SONIFICATION OF STATISTICAL GRAPHS

S. Camille Peres and David M. Lane

Rice University, Houston, TX
peres@rice.edu, lane@rice.edu

ABSTRACT

Two experiments are presented that compare the effectiveness of different parameters of sound for the auditory presentation of box plots. Temporal mapping was found to be better than pitch or panning mapping. In the first experiment, the mapping condition that used two dimensions (the redundant condition) did not result in a better performance than those mappings that used one dimension. However, subjects showed a strong preference for the redundant condition. Finally, subjects' overall level was not very high and performance did not increase with practice as much as might have been expected.

1. INTRODUCTION

Statistical graphs are an important tool for communicating parameters of a dataset and it is well documented that they can be effective when presented visually [1]. However, sometimes a visual presentation is not practical such as with devices without displays or with small displays (i.e. cell phones and PDAs), for tasks when the eyes are busy [2], and for individuals with visual disabilities [3]. Sonification, or the representation of data through sound or non-speech audio, could be effective in these situations and this paper focuses on the use of sonification for one type of statistical graph: box plots.

John Flowers and his colleagues conducted a series of studies comparing the perception of auditory and visual presentations of statistical graphs [4-7]. They used different dimensions of sound both independently and together to represent the parameters of a data set. For instance, a frequency polygon was represented by using pitch to represent the Y-axis, and loudness for the values on the X-axis. Of interest was the relationship between statistical properties of data and similarity judgments of visual and auditory graphs. In general, they found that the perception of auditory and visual displays was similar: judgments of both types of graphs were influenced by skew, spread and central tendency. However, judgments of visual graphs were relatively more influenced by skew whereas judgments of auditory graphs were relatively more influenced by central tendency.

This pioneering research by Flowers et al. provides pertinent information regarding auditory displays of statistical data. However, many issues regarding the designs of these displays remain unresolved. One issue involves the relative effectiveness of various sound dimensions for representing data. Although research [8, 9] suggests that spatial location and temporal design may be effective dimensions for use in auditory designs, there is relatively little data on this topic. One of the goals of the present research was to compare the relative effectiveness of displays using the temporal aspects of sound, spatial location (panning), and pitch. Another goal was to investigate the effects of using redundant dimensions. There is reason to believe that in some contexts, using two sound dimensions together in a redundant fashion is a better representation of the data than could be achieved by the dimensions used individually [10]. Finally, we were interested in the subjective impressions of auditory graphs. It is often assumed that subjects will prefer conditions where they have a better performance, but previous research has provided counter examples [3].

2. EXPERIMENT 1

This experiment compared the effectiveness of pitch, panning, and redundant sound dimensions for presenting the statistical information normally contained in box plots. Specifically, the box plot represented the "five number summary" of a distribution, which consists of the minimum, 25th percentile, median, 75th percentile and the maximum of the distribution. Box plots are widely used and important graphical displays. Moreover, their simplicity makes them well suited for testing the basic principles of sonification that we were concerned with.

2.1. Method

Task. A simple matching task was used in this study. Subjects listened to a sonified box plot and then selected a visual representation of the box plot from a set of four. We assumed that the visual representations would be easily interpreted. The selected box plot played and subjects were told whether or not their choice was correct. If they were incorrect, they continued selecting box plots until they had selected the correct one. For purposes of data analysis, the response on a trial was considered correct only if the subject selected the correct graph on the first trial.

The stimuli consisted of twenty box plots that varied in skew, location (central tendency), and spread. On a given trial, one of the three distracters differed from the target in skewness, one in either location or spread (determined randomly), and if this distracter differed in location, then the third distracter differed in both skew and location. If the second distracter differed in spread, then the third distracter differed in skew and spread. The target and the distracters were determined randomly on each trial.

There were three different sound dimension conditions: pitch, panning, and a redundant condition (pitch and panning). For the pitch condition, the values from the box plots were mapped to a note on the equal tempered musical scale in the range of 16 notes below and 16 notes above 440 Hz (middle C). For the panning condition, the values from the box plots were mapped to an amplitude scale that lateralized the values to points in space on an axis that goes through the ears. The redundant condition used both the pitch and panning transformations. The sounds were played

in the following sequence: the absolute minimum of the scale, the values of the box plot (minimum, lower 25th, median, upper 75th, and maximum), and the absolute maximum of the scale. The box plot sounds were played without pauses and there was a 1-second pause separating the absolute minimum and maximum values from the other sounds.

Subjects. Fifty six undergraduate students (47 females and 9 males) were randomly assigned to the three conditions with the constraint that the groups were equal size as possible given the total sample size: The pitch and redundant conditions each had 19 subjects; the panning condition had 18 subjects.

Procedure. After a fifteen-minute training session, each subject completed 50 trials. The training session took approximately fifteen minutes and the experiment took an average of 28 minutes (range: 12.7 to 50.4 minutes). After the experiment, the subjects provided some demographic information and answered several questions regarding their impression of the experiment. Subjects used a five-point scale to rate the sounds on how pleasant, annoying and difficult the tasks were, with 1 being the most positive rating and 5 the lowest.

Computers with Internet access in a laboratory environment were used for the experiment and an interactive web site was used for training and data collection. Sounds were presented to subjects over headphones.

2.2. Results

The trials were grouped into five blocks of ten and the mean proportions correct for the blocks were M_1=0.44, M_2=0.50, M_3=0.49, M_4=0.50, and M_5=0.53. These values show that the subjects improved at the task. However, the learning occurred primarily in the first block of ten trials, reaching an asymptote of only 0.50 (with chance being 0.25). The linear component of trend was significant, F (1,55) = 5.59, p = 0.025. Note that a significant linear component does not mean that the relationship is linear. There was no evidence that the shape of the learning curve differed as a function of sound dimension, F (2,53) = 0.21, p = 0.815.

The pitch condition had the highest average proportion correct (M=0.58) and the panning condition had the lowest (M=0.37). The redundant condition (M=0.50) was intermediate. The differences between the conditions were significant, F (2,53) = 10.82 p < .001, and the Tukey HSD (p < .05) showed that the panning condition was significantly lower than the other two conditions which did not differ significantly from each other.

Subjects' choices on error trials revealed the aspect of the distribution on which the error was based. For example, if a subject chose a box plot that was identical to the target in all respects except for location, this was deemed a location error. Analysis of the errors (adjusting for the fact that there were more opportunities to make skew errors than other errors) showed a lower proportion of location errors than either spread or skew errors. The means as a function of condition are shown in Table 1. Differences among conditions were significant for skew, F(2,52) = 18.48, p < 0.001, and location F(2,52) = 7.52, p < 0.01, but not for spread F(2,52) = 1.09, p = 0.343. For the skew and location conditions, the Tukey HSD test showed significant differences (p < 0.05) between the panning condition and each of the other two that did not differ significantly from each other.

As can be seen in Table 2, subjects rated the redundant condition less difficult, less annoying, and less unpleasant than the other conditions (p<0.001 for all comparisons). Note that although the subjects had a strong preference for the redundant condition, performance did not mirror this preference.

Table 1. Mean proportion of the three errors types by sound condition.

	Pitch	Panning	Redundant
Spread	0.30	0.35	0.32
Skew	0.26	0.46	0.34
Location	0.13	0.30	0.20

Table 2. Mean subjective ratings.

	Pitch	Panning	Redundant
Difficulty	2.21	2.94	1.95
Annoying	2.42	3.06	1.05
Unpleasant	2.16	2.33	1.00

2.3. Discussion

Subjects found the task difficult, improving over the first 10 trials and showing no subsequent improvement. The mean proportion correct was about 0.50 (chance was 0.25). However, there was considerable variability: the proportion correct ranged from 0.13 to 0.90.

The pitch condition had the best performance, although it was not significantly better than the redundant condition. Performance in the panning condition was worse than in either of the other conditions. However, one should keep in mind the specific mapping for each dimension may determine, in part, whether or not there would be a redundancy gain.

There was an interesting disassociation between subjective impressions and performance. Subjective impressions in the redundant condition were considerably higher than for the pitch condition even though performance in the redundant condition was slightly (though not significantly) lower than in the redundant condition.

3. EXPERIMENT 2

This experiment was conducted to extend the findings of Experiment 1 and compare the effectiveness of pitch and temporal sound dimensions for representing the "five-number-summary" box plots. Experiment 1 found that performance asymptoted very quickly. In this experiment we used 100 rather than 50 trials to see if subjects had actually reached an asymptote, or whether further practice would lead to further increases in performance. Finally, the presentation time used to play a box plot was manipulated, one time being twice as long as the other.

3.1. Method

Design. A Sound Dimension (pitch and temporal) x Presentation Time (short and long) factorial design was

employed. Both sound dimension and presentation time were manipulated as between-subjects variables.

Task. Subjects were presented with a task identical to that in Experiment 1. The pitch condition was mapped using the same method as in Experiment 1. In the temporal condition, the distances between the values of the box plot were represented by the time between the onsets of the sounds. The frequencies of all the sounds remained constant at 440 Hz.. The sounds of the box plots were played in the same sequence as in Experiment 1. The box plots with the long presentation time played for 9 sec. while those with the short presentation time played for 4.5 sec.. This resulted in four conditions: pitch-long (PL), pitch-short (PS), temporal-long (TL), and temporal-short (TS).

Subjects. Forty-eight undergraduate students were randomly assigned to one of the four conditions with the constraint that the groups were equal size as possible given the total sample size. The data from ten subjects had to be eliminated because of a problem with the interactive web site. This left a total of 38 subjects, with 10 subjects in both the PL and TS conditions and 9 subjects in the PS and TL conditions. There were 16 females and 22 males.

Procedure. Subjects completed 100 trials, which took approximately one hour. The testing design and environment were identical to Experiment 1 and, as in the first experiment, subjects answered several questions regarding their impressions of the box plots along with demographic information.

3.2. Results

The proportion of correct responses increased from 0.50 to 0.59 over the 10 blocks and the linear component of trend was significant, $F(1,34) = 7.70$, $p = 0.009$. Performance appeared to reach asymptote at the fifth block. The linear component of trend did not interact significantly with either sound dimension, $F(1,34) = 3.42$, $p = 0.073$, or presentation time, $F(1,34) = 0.001$, $p = 0.975$.

For the pitch condition, performance increased over the 100 trials for the short-presentation-time condition and decreased slightly for the long-presentation-time condition. For the temporal condition, performance for the long-presentation-time condition improved more than for the short-presentation-time condition. The Trials (linear) x Sound Dimension x Presentation Time interaction was significant, $F(1,34) = 4.98$, $p = 0.032$.

Skew errors decreased significantly over the 100 trials whereas the proportion of spread and location errors did not. For spread errors, the shape of the learning curve was different for the sound dimensions. The proportion of spread errors decreased from 0.35 to 0.22 in the temporal condition whereas it increased from 0.31 to 0.38 in the pitch condition. The Sound Dimension x Trials (linear) interaction was significant for spread errors, $F(1,34) = 6.75$, $p = 0.013$. The Sound Dimension x Trials interaction was not significant for skew or location errors.

Table 3 shows the mean proportion correct as a function of sound dimension and presentation time. The mean proportion correct for the temporal group (M=0.65) was significantly higher than for the pitch group (M=0.49), $F(1,34) = 6.06$, $p = 0.019$. Subjects in the long-presentation-time condition (M=0.62) had a higher mean proportion of correct responses than those in the short condition (M=0.52). However, this difference was not significant. Finally, there was not a significant Sound Dimension x Presentation interaction.

Table 3. Mean proportion correct as a function of sound dimension and presentation time.

	Pitch	Temporal
Long Presentation	0.52	0.71
Short Presentation	0.45	0.56

The mean proportions of errors as a function of sound condition and presentation time are shown in Table 4. As in Experiment 1, subjects made substantially fewer location errors (M=0.16) than either spread (M=0.32) or skew errors (M=0.29). Separate Presentation Time x Sound Dimension ANOVAs were run for skew, spread and location errors. The only significant effects were found for skew errors for which performance in the temporal condition was higher than in the pitch condition, $F(1,34) = 8.80$, $p < 0.01$, and performance in the long-presentation-time condition was higher than in the short-presentation-time condition, $F(1,34) = 7.64$, $p < 0.01$. The Sound Condition x Presentation Time interaction did not approach significance, $F < 1$. There were no significant effects for spread or location errors.

Table 4. Mean proportion of the three errors types as a function of sound condition and presentation time.

	Pitch	Temporal
Short		
Spread	0.37	0.31
Skew	0.40	0.29
Location	0.20	0.18
Long		
Spread	0.35	0.24
Skew	0.29	0.17
Location	0.17	0.09

The subjects rated the box plots on the same questions used in Experiment 1. As can be seen in Table 5, subjects gave more favorable ratings to the temporal conditions than to the pitch conditions on all three questions. However, the only significant difference occurred on the ratings of difficulty for which the temporal dimension was rated easier than the pitch dimension, $F(1,34) = 9.84$, $p < 0.01$.

Table 5. Mean subjective ratings.

	PL	PS	TL	TS
Difficulty	2.00	2.11	0.89	1.50
Unpleasant	2.50	2.44	2.00	1.80
Annoying	2.60	2.33	2.11	1.80

3.3. Discussion

As in the first experiment, subjects found the task difficult and most reached an asymptote at about the 50th trial of about 0.60 correct. However, the performance for the subjects in temporal-long (TL) group continued to increase

throughout the 100 trials. The subjects in the pitch-short (PS) group did not perform as well as the TL group overall, but they did show the same effects of practice.

Performance in the temporal conditions was substantially higher than performance in the pitch condition. This finding is consistent with the pattern of responses found in the subjective ratings.

As in Experiment 1, there were fewer location errors than skew or spread errors. Also consistent with Experiment 1 was the finding that sound dimension had a larger effect on skew errors than on spread errors.

4. GENERAL DISCUSSION

Subjects generally found this task difficult and they did not get much better with practice. The best performance achieved in either of the two experiments was in the temporal-long condition in which the mean proportion correct was 0.70. There was considerable variability in performance with some subjects performing approximately at chance and some performing *very* well. In the first experiment, the best three subjects were correct 78%, 80%, and 90% of the time; for the second experiment, the four top performing subjects were correct 87%, 87%, 88%, and 95% of the time. Designers of auditory statistical displays should be cognizant of this high variability.

Although Lorho et al. [8] found that subjects were relatively good at locating sounds spatially, the present data suggest that spatial location is not a good mapping technique for sonifying statistical graphs. Of the three methods examined here, temporal mapping was clearly the best. These results considered in conjunction with previous work [9] suggest that temporal mapping can be effective when sonifying box plots.

Experiment 1 found that the redundant use of the panning and pitch dimensions did not result in better performance than the pitch dimension used alone; performance in the redundant condition was slightly but not significantly worse than in the pitch condition. Consistent results have been obtained [11] using a simple task in which subjects map sounds to absolute numeric values. This research found that performance was significantly lower when temporal and pitch information was presented redundantly than when temporal information was presented alone. This provides further evidence that redundancy is not better and may be worse than the better of the two individual dimensions in this type of task. Naturally, there may be other contexts for which an auditory design using dimensions of sound redundantly would be effective.

In Experiment 1, subjects strongly preferred the redundant condition to the pitch condition even though performance was slightly better in the pitch condition. These findings are similar to those of Petrie et al. [3], who also found a disassociation between performance and preference on an auditory task.

One must be careful when generalizing these results beyond the stimuli tested. The specific values of location, skew, and spread tested no doubt affected the relative difficulty of these dimensions. Moreover, further research is needed to see if these results hold on tasks with dependent variables other than the ability to match the auditory and visual displays. It is also an open question whether the finding that the temporal dimension is better than pitch would hold in other contexts. We suspect that extracting information from the temporal presentation is more attention demanding than from the pitch dimension. If this is the case, then the use of the temporal dimension in dual task situations where visual and auditory information must be interpreted simultaneously might result in poorer performance than would the pitch dimension. Research is currently being developed to investigate this question.

Acknowledgements

We would like to thank Anikó Sándor and Adan Galvan for their work on these projects. This research was supported by NSF Grant IIS-9906818.

5. REFERENCES

[1] W. Cleveland, *The Elements of Graphing Data*. Murray Hill, New Jersey: AT&T Bell Laboratories, 1994.

[2] G. Kramer, "Some Organizing Principles for Representing Data with Sound," presented at Auditory display: Sonification, audification and auditory interfaces. Proceedings of the First International Conference on Auditory Displays (ICAD), Reading, MA, 1994.

[3] H. Petrie and S. Morley, "The Use of Non-Speech Sounds in Non-Visual Interfaces to the MS-Windows GUI for Blind Computer Users," presented at 1998 International Conference on Auditory Display (ICAD), Glasgow, UK, 1998.

[4] J. Flowers and T. Hauer, ""Sound" alternatives to visual graphics for exploratory data analysis," *Behavior Research Methods, Instruments and Computers*, vol. 25, pp. 242-249, 1993.

[5] J. Flowers and T. Hauer, "Musical versus Visual Graphs: Cross-Modal Equivalence in Perception of Time Series Data," *Human Factors*, vol. 37, pp. 553-569, 1995.

[6] J. H. Flowers, D. C. Buhman, and K. D. Turnage, "Cross-modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples," *Human Factors*, vol. 39, pp. 341-351, 1997.

[7] J. H. Flowers and T. A. Hauer, "The ear's versus the eye's potential to access characteristics of numeric data. Are we too visuocentric?," *Behavior Research Methods, Instruments and Computers*, vol. 24, pp. 258-264, 1992.

[8] G. Lorho, J. Marila, and J. Hiipakka, "Feasibility of multiple non-speech sound presentations using headphones," presented at International Conference on Auditory Displays (ICAD), Espoo, Finland, 2001.

[9] J. M. Watkins, C. D. LeCompte, N. M. Elliott, and B. S. Fish, "Short-Term Memory for the Timing of Auditory and Visual Signals," *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 18, pp. 931-937, 1992.

[10] G. Kramer, "Mapping a single data stream to multiple auditory variables: A subjective approach to creating a compelling design," presented at International Conference on Auditory Displays (ICAD), Palo Alto, California, USA, 1996.

[11] A. Sándor and D. M. Lane, "Sonification of Absolute Values with Single and Multiple Dimensions," presented at International Conference on Auditory Display (ICAD), Boston, MA, 2003.