&T/SPIE International Symposium on Electronic Imaging: Science & Technology*, pp. 294-301, 1995.
- [72] Gulick, W.L., Gescheider, G.A., and Frisina, R.D., *Hearing: Physiological acoustics, neural coding, and psychoacoustics*, Oxford University Press, 1989.
- [73] Gutin, G., Punnen, A.P., *Traveling Salesman Problem and Its Variations*, Kluwer Academic Publishers, 2002
- [74] Handel, S., *Listening: an Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press, 1989.
- [75] Hartmann, W.M., *Sounds, Signals, and Sensation: Modern Acoustics and Signal Processing*, New York: Springer Verlag, 1997.
- [76] Hart, S. G. and Wickens, C. D., “Workload assessment and prediction,” In H. R. Boher (Ed.), *Manprint: an Emerging Technology, Advanced Concepts for Integrating People, Machine, and Organizations*, New York: Van Nostrand Reinhold, pp. 257–296, 1990.
- [77] Heller, M.A., “Tactile memory in sighted and blind observers: the influence of orientation and rate of presentation,” *Perception*, 18, pp. 121-133, London: Pion, 1989.
- [78] Hermann, T., Krause, J., and Ritter, H., “Real-time control of sonification models with a haptic interface”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 82-86, Kyoto, Japan, International Community for Auditory

Display, July 2–5, 2002.

- [79] Hermann, T., Meinicke, P., and Ritter, H., “Principal Curve Sonification”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 81-86, International Community for Auditory Display, 2000.
- [80] Hermann, T. and Ritter, H., “Listen to your data: model-based sonification for data analysis,” in M. R. Syed, (Ed.) *Advances in Intelligent Computing and Multimedia Systems*, pp. 189-194, International Institute for Advanced Studies in System Research and Cybernetics, 1999.
- [81] Howard, I.P. and Templeton, W.B., *Human Spatial Orientation*, John Wiley & Sons, 1966.
- [82] Hunt, A., Hermann, T., Pauletto, S., “Interacting with sonification systems: closing the loop,” *Proceedings of the IEEE International Conference on Information Visualization*, pp. 879-884, Washington, D.C.: IEEE Computer Society, 2004
- [83] Jacko, J.A. and Sears, A., “Designing interfaces for an overlooked user group: considering the visual profiles of partially sighted users,” *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 75-77, New York: Association for Computing Machinery, 1998.
- [84] Jacob, R.J.K. and Sibert, L.E., “The perceptual structure of multidimensional input device selection,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 211-218, New York: Association for Computing Machinery, 1992.
- [85] Jacobson, R.D., “Navigating maps with little or no sight: An audio-tactile

- approach,” *Proceedings of the Workshop on Content Visualization and Intermedia Representations (CVIR)*, 1998.
- [86] James, J., “Lessons from developing audio HTML interfaces,” *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 27-34, New York: Association for Computing Machinery, 1998.
- [87] Jeong, W., *Adding Haptic and Auditory Display to Visual Geographic Information*, PhD Thesis, Florida State University, 2001.
- [88] Kamel, H.M., Roth, P., “Graphics and user’s exploration via simple sonics (GUESS): providing interrelational representation of objects in a non-visual environment,” *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 261-266, Espoo, Finland, International Community for Auditory Display, July29-August1, 2001.
- [89] Kennel, A., “AudioGraf: diagram reader for the blind,” *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 51-56, New York: Association for Computing Machinery, 1996.
- [90] Kramer, G., “Audification: using sound to understand complex systems and navigate large data sets,” *Proceedings of the Santa Fe Institute Science Board*, Santa Fe Institute. 1990.
- [91] Kramer, G., “Some Organizing Principles for Representing Data With Sound,” In G. Kramer (Ed.), *Auditory Display: Sonification, Audification and Auditory Interfaces*, Proceedings Volume 18, Santa Fe Institute, Studies in the Sciences of Complexity, Addison-Wesley Publishing Company, 1994.
- [92] Kramer, G., “Sonification of financial data: an overview of spreadsheet and

- database sonification,” *Proceedings of Virtual Reality Systems*, New York, NY: SIG Advanced Applications, 1993.
- [93] Kramer, G., and Ellison, S., “Audification: the use of sound to display multivariate data,” *Proceedings of the International Computer Music Conference*, pp. 214-221, International Computer Music Association, 1991.
- [94] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J., *Sonification Report: Status of the Field and Research Agenda*, International Community for Auditory Display, 1997.
- [95] Krygier, J., “Sound and geographic visualization”, In MacEachren, A. M., and Taylor, D. R. F. (Eds.), *Visualization in Modern Cartography*. Oxford, UK: Pergamon, pp. 146-166, 1994.
- [96] Landau, S. and Gourgey, K., “Development of a talking tactile tablet”, *Information Technology and Disabilities E-Journal*, VII (2), Lake Forest, CA: Equal Access to Software and Information, 2001.
- [97] Leplâtre, G. and Brewster, S.A. (2000), “Designing non-speech sounds to support navigation in mobile phone menus,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2000.
- [98] Lessard N., Pare M., Lepore F., and Lassonde M., “Early-blind human subjects localize sound sources better than sighted subjects,” *Nature*, 395(6699), pp. 278-80, London: Nature Publishing Group, September, 1998.
- [99] Lewald J., “Opposing effects of head position on sound localization in blind and sighted human subjects,” *European Journal of Neuroscience*, 15(7), pp.1219-

1224, UK: Blackwell Publishing, April 2002.

- [100] Lodha, S.K., Beahan, J., Heppe, T., Joseph, A., and Zane-Ulman, B., "MUSE: a musical data sonification toolkit," *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 36-40, International Community for Auditory Display, 1997.
- [101] Lodha, S.K., Whitmore, D., Hansen, M., and Charp, E., "Analysis and user evaluation of a musical-visual system: does music make any difference?" *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 167-172, International Community for Auditory Display, 2000.
- [102] Loge, K., Dram, A., and Inman, D., "Teaching orientation and mobility skills to blind children using computer generated 3-D sound environments", *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 1-5, Atlanta, GA, USA, International Community for Auditory Display, April 2000.
- [103] Lumbreras, M., and Sanchez, J., "Interactive 3D sound hyperstories for blind children", *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 318-325, New York: Association for Computing Machinery, 1999.
- [104] Mansur, D.L., Blattner, M.M., and Joy, K. I., "Sound graphs: a numerical data analysis method for the blind," *Journal of Medical System*, 9(3), pp.163-174, USA: Kluwer Academic/Plenum Publishers, 1985.
- [105] Maucher, T., Meier, K., and Schemmel, J., "An interactive tactile graphics display," *Proceedings of the IEEE International Symposium on Signal Processing and its Applications (ISSPA)*, pp. 190-193, Washington, D.C.: IEEE Computer

Society, 2001.

- [106] McAdams, S. and Bigand, E. (Eds.), *Thinking in sound: the cognitive psychology of human audition*. Oxford: Clarendon Press, 1993.
- [107] MacEachren, A.M., Wachowicz, M., Haug, D., Edsall, R., and Masters, R., “Constructing knowledge from multivariate spatiotemporal data: integrating geographic visualization with knowledge discovery in database methods”, *International Journal of Geographic Information Science*, 13(4), pp. 311-334, Abingdon, UK: Taylor & Francis, 1999.
- [108] MacEachren, A.M., Dai, X., Hardisty, F., Guo, D., and Lengerich, G., “Exploring high-D spaces with multiform matrices and small multiples”, *Proceedings of the IEEE Symposium on Information Visualization*, pp. 31-38, Washington, D.C.: IEEE Computer Society, 2003
- [109] Mackinlay, J.D., *Automatic Design of Graphical Presentations*, PhD thesis, Stanford University, Department of Computer Science, 1986. Technical Report STAN-NCS-86-1138.
- [110] McGookin, D.K. and Brewster, S.A., “Dolphin: the design and initial evaluation of multimodal focus and context,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2002.
- [111] Meijer, P.B.L., “An experimental system for auditory image representations,” *IEEE Transactions on Biomedical Engineering*, 39(2), pp. 112-121, Washington, D.C.: IEEE Computer Society, February 1992.
- [112] Middlebrooks, J. C., “Virtual localization improved by scaling non-individualized

- external-ear transfer functions in frequency”, *Journal of the Acoustic Society of America*, 106(3), pp. 1493-1510, 1999.
- [113] Moore, B.C.J., *An Introduction to the Psychology of Hearing*, Academic Press, 1982.
- [114] Moore, F. R., *Elements of Computer Music*, Prentice Hall, 1990.
- [115] Moore, B.C.J. (Ed.), *Handbook of Perception and Cognition (Vol. 6): Hearing*, New York: Academic Press, 1995.
- [116] Moore, B.C.J., *An Introduction to the Psychology of Hearing*, 4th edition, Orlando, FL: Academic Press, 1997.
- [117] Mountford, S.J. and Gaver, W., “Talking and listening to computers,” In B. Laurel (Ed.), *The Art of Human-Computer Interface Design*, pp. 319-334, Massachusetts: Addison-Wesley, 1990.
- [118] Müller–Tomfelde, C. and Steiner, S., “Audio-enhanced collaboration at an interactive electronic whiteboard,” *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 267-271, International Community for Auditory Display, 2001.
- [119] Mynatt, E. D. and Edwards, W. K., “Mapping GUIs to auditory interfaces,” *Proceedings of the ACM Symposium on User Interface Software and Technology, Audio and Asynchronous Services*, pp. 61-70, New York: Association for Computing Machinery, 1992.
- [120] Mynatt, E.D., Weber, G., “Nonvisual presentation of graphical user interfaces: contrasting two approaches,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 166-172, New York: Association for

Computing Machinery, 1994.

- [121] National Cancer Institute, *Cancer Mortality Maps and Graphs*, available at <http://www3.cancer.gov/atlasplus/>, (accessed on 23/10/2003).
- [122] National Federation of the Blind, *Blindness Statistics*, available at <http://www.nfb.org/stats.htm>, (accessed on 10/19/2005).
- [123] Neuhoff, J.G., Kramer, G., Wayand, J., “Sonification and the interaction of perceptual dimensions: can the data get lost in the map?” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2000.
- [124] Norman, K.L., Zhao, H., Shneiderman, B., and Golub, E., “Dynamic query choropleth maps for information seeking and decision making”, *Proceedings of the International Conference on Human - Computer Interaction (HCII)*, vol 2, *Theory and Practice*, pp. 1263-1267, Mahway, NJ: Lawrence Erlbaum Associates, 2003.
- [125] North, C. and Shneiderman, B., “Snap-together visualization: a user interface for coordinating visualizations via relational schemata”, *Proceedings of Advanced Visual Interfaces*, pp. 128-135, New York: Association for Computing Machinery, 2000.
- [126] Oogane, T. and Asakawa, C., “An interactive method for accessing tables in HTML,” *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 126-128, New York: Association for Computing Machinery, 1998.
- [127] Parente, P. and Bishop, G., “BATS: the blind audio tactile mapping system,”

Proceedings of the ACM Southeast Regional Conference, New York: Association for Computing Machinery, March 2003.

- [128] Parkes, D., “Audio tactile systems for designing and learning complex environments as a vision impaired person: static and dynamic spatial information access”, Steele, J., and Hedberg, J. G., (Eds.), *Learning Environment Technology: Selected Papers from LETA 94*. Canberra, AJET Publications, pp. 219-223, 1994.
- [129] Perez-Quinones, M.A., Capra, R.G., and Shao, Z., “The ears have it: a task by information structure taxonomy for voice access to web pages,” *Proceedings of Interact*, Amsterdam, The Netherlands: IOS Press, 2003.
- [130] Pauletto, S., Hunt, A., “A toolkit for interactive sonification,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2004.
- [131] Perlman, G., “Human-computer interaction on-line bibliography,” available at <http://www.hcibib.org>, accessed on 09/08/2004.
- [132] Pirhonen, A., Brewster, S.A., and Holguin, C., “Gestural and audio metaphors as a means of control for mobile devices,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 291-298, New York: Association for Computing Machinery, 2002.
- [133] Pitt, I.J., and Edwards, A.D.N., “Pointing in an auditory interface for blind users,” *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 280-285, Washington, D.C.: IEEE Computer Society, 1995.
- [134] Poll, L.H. and Eggen, B.H., “Non-visual interaction with GUI objects,” *Proceedings Of the HCI'96 Conference on People and Computers*, XI, pp. 159-

- 168, Surrey, UK: Springer-Verlag, 1996.
- [135] Radeau, M., "Cognitive impenetrability in auditory-visual interaction," *Analytic Approaches to Human Cognition*, pp. 41-55. Amsterdam, The Netherlands: Elsevier Science Publishers B.V., 1992.
- [136] Ramloll, R., *Supporting Co-operative Work Through Ubiquitous Awareness-Filtration Mechanisms*, PhD Thesis, Lancaster University UK, 2000.
- [137] Ramloll, R., and Brewster, S.A., "An environment for studying the impact of spatialising sonified graphs on data comprehension," *Proceedings of the IEEE International Conference on Information Visualization*, pp. 167-174, July 10-12, London, England, Washington, D.C.: IEEE Computer Society, 2002.
- [138] Ramloll, R., Yu, W., Brewster, S.A., Riedel, B., Burton, M., Dimigen, G., "Constructing sonified haptic line graphs for the blind student: first steps," *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 17-25, New York: Association for Computing Machinery, Nov 13-15, 2000.
- [139] Ramloll, R., Yu, W., Riedel, B. and Brewster, S.A., "Using non-speech sounds to improve access to 2-D tabular numerical information for visually impaired users," *Proceedings of BCS IHM-HCI*, Lille, France, pp. 515-530, France: Cepadues Editions, 2001.
- [140] Ramstein, C., Martial, O., Duftesne, A., Carrum, M., Chassb, P., and Mabileuff, P., "Touching and hearing GUI's: design issues for the PC-access system," *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 2-9, New York: Association for Computing

Machinery, 1996.

- [141] Reenskaug, T., “The Model-View-Controller (MVC): its past and present”, available at http://heim.ifi.uio.no/~trygver/2003/javazone-jaoo/MVC_pattern.pdf, (accessed on 01/12/2006).
- [142] Peterson, L.R. and Peterson, M.J., “Short-term retention of individual verbal items,” *Journal of Experimental Psychology*, 58, pp. 193-198, Washington, D.C.: American Psychological Association, 1959.
- [143] Roberts, L.A., and Sikora, C.A., “Optimising feedback signals for multimedia devices: Earcons vs. Auditory icons vs. Speech,” *Proceedings of the International Ergonomics Association Congress*, pp. 224-226, Tampere, Finland, 224-226, 1997.
- [144] Röber, N. and Masuch, M., “Interacting with sound: an interaction paradigm for virtual auditory worlds,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2004.
- [145] Roder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., Hillyard, S.A., and Neville, H.J., “Improved auditory spatial tuning in blind humans,” *Nature*, 400(6740), pp. 162-166, London: Nature Publishing Group, July 1999.
- [146] Rock, I., and DiVita, J., “A case of viewer-centered object perception”, *Cognitive Psychology*, 19, pp. 280-293, Elsevier, 1987.
- [147] Roffler, S. and Butler, R., “Localization of tonal stimuli in the vertical plane”, *Journal of the Acoustical Society of America* 43, pp. 1260-1266, 1968.
- [148] Ronkainen, S., “Earcons in motion – defining language for an intelligent mobile device,” *Proceedings of the International Conference on Auditory Display*

- (*ICAD*), International Community for Auditory Display, 2001.
- [149] Roth, S. F., Chuah, M. C., Kerpedjiev, S., Kolojejchick, J. A., and Lucas, P., “Towards an information visualization workspace: combining multiple means of expression”, *Human-Computer Interaction*, 12(1-2), pp. 131-185, Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1997.
- [150] Roth, P., Petrucci, L., Assimacopoulos, A., and Pun T., “AB-Web: active audio browser for visually impaired and blind users,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 1998.
- [151] Roth, P., Petrucci, L., and Pun, T., “From dots to shapes: an auditory haptic game platform for teaching geometry to blind pupils,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2000.
- [152] Saue, S., “A model for interaction in exploratory sonification displays,” *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2000.
- [153] Scaletti, C., “Sound synthesis algorithms for auditory data representations,” In G. Kramer, (Ed.), *Auditory Display*, Addison-Wesley, 1994.
- [154] Schneider, J., Strothotte, T., “Constructive exploration of spatial information by blind users,” *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pp. 188-192, New York: Association for Computing Machinery, 2000.
- [155] Section 508 of the US Rehabilitation Act, <http://www.section508.gov/>, (accessed

on 10/18/2005)

- [156] The Sendero Group, *Sendero Group Homepage – GPS for the blind*, <http://www.senderogroup.com>, 2003.
- [157] Shinn-Cunningham, B.G., “Distance cues for virtual auditory space”, *Proceedings of the IEEE International Symposium on Multimedia Information Processing*, pp. 227-230, Washington, D.C.: IEEE Computer Society, 2000.
- [158] Shinn-Cunningham, B.G., Zurek, P.M., Stutman, E.R., and Berkovitz, R., “Perception of azimuth for sources simulated using two loudspeakers in natural listening environments,” *Proceedings of the 19th Association for Research in Otolaryngology (ARO) Midwinter Meeting*, St. Petersburg Beach, FL, 1996.
- [159] Shneiderman, B., “The eyes have it: a task by data type taxonomy for information visualizations,” *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336, Washington, D.C.: IEEE Computer Society, 1996.
- [160] Shneiderman, B., and Plaisant, C., *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 4th Edition, Addison Wesley Longman Inc., 2005.
- [161] Sikora, C.A., Roberts, L.A., and Murray, L., “Musical vs. real world feedback signals,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 220-221, New York: Association for Computing Machinery, 1995.
- [162] Simons, D. J. and Wang, R. F., “Perceiving real-world viewpoint changes”, *Psychological Science*, 9(4), pp. 315-320, UK: Blackwell Publishing, 1998.
- [163] Singer, M. J., Allen, R. C., McDonald, D. P., and Gildea, J. P., “Terrain

- appreciation in virtual environments: spatial knowledge acquisition,” *ARI Technical Report 1056*, Alexandria, VA, U.S. Army Research Institute for the Behavioral and Social Sciences, 1997.
- [164] Smith, S., Grinstein, G.G., and Bergeron, R.D., “Stereophonic and surface sound generation for exploratory data analysis,” *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 125-132, New York: Association for Computing Machinery, 1990.
- [165] Smith, D. and Walker, B., “Tick-marks, axes, and labels: the effects of adding context to auditory graphs”, *Proceedings of the International Conference on Auditory Display (ICAD)*, Kyoto, International Community for Auditory Display, July 2-5, 2002.
- [166] Sripada, S.G., Reiter, e., Hunter, j., and Yu, j., “Generating English summaries of time series data using the gricean maxims,” *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 187-196, New York: ACM Press, 2003.
- [167] Stevens, R. D., Brewster, S. A., Wright, P. C., and Edwards, A. D. N., “Design and evaluation of an auditory glance at algebra for blind readers,” In G. Kramer (Ed.), *Auditory Display: The Proceedings of the Second International Conference on Auditory Display*. Addison-Wesley, 1994.
- [168] Stevens, R.D., *Principles for the Design of Auditory Interfaces to Present Complex Information to Blind People*, PhD Thesis, Department of Computer Science, The University of York, Heslington, York, UK. YO1 5DD., July 1996.
- [169] Strelow E.R., Brabyn J.A., “Locomotion of the blind controlled by natural sound

- cues,” *Perception*, 11(6), pp. 635-40, London: Pion, 1982.
- [170] Sun Microsystems, *Java Web Start Technology*,
<http://java.sun.com/products/javawebstart/>, (accessed on 09/06/2003)
- [171] Thurlow, W.R., and Jack, C.E., “Certain determinants of the ‘ventriloquism effect’”, *Perceptual and Motor Skills* 36(3, Pt. 2), pp. 1171-1184, 1973.
- [172] Tukey, J. W., *Exploratory Data Analysis*, Addison-Wesley, 1977.
- [173] Vecchi, T., “Visuo-spatial imagery in congenitally totally blind people”, *Memory*, 6(1), pp. 91-102, England: Taylor & Francis, Jan 1998.
- [174] Vickers, P., *CAITLIN: Implementation of the Musical Program Auralisation System to Study the Effects on Debugging Tasks as Performed by Novice Pascal Programmers*, PhD Thesis, Loughborough University, UK, 1999.
- [175] Vickers, P. and Alty, J.L., “Towards some organizing principles for musical program auralisations”, *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 1998.
- [176] VirTouch, *Tactile Solutions for the Blind - VTPlayer*,
<http://www.virtouch2.com/index.htm> (accessed on 10/06/2005).
- [177] VisuAide, *GPS Orientation Solutions – Trekker*, 2003,
<http://www.visuaide.com/gpssol.html> (accessed on 23/10/2003)
- [178] W3C, *Web Accessibility Initiative (WAI)*, <http://www.w3.org/WAI/>
- [179] W3C, *Web Content Accessibility Guidelines*,
<http://www.w3.org/WAI/intro/wcag.php>
- [180] Walker, A., Brewster, S.A., McGookin, D., and Ng, A., “Diary in the sky: a spatial audio display for a mobile calendar”, *Proceedings of the British HCI*

- Group Annual Conference (IHM-HCI)*, Lille, France, Springer, pp. 531-540, 2001.
- [181] Walker, B.N. and Cothran, J.T., “Sonification sandbox: a graphical toolkit for auditory graphs”, *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2003.
- [182] Walker, B.N. and Ehrenstein, A., “Congruency effects with dynamic auditory stimuli: design implications”, *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 1997.
- [183] Walker, B.N. and Lane, D.M., “Psychophysical scaling of sonification mappings: a comparison of visually impaired and sighted listeners”, *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2001.
- [184] Walker, B., “Magnitude estimation of conceptual data dimensions for use in sonification”, *Journal of Experimental Psychology*, pp. 211-221, Washington, D.C.: American Psychological Association, April 8, 2002, also as PhD Thesis at Rice University, 2000.
- [185] Walker, B., Kramer, G., “Mappings and metaphors in auditory displays: an experimental assessment”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 71-74, International Community for Auditory Display, 1996.
- [186] Wang, Z., and Ben-Arie, J., “Conveying visual information with spatial auditory patterns”, *IEEE Transactions on Speech and Auditory Processing*, 4, pp. 446-455, Washington, D.C.: IEEE Computer Society, 1996.

- [187] Wenzel, E.M., Arruda, M., Kistler, D.J., and Wightman, F.L., “Localization using nonindividualized head-related transfer functions”, *Journal of the Acoustical Society of America*, 94(1), pp. 111-123, 1993.
- [188] Wightman, F.L., and Kistler, D.J., “Headphone simulation of free-field listening. I. Stimulus synthesis”, *Journal of the Acoustic Society of America*, 85, pp. 858-867, 1989.
- [189] Willuhn, D., Schulz, C., Knoth-Weber, L., Feger, S., and Sallet, Y., “Developing accessible software for data visualization”, *IBM Systems Journal*, 42 (4), <http://www.research.ibm.com/journal/sj/424/willuhn.html>, 2003.
- [190] Wilson C. M. and Lodha S. K., “Listen: a data sonification toolkit”, *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 35-40, International Community for Auditory Display, 1996.
- [191] Winberg, F. and Hellström, S.O., “Qualitative aspects of auditory direct manipulation: a case study of the towers of Hanoi”, *Proceedings of the International Conference on Auditory Display (ICAD)*, International Community for Auditory Display, 2001.
- [192] Yu, W. and Brewster, S.A., “Comparing two haptic interfaces for multimodal graph rendering”, *Proceedings of the IEEE Symposium On Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Orlando, Florida, Washington, D.C.: IEEE Computer Society, March 24-25, 2002.
- [193] Zhao, H. and Shneiderman, B., “Colour-coded pixel-based highly interactive Web mapping for geo-referenced data exploration”, *International Journal of Geographical Information Science*, 19(4), pp. 413-428, Abingdon, UK: Taylor &

Francis, 2005.

- [194] Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R., "Sonification of geo-referenced data for auditory information seeking: design principle and pilot study", *Proceedings of the International Conference on Auditory Display (ICAD)*, Sydney, Australia, International Community for Auditory Display, July 6-10, 2004.
- [195] 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*, pp. 26-35, Washington, D.C.: IEEE Computer Society, Apr-Jun 2005.
- [196] Zhao, H., Plaisant, C., and Shneiderman, B., "I hear the pattern - interactive sonification of geographical data patterns", (*Extended Abstracts*) *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1905-1908, New York: Association for Computing Machinery, 2005.
- [197] Zhao, H., Plaisant, C., Shneiderman, B., Lazar, J., "A framework for auditory data exploration and evaluation with geo-referenced data sonification", *HCIL Technical Report* (HCIL-2005-28), Human-Computer Interaction Lab., University of Maryland, College Park, Maryland, U.S.A.
- [198] Zhao, H., Plaisant, C., Shneiderman, B., Delogu, F., Palmiero, M., Pasqualotto, E., Olivetti, M., "iSonic: interactive sonification for geo-referenced data exploration for the vision impaired", (*Demonstration*) *Proceedings of Interact*, Toronto, Canada: Human-Computer Interaction Resource Network (HCIRN), 2005.

- [199] Zhang, J. and Norman, D. A., “Representations in distributed cognitive tasks”, *Cognitive Science* 18, pp. 87–122, Cognitive Science Society Inc., 1994.
- [200] Zheng, X. S., McConkie, G.W., and Schaeffer, B., “Navigational control effect on representing virtual environments”, *Proceedings of the Human Factors and Ergonomics Society Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society (HFES), October 2003.
- [201] Zimler, J., and Keenan, J.M., “Imagery in the congenitally blind: how visual are visual images?” *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, pp. 269-282, Washington, D.C.: American Psychological Association, 1983.
- [202] Zotkin, D..N., Duraiswami, R., and Davis, L.S., “Customizable auditory displays,” *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 167-176, Kyoto, International Community for Auditory Display, 2002.
- [203] Zotkin, D. N., Duraiswami, R., Davis, L. S., Mohan, A., and Raykar, V., “Virtual audio system customization using visual matching of ear parameters”, *Proceedings of IEEE International Conference on Pattern Recognition (ICPR)*, Quebec City, Canada, pp. 1003-1006, Washington, D.C.: IEEE Computer Society, 2002,
- [204] Zotkin, D..N., Hwang, J., Duraiswami, R., and Davis, L.S., “HRTF personalization using anthropometric measurements,” *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 157-160, Mohonk, Washington, D.C.: IEEE Computer Society, 2003.

- [205] Zotkin, D. N., Duraiswami, R., Grassi, E., Gumerov, N.A., “Fast head related transfer function measurement via reciprocity,” to appear in *Journal of Acoustic Society of America*.