INTERACTING WITH AN INFORMATION SPACE USING SOUND: ACCURACY AND PATTERNS

Jason Holmes

School of Library and Information Science,
Kent State University,
Kent, Ohio
USA.
jholmes@slis.kent.edu

ABSTRACT

Human auditory perception is suited to receiving and interpreting information from the environment but this knowledge has not been used extensively in designing computer-based information exploration tools. It is not known how accurate humans can be in navigating an auditory display. Furthermore, it is not known if listeners will conform to known pattern search techniques in a search task using sound alone. An auditory display was created using PD (Pure Data), a graphical programming language used primarily to manipulate digital sound. The visual interface for the auditory display was a blank window. The auditory interface was based on ground level ozone concentration data. When the cursor is moved around in this window, the sound generated changes based on the underlying data value at any given point. An experiment was conducted to determine how accurately subjects were able to locate the highest concentration level using the auditory display. The four attributes of sound tested were frequency-sine waveform, frequency-sawtooth waveform, loudness and tempo. Results indicate that sonic display of data yields less resolution than visual. It is also shown that people will generally utilize recognizable search patterns when exploring the information space.

1. INTRODUCTION

There are two inter-related concepts of accuracy explored in this paper. One is how accurately sound conveys the data values it is meant to represent. There is currently no standard vocabulary to guide the communication of data through sound. It is assumed that there is a real relationship between the underlying data and the representation of that data through the control of various aspects of sound. If a data value doubles, then the sound parameter (pitch, loudness, tempo) would also double. See Walker, Kramer, Lane[1] and Walker[2] for more on the issues with perception and psychophysical scaling in sonification. As a design objective for the experiment reported on here, I have attempted to maintain a simple algorithmic relationship between data and the sound generated.

The other type of accuracy that is under primary consideration in this study is the capability of human subjects to accurately identify high data points in a large data set with a computer display using sound only. In navigating a computer display of data visually, accuracy is seldom a concern. Using a scrollbar or clicking a 10x10 pixel icon using one's vision is trivial from the perspective of the accuracy needed to accomplish this task. Designers of auditory displays, on the other hand, are in need of research into the accuracy that is possible in this environment. Establishing the accuracy with which humans can navigate using sound alone is an early step in integrating sound into a multi-modal information system.

2. RESEARCH DESIGN

Each subject was asked to complete the task of finding the point with the highest concentration of ozone with each aspect of sound represented. The treatments were comprised of four sound types; frequency-sine waveform, frequency-sawtooth waveform, tempo, and loudness.

Of concern when using a repeated-measures design is the possibility of training effects confounding the experiment. To counter balance these effects, two precautions were taken. First, the treatments were randomly presented to counteract learning based on the presentation order of the sounds alone.

Second, the underlying data used to generate the sound were rotated so that the high point was in a different place for each treatment. Each subject heard all four sound treatments but the high point moved each time. Map A is the original data. Map B is same data transposed by switching the columns and rows. Map C was created by taking the data in Map A and flipping them horizontally. Map D was created by taking the data in Map B and flipping them vertically. Figure 1 illustrates the rotation of the data values. This rotation of the underlying data counteracts the subject's natural proclivity to simply return to the point they thought was the highest in the previous treatment. The order of pattern rotation of the data for each subject was map A first, map B second, map C third, and map D fourth.

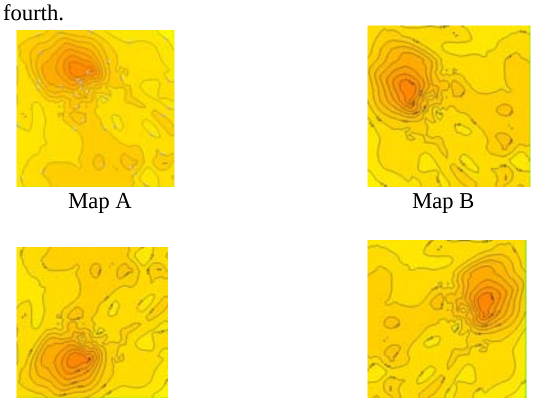


Figure 1: *Contour maps showing the different locations of the high point*

2.1. Limitations

Hearing tests were not performed on the subjects prior to the experiment. Subjects were excluded from the study if they did not have self-described normal hearing. Subjects were not given any indication as to the location or size of the target for which they were searching. They were also not given any suggestion as to the topography of the information space. It was assumed that the subjects would explore the information space before making their determination of the location of the high point. Only 24 subjects participated in the experiment due to funding constraints.

2.2. Source Data

The Texas Natural Resources Conservation Commission provided data for the ground level ozone concentration levels in Dallas-Fort Worth for the date of June 20, 1995 . Values for the concentration level range between 0 and 0.135315. These values are too small to directly control sound, so rather than scale them one time to get a usable range for frequency, then again (and differently) to get a usable duration range, a transformation of the values was needed. A maximum/minimum transformation was performed to achieve values between 0 and 1. By transforming the data, it was possible to use one set of values that can be linearly transformed as needed to accommodate the various ranges needed by each of the sound generators. This transformation is also the basis of the measurement of numeric accuracy. The following formula was used in the maximum/minimum transformation.

$$Z_i = \frac{(X_i - X_{min})}{X_{max} - X_{min}} \quad (1)$$

Where Z_i is the transformed value of X_i. [3]

2.3. Apparatus

Subjects sat at a table with a 1 Ghz Intel® Pentium III® computer using the Microsoft® Windows® 2000 operating system in a lab at Kent State University. The computer monitor is 19 inches with a resolution of 640x480. The subjects used headphones that can be adjusted to comfort and fit. A mouse designed for ambidextrous use was the pointing device.

The experiment was written using PD (version .34) [4], a visual programming environment with an object-oriented language used primarily to manipulate and control sound production and GEM (version .84) [5], a runtime add-in to PD used to generate real-time computer graphics. The experiment was controlled remotely using GrIPD (version 0.1.)[6], an interface design tool for PD. All three of these programs are in the beta/experimental stage but the portions of the program used from this experiment are stable. They are available for free under the conditions of the GNU Public License (GPL).

Using PD and GEM, an interface was developed that consists of a 480 x 480 pixels black window that constitutes the perimeter of the information space. This visually empty window was centered on the screen. Each of the transformed values for ozone is read into a 10 x 10 pixel area in the information space. This pixel size, 10 x 10, is a commonly used size for icons in a standard visual interface. The screen resolution was set to 640 x 480 to give the maximum physical size for each square in the coordinate. Using the mouse to move the cursor around in a black box, the sounds the subject experiences changed depending on the data value for the point they are currently on and the treatment group. Sound is stopped when the mouse pointer is outside the information space. The trial is over when the subject clicks on any mouse button.

2.4. Mapping Sound using PD

The overall design objective for encoding sound in this experiment was to create simple sounds that were in the normal range of hearing for most people. The data were encoded to the range of sound parameters that would be easy to hear and distinguish changes in the sound as representative of the underlying data.

2.4.1. Frequency-Sine Waveform and Sawtooth Waveform

The design goal was to place the frequency in the low-normal range of human hearing. For both of these experimental treatments, the range of frequency was from 0 to 2000Hz. The underlying data values range from 0 to 1. Consequently, a cell value of 1 would generate a sound with a frequency of 2000Hz; a cell value of .5 would generate at sound with a frequency of 1000Hz, and so on.

2.4.2. Amplitude

The design goal with amplitude was to create a broad range of possible loudness levels. The upper bound of this range needed to be set with the comfort and safety of the research subjects in mind. Decibel levels that are too high can be annoying at best and painful or destructive to the ear at worst. Bearing these design considerations in mind, the amplitude experimental treatment has a range of 0 to 80 dB at a frequency of 800Hz. Again, the underlying data values range from 0 to 1. A cell value of 1 would generate a sound with amplitude of 80dB; a cell value of .5 would generate a sound with amplitude of 40dB, and so on.

2.4.3. Tempo

The design goal with tempo was essentially to replicate a Geiger counter. A faster tempo of sounds indicates a higher underlying data value. The tempo experimental treatment was created using the MIDI components in PD. The pitch used was 84, or C above middle C in musical terms, and the instrument used was number 116, woodblock. This combination provided a reproduction of the “click” sound used in a Geiger counter. The duration (or tempo) of the sound was controlled using the underlying data values. The range of duration was from 50 milliseconds to 500 milliseconds. A cell value of 1 would generate a rapid stream sounds with a duration of 50 milliseconds; a cell value of 0 would generate a slow stream of sounds that had a duration of 500 milliseconds.

2.5. Variables and Measurement

One factor, aspect of sound, with 4 levels, frequency-sine waveform, frequency-sawtooth waveform, loudness and tempo, are the independent variables in the experiment. There are three dependent variables under consideration: distance accuracy, numeric accuracy, and time on task. Time on task is used to measure the efficiency of each sound in accomplishing the task. It is measured as the beginning of each trial until the time that the subject has selected a point they think had the highest concentration. The unit of measure is seconds.

The measurement numeric accuracy is the difference between the value in the cell selected by the subject and 1. Scaling of the original data values is discussed below.

$$1 - x = \text{Error} \tag{2}$$

Selecting the cell that contains the highest concentration would result in a score of zero error. Selecting any other cell would result in a positive number between zero and one, the magnitude of which reflects the degree of error. There is only one cell that contains the highest concentration. A numeric accuracy value of .25 would mean that the subject clicked on a point that had only 75% of the intensity of the actual high point.

Measurement of distance accuracy is the Euclidean distance from the point selected by the subject as the highest concentration to the point of actual highest concentration. The formula used to calculate the distance between two points, O_1 and O_2, in a 2-d grid is

$$D(O_1, O_2) = \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2} \tag{3}$$

2.6. Subjects

Subjects were drawn from graduate students in the School of Library and Information Science. Because previous studies [7],[8] have suggested that dominant hand may be a demographic factor affecting performance in sonification studies, I attempted to have equal representation of left and right-handed subjects. The distribution of the sample was nine left-handed women, six right-handed women, three left-handed men and six right-handed men.

Subjects were solicited for participation in the study through School's general listserv. Subjects must have self-described "normal" hearing. Because the sound mappings are intentionally designed to coincide with the normal range and parameters of hearing, pre-testing for hearing ability was not conducted.

2.7. Data Gathering

The time on task was collected to determine the relative effectiveness of each treatment. Using GEM, an add-on to PD, the X position, Y position, and click state of the mouse were also collected and stored in a text file for further analysis. By recording the position of the mouse throughout the experiment, the path or search pattern can be recreated and analyzed for each subject on each trial.

The data of most importance in this experiment is the coordinates at the time of the mouse click. These data provide the basis for analysis of both measures of accuracy outlined above. When the subject clicks on the point he or she believes to be the highest concentration, the sound ceases and the timer stops.

3. RESULTS

Twenty-four subjects completed 96 total trials (4 trials each). Only one subject on one trial was able to click on the precise highest data point. Subjects did not express any difficulty in understanding the task or the interface. Overall, performance was good. 54% of the trials were within 50 pixels of the highest data point. Table 1 shows descriptive statistics for accuracy aggregated over all 24 subjects.

	Distance Accuracy	Numeric Accuracy	Time (seconds)
Mean	6.77	.26	33.5
Median	5.00	.22	23.5
Minimum	0	0	5.0
Maximum	32.00	.73	248.5

Table 1: *Descriptive Statistics of Aggregate Performance*

Figures 2-5 illustrate the points clicked by each subject on that map. The top number for each point is the subject number. The bottom number is the treatment, or aspect of sound used. 1 is frequency-sine waveform, 2 is frequency-sawtooth waveform, 3 is duration, and 4 is amplitude. The relative size of the points corresponds to aspect of sound with 1 having a small diameter and 4 having a larger diameter. All subjects had the same order of presentation of the maps; map A first, map B second, map C third, and map D fourth. The "X" marks the high point.

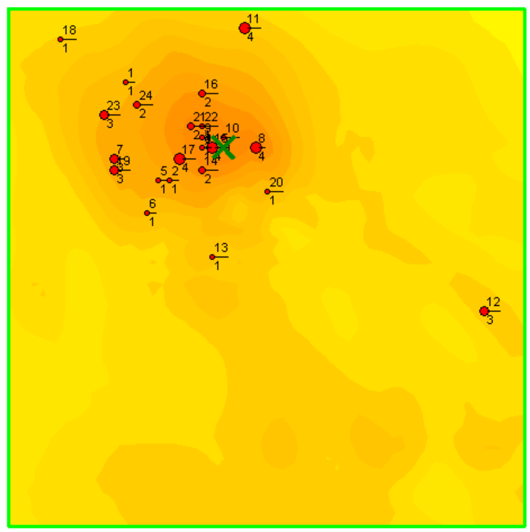


Figure 2. *Results map A*

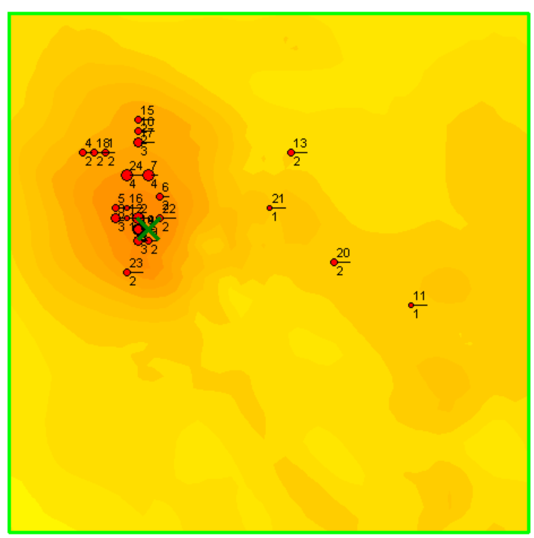


Figure 3. *Results map B*

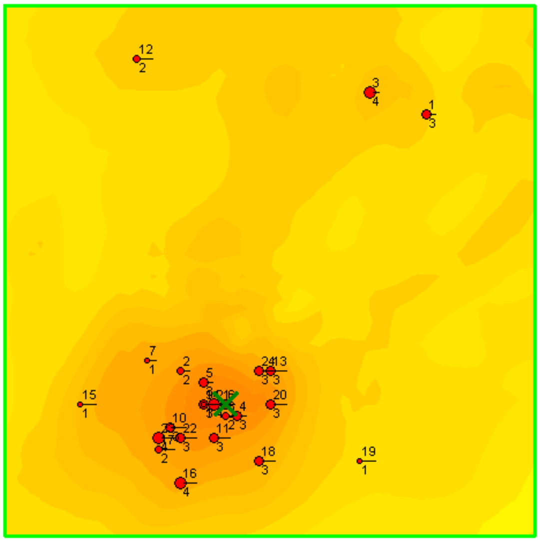


Figure 4. *Results map C*

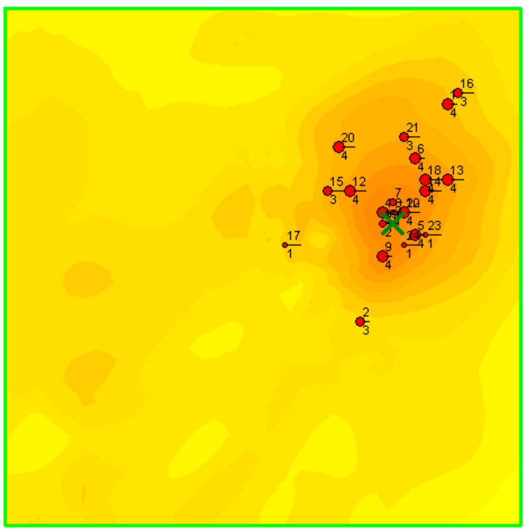
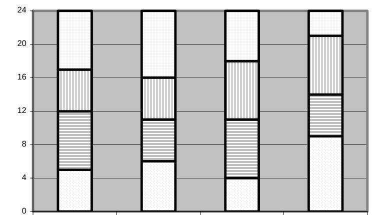


Figure 5. *Results map D*

The density of points around the high point appears to be lowest in Map A. The distribution of points around the high point in maps B and C is tighter with a few outliers. Map D has the highest density of points near the high point. The outliers are closer than in the previous maps also, suggesting that by the fourth trial, the subjects were beginning to better understand the sonic space regardless of the sound they heard.

Although ANOVA analysis showed there was no statistical difference in the treatment used, closer observation of the data reveals some interesting findings. Figure 6 illustrates the frequency of treatment for distance accuracy divided into quartiles. The distribution of treatment frequencies in the first quartile is fairly even indicating there is no benefit in accuracy of using one treatment over another. It may be the case, however, that a treatment may have led to more inaccuracy than the others. Of particular note is the high occurrence of the Frequency-Sin treatment in the 4th quartile. 9 out of the 24 least accurate trials used Frequency-Sin. Only 3 of the 24 least accurate trials were mapped to Loudness. This is contrary to previous studies suggesting that amplitude is not a good conveyor of information [9].



	1st quartile	2nd quartile	3rd quartile	4th quartile
Loudness	7	8	6	3
Tempo	5	5	7	7
Frequency-Sawtooth	7	5	7	5
Frequency-Sin	5	6	4	9

Figure 6. *Distribution of distance accuracy by quartile*

3.1. Discussion of Accuracy and Sound Mappings

A simplistic mapping approached was used to encode that underlying data values to each of the 4 sound treatments. The intention was to reduce any bias or preconception the researcher may have and simply let the data drive the sound with as little interference as possible. A change in one unit of the underlying data will always produce a change of one unit in frequency, or loudness, or tempo. A consequence of this approach is that there is not necessarily an equally discriminable quantity as the step size for representing data points. In other words, a one unit increase in loudness (1 dB) is *perceptually* different than a one unit change in the frequency (1 Hz). There is no perceptual scaling of the sounds. Therefore, comparisons across sound treatments must be made with this in mind.

The outliers in the accuracy measures were intriguing. Looking at Figure 3, it is interesting that 4 different subjects selected point nowhere near the high point using the two frequency based sounds. In Figure 4, there are 3 different subjects who are again nowhere near the highpoint. These observations prompted the researcher to analyze the search patterns that would lead to such inaccurate trials.

4. SEARCH PATTERNS OBSERVED

None of the subjects participating in the experiment had used sound to navigate an information space before. They were given a task they had never done using a novel interface and an unfamiliar form of data representation. How do people begin searching in these unfamiliar conditions? What patterns or techniques are employed to accomplish the experimental task? In looking at the search patterns, I am not attempting to establish a link between the search pattern used and the resulting measurement of accuracy. I am simply making a subjective classification and qualitative assessment of the types of search patterns employed to accomplish the task. A subject may employ one of the search techniques and still not be very accurate, or they may be very accurate in spite of using no detectable systematic pattern.

Several types of search techniques and search patterns were used to complete the task of finding the point with the highest concentration of ozone. I have identified 4 basic categories of search pattern by plotting the movement of the cursor as the subjects attempted to find the high point. The large dot indicates the high point. The point where the subject began is marked with a "0". The point where the subject clicked when they thought they found the high point is marked with a "1". In some instances, the "1" marker will be obscured by the large high point dot. Also note that the distances between the dots in the path indicate the speed of the mouse. The farther apart the dots are, the faster the mouse was moving.

Some of the terms and their basic definitions used in the following sections are taken from search theory, a sub-field within operations research. Much of search theory is focused on development of efficient and effective search patterns for reconnaissance of missing ships, planes, and people [10]. The patterns developed by search theory are visual search patterns of physical space, but there is some crossover in the types of patterns used in the auditory interface used in the experiment to search in a virtual space. A particularly important point of departure when comparing physical, visual search patterns described in search theory to virtual search patterns is the concept of effort and expense. There is essentially no effort or expense associated with the exploration of the virtual sonic space used in the experiment. There is no penalty for going back over the same area many times, nor is there any particular urgency or stress associated with failure to find the target or the length of time it takes to accomplish the task.

There is no external need to employ any particular search technique, and yet the subjects do tend to fall into predictable and known patterns of searching.

4.1.1. *Parallel Sweep*

The parallel sweep is used when uniform coverage of an area is desired and the area is unfamiliar. It is an efficient method of searching a large area in a minimum amount of time. Several subjects used the horizontal parallel sweep similar to the one seen in figure 9.

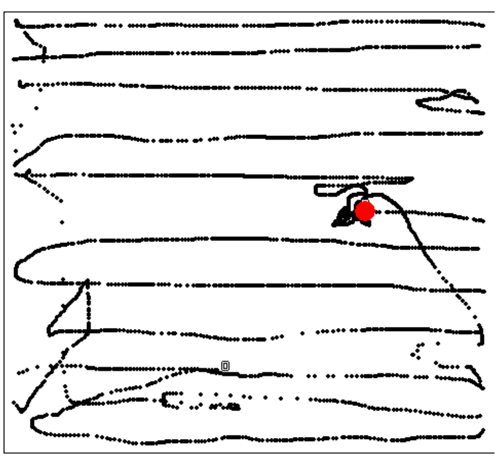


Figure 9. *Example of parallel sweep: Subject 04 map D*

4.1.2. *Quadrant Search*

The quadrant search pattern is one in which the searcher mentally breaks down the screen into quadrants to divide the area into a more manageable size. Within the quadrants, the searcher may use another pattern to search each quadrant, such as a parallel sweep. Figure 10 illustrates the segmenting of the space as well as the search sub-pattern of a parallel sweep in the bottom left quadrant.

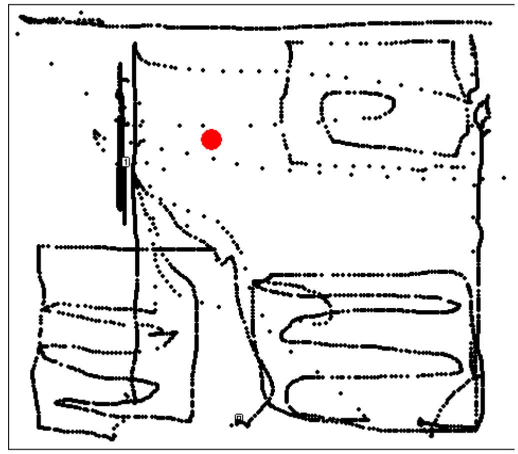


Figure 10. *Example of quadrant search: Subject 19 map A*

4.1.3. *Sector Search*

A sector search pattern begins once the approximate location of the high point is located. In this pattern, the searcher explores out from the approximate location of the high point and returns again, then conducts another exploration in another area, and returns again. This is repeated until they are confident that the space is adequately explored. Figure 11 illustrates the sector search pattern. Notice the density of activity around the high point. The path depicted in figure 11 begins as a perimeter search, but once the approximate location of the high point is found, the subject begins a sector search radiating out from the high point.

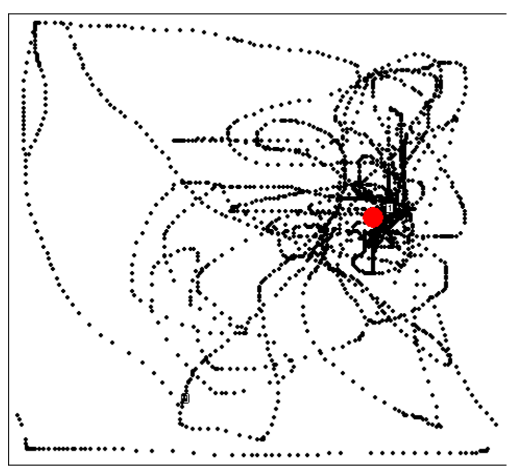


Figure 11. *Example of sector search: Subject 22 map D*

4.1.4. Perimeter Search

The perimeter search is one in which the boundaries of the space are explored, but little or none of the middle is traversed. Some subjects used a circle to circumscribe the border (Figure 12). Other subjects used a square shaped pattern turning at a 90-degree angle following the border created by the window as shown in Figure 13. This type of search pattern would typically lead to inaccuracy given that the high point on all of the maps are not located at the perimeter.

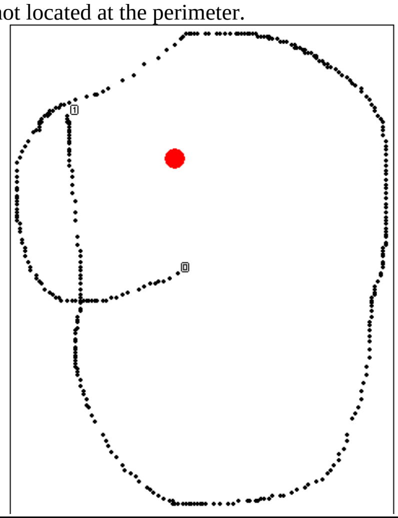


Figure 12. *Example of perimeter search: Subject 02 map A*

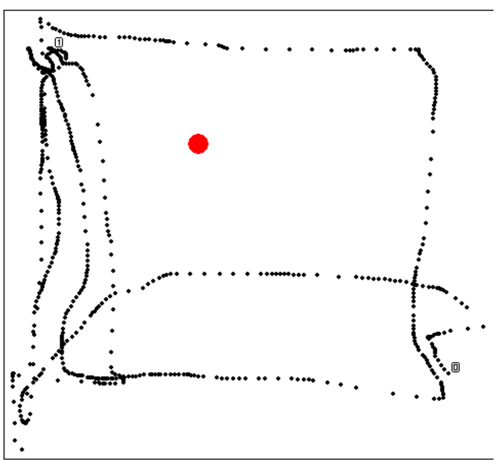


Figure 13. *Example of perimeter search: Subject 18 map A*

4.1.5. No Discernable Systematic Technique

For some searchers, no discernable systematic technique was employed in exploring the space to accomplish the task. For these search patterns, there is no attempt to thoroughly explore the information space. Figure 14 illustrates the path used in the only trial to actually select the high point exactly.

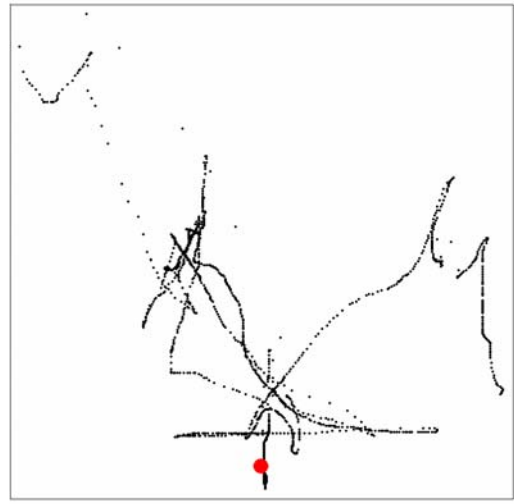


Figure 14. *Example of no discernable systematic technique: Subject 06 map C*

4.1.6. Low Effort Searches

Some subjects were not very accurate because they did not explore the space sufficiently to hear what the sounds were in other areas. This subject did not hear the sound at the highest point. Even though there was essentially no effort involved in exploration, some subjects simply chose not to explore. Figure 15 illustrates a typical low effort search. Very little coverage of the area is attempted.

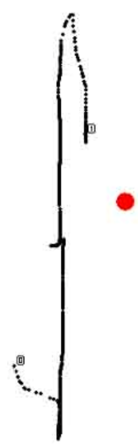


Figure 15. *Example of low effort*

4.2. Analysis of Search Patterns

The search patterns of each subject were analyzed to see if there were any tendencies based on demographic characteristics. 96 total patterns were analyzed. The most common technique employed was the perimeter search (41%). The next most common was the sweep pattern (36%), followed by quadrant (9%), no distinguishable pattern (8%), and sector (5%).

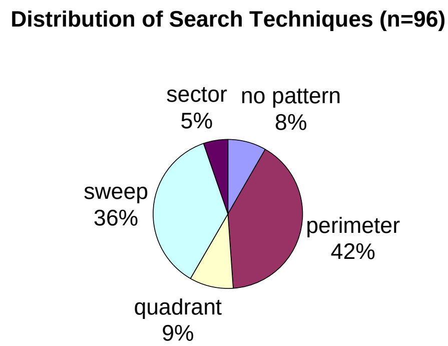


Figure 16: *Distribution of Search Techniques*

4.3. Discussion of Search Patterns

The subjects tended to use the same technique regardless of the sounds they heard. 8 subjects (33%) used one search technique exclusively. Of these 8 subjects, 7 used the perimeter technique, 1 used sweep exclusively. Another 9 subjects (38%) used the same technique in 3 out of the 4 trials. This consistency in the application of a searching technique has several notable points. First, the same technique was employed regardless of the sound treatment. This would indicate that the subjects brought with them a technique that was not altered by the change in the treatments used in the auditory interface. The subjects were given no experimental feedback that might prompt them to change their search pattern to one that might be more effective. Left to their own means, the subjects tended to continue with the application of the search pattern they felt most comfortable.

Second, the most common type of search pattern (perimeter search) was also the least effective given that the location of the highest point on all four map configurations was located towards the interior of the information space. In these cases, the subject was less likely to hear a change in the sounds they were hearing because of the low intensity of the sounds generated at the borders of the information space. Because they typically did not explore the interior they would not hear the more intense sounds that might lead them to the highest point. The perimeter technique is the least thorough of the systematic techniques.

Even though there was essentially no effort involved in exploring more thoroughly by applying a different search pattern, the subjects tended to use the perimeter search pattern. This could be accounted for by assumptions the subjects made about the nature of the information space. If a subject assumed that the information space was divided into horizontal or vertical bands of sound, the perimeter search would be an effective technique in discovering the target. There was no indication given by the researcher as to the nature of the information space or the target.

It would seem that some subjects took the experimental task seriously by systematically exploring the information space. Other subjects did not seem to be interested in exploration, but instead made a quick "stab" in the general direction of the high point. The case could be made that those subjects who explored liked the sonic interface and those who did not explore did not like the interface. It may well be the case that auditory display is not for everyone. Some will like it and make use of it, others will not.

5. CONCLUSION

This paper has reported on an experiment conducted to investigate the accuracy with which subjects could navigate an auditory information space. The paper also analyzed the search patterns employed by subjects when attempting this task. Only 1 out of 96 trials was able to locate the exact high point indicated by the most intense sound. 54% of the trials were within 50 pixels of the highest data point. It may be that the design of the experiment was flawed. Perhaps a 10x10 pixel cell is not a large enough target. A larger target may have improved accuracy, but the precision of the measurement would have been decreased.

The 2 most commonly occurring search patterns were the perimeter search (41%) and the sweep search (36%). Subjects tended to use the same search technique for each trial, regardless of the sound they were hearing to navigate.

Kramer suggests, "Since our auditory systems evolved to hear complex sounds, might complex sonifications convey data more effectively than those employing pure waveforms?" [11] This seems to have some face validity and is supported by Kramer anecdotally. Based on the findings presented in this paper, there may be some validity to using more complex sounds over the simple sounds used here. Listening to a sine tone, even when the frequency or loudness is changing, is fairly uninteresting and not engaging for very long. More complex sounds may hold the attention of the listener longer thus encouraging them to make more use of the auditory display. Myriad complex sounds can be generated electronically providing a rich source for sonification. Based on the search patterns (or lack of a pattern) presented here, it is plausible that those subjects who did not explore were simply not engaged enough by the sound to bother with accomplishing the task. Future experiments are being planned to investigate the issues raised by this research.

6. REFERENCES

[1] Walker, B. N., Kramer, G., and Lane, D. M., "Psychophysical scaling of Sonficiation Mappings," *Proceedings of the Proceedings of the Sixth International Conference on Auditory Display, ICAD 2000,* Atlanta, GA, 2000.

[2] Walker, B. N., *Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification*, doctoral dissertation, Dept. Psychology, Rice University, 2000.

[3] Einax, J.W., Zwainger H.W., and Geib, S., *Chemometrics in environmental analysis.* VCH, Weinheim, 1997.

[4] Puckette, M. S., Pd (Pure Data). [Online]. Available: http://crca.ucsd.edu/~msp/software.html.

[5] Danks, M., GEM (Graphics Environment for Multimedia). [Online]. Available: http://gem.iem.at/.

[6] Sarlo, J., GrIPD:Graphical Interface for Pure Data. [Online]. Available: http://crca.ucsd.edu/~jsarlo/gripd/2001.

[7] Mansur, D. L., Blattner, M. M., and Joy, K. I., *The Representation of Line Graphs through Audio Images*, UCRL-91586, University of California, Davis, CA, 1984.

[8] Walker, B. N., *Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification*, doctoral dissertation, Dept. Psychology, Rice University, 2000.

[9] Bly, S., *Presenting Information in Sound*, UCRL-87087, Lawrence Livermore National Laboratory, Livermore, CA, 1982.
[10] Civil Air Patrol - Reserve Air Patrol. *Mission observer course manual*, Civil Air Patrol, 1999.
[11] Kramer, G., "Mapping a single data stream to multiple auditory variables: A subjective approach to creating a compelling design," Proceedings of the Third International Conference on Auditory Display, ICAD /06, Palo Alto, CA, 1996.