SONIFICATION OF ABSOLUTE VALUES WITH SINGLE AND MULTIPLE DIMENSIONS

Anikó Sándor and David M. Lane

Department of Psychology
Rice University
6100 Main Street
Houston, TX 77005, USA
aniko@rice.edu lane@rice.edu

ABSTRACT

Although auditory displays are effective for the representation of patterns in data, they are generally thought to be less effective for the communication of absolute values [1, 2]. Nonetheless, there are times when it is desirable to represent absolute values with sound. Two experiments were conducted to investigate the limits of representing absolute values with sound and to compare several ways of representing the values.

Temporal representations of data led to better performance than the representation with pitch or the redundant use of temporal and pitch representations. We introduce the term "Mappable Difference" to refer to the smallest difference in a dimension that can be consistently mapped to a numeric value. Knowing the "Mappable Difference" in an auditory dimension can potentially aid display designers in determining the number of absolute values that can be represented by sound.

1. INTRODUCTION

It has been well established that sound can be an effective way to represent patterns in data [1, 3-6]. However, the consensus is that sound is less suitable for the representation of absolute values. Nonetheless, there are times when it is desirable to represent absolute values with sound. For example, an observer may wish to not only know how Y is changing as a function of X, but also whether Y has reached a critical value.

One of the goals of this research was to gather initial data on the limits of representing absolute values with sound. Specifically we were interested in determining the smallest size difference on a dimension such as pitch that can be accurately mapped to a difference on a numeric scale. This information can guide a designer of auditory displays on, among other things, the number of distinct values they should represent with a given sound dimension.

The second purpose of the current experiments was to test whether redundant dimensions would be better than a single dimension in a simple task in which observers are to map sounds to absolute values. The development of a good representation of data with auditory displays requires that one choose not only the type of transformation between the visual and auditory information, but also the auditory dimension to use. In general, studies in sonification have used pitch and timbre combined with some temporal aspects of the sounds (usually tempo) to represent data. Usually only one dimension is used to represent the relevant information. However, there is good reason to believe that using two sound dimensions together in a redundant way could enhance performance compared to using the dimensions individually [7]. Finally, we were interested in whether people with more musical training perform better that those with less musical training, and whether the same results would be found for different representations.

2. EXPERIMENT 1

2.1. Method

A between-subjects design consisting of three methods of representing values was used. Values were represented either by pitch, by temporal intervals, or the redundant combination of the two. Forty-eight undergraduate students with normal hearing participated in the experiment for course credit.
http://psych.rice.edu/sonification/ICAD2003_AS/experiment1/

2.1.1. *Task and Procedure*

On each trial subjects were presented with a sequence of three tones representing a value from 1 to 50 and asked to indicate the value by clicking on a visually-presented axis. In the pitch condition, the middle tone represented the value with the first and last tones representing the end points of the scale. A unit was represented by the pitch difference between two consecutive musical tones from the equal tempered musical scale. The frequencies of the sounds were c2 (65.41 Hz) to d#6 (1244.51 Hz) using a sampled grand piano sound. The duration of the sequences was approximately 3.5 seconds, and the length of each tone was 0.25 seconds with a pause of 0.25 seconds between the tones.

In the temporal condition, the tones were played with the same pitch (c4 or 261.63 Hz) and the values were represented by the temporal difference between two tones. Higher values were played with a longer pause after the minimum value than lower values. Each unit on the scale was represented by 0.0625 seconds. The individual tones had the same duration of 0.25 seconds as in the pitch condition.

In the redundant condition, the values were represented both by the pitch and the timing of the sounds.

Subjects received instructions on a web page. There was a total of 50 sounds presented twice in random order so that each subject performed 100 trials. For each trial, subjects listened to the target sound and their task was to indicate the value that the second tone represented on the axis by clicking on one tick mark. After the choice was made, a feedback axis was displayed for 0.5 seconds under the original axis with the correct choice marked on it. Before performing the task, subjects filled out a form asking them how much musical training they had had.

The experiment was run on Apple Macintosh computers in the Sonification Lab at Rice University, and the sounds were presented trough headphones.

2.2. Results

For each subject the first 10 trials were considered training and were excluded from the analysis. Subjects for whom the correlation between the sonified values and their choices was more than three standard deviations below or above the overall mean (mean and standard deviation calculated without these values) were considered outliers. Five subjects were removed from the analysis: one from the frequency condition, one from the redundant condition, and three from the temporal condition.

2.2.1. Correlations

Mean correlations between the sonified values and subjects' choices are shown in Table 1. Because the sampling distribution of Pearson's correlation is not normal, the correlations were transformed to Fisher's z' scores for use in significance tests. An analysis of variance revealed a significant effect of condition, $F(2,40) = 14.78$, $p < 0.001$.

Table 1. *Summary for the three groups*

Condition	N	Mean Fisher's z'	Mean correlation (SD)
Pitch	15	1.51	0.89 (0.05)
Redundant	15	1.70	0.92 (0.03)
Temporal	13	2.12	0.96 (0.02)

The Tukey's HSD test found the mean correlation for the temporal condition to be significantly higher than both pitch ($p < 0.001$) and the redundant conditions ($p < 0.001$). However, the difference between the redundant and pitch conditions was not significant ($p = 0.20$).

2.2.2. Absolute Differences

The mean absolute value of the differences between the choices and the sonified values was computed for each subject for each condition (see Table 2). Consistent with the findings using correlations as the dependent variable, the pitch condition had the highest mean absolute difference and the temporal condition the lowest (*note*. higher values reflect worse performance than lower values).

Table 2. *Summary for the absolute differences by condition*

Condition	N	Median	Mean (SD)
Pitch	15	4.66	4.83 (1.05)
Redundant	15	3.94	3.97 (1.03)
Temporal	13	2.51	2.82 (0.88)

There was a significant effect of condition, $F(2,40) = 14.13$, $p < 0.001$. Again, Tukey's HSD test found a significant difference between the temporal and both pitch ($p < 0.001$) and redundant conditions ($p = 0.011$), but not between the pitch and redundant conditions ($p = 0.059$).

2.2.3. Accuracy

To explore subjects' accuracy in more detail, the proportion of trials having a mean absolute difference lower than a specific accuracy criterion were calculated for each condition. The criteria ranged from 0 to 9 (absolute difference of zero, absolute difference less or equal with one and so on). When the strictest criterion of zero was used the subject's response had to be exactly correct. Using this criterion, the proportions correct in the pitch, redundant, and temporal conditions were 0.09, 0.14 and 0.18 respectively. The proportion correct increased more rapidly for the temporal condition than for the two other conditions. Using a criterion for correctness of 6, the proportions correct in the pitch, redundant, and temporal conditions were 0.74, 0.80, and 0.91 respectively.

Table 3 shows the level of the correctness criterion required in order to achieve various proportions correct.

Table 3. *Accuracy for specific proportions for the three conditions*

	Mappable Difference		
Proportion	Pitch	Redundant	Temporal
0.1	0.0	0.0	0.0
0.2	0.8	0.3	0.1
0.3	1.5	0.8	0.5
0.4	2.2	1.5	0.9
0.5	3.0	2.2	1.4
0.6	4.0	3.1	2.0
0.7	5.5	4.3	2.9
0.8	7.0	6.0	4.2
0.9	10.0	9.0	5.8

For example, in the temporal condition, one must use a criterion of at least 2.0 to be correct at least 0.6 of the time. To be correct at least 0.9 of the time one must relax the criterion to 5.8. We call the correctness criterion needed to achieve a certain level of accuracy the "Mappable Difference" (MD). There is a direct relationship between MD and the number of values that can be distinguished accurately. Recalling that we used a 50-point scale, the number of values that can be distinguished accurately is equal to 50/(2MD). Thus, since the MD for 0.9 accuracy is 5.8, 4.3 distinct values can be mapped with an accuracy of 0.9 with the temporal mapping used in this experiment. However, if we use representation by pitch, with the same level of accuracy we could represent only 2.5 values.

3. EXPERIMENT 2

Experiment 1 found that subjects performed better in the temporal condition than in the redundant condition. Experiment 2 was designed to investigate in more detail the basis of responses in the redundant condition. Among the many possibilities are (a) subjects base their responses on only one dimension, with some subjects using one dimension and other subjects using the other dimension, (b) subjects base their response on one dimension, but this dimension changes from trial to trial, and (c) subjects base their response on information from both dimensions.

3.1. Method

Seventy-five subjects with normal hearing participated in the experiment either for payment or for course credit.
http://psych.rice.edu/sonification/ICAD2003_AS/experiment2/

3.*1.2.* *Stimuli and Procedure*

The method of representing the values (pitch, timing and redundant) was used as a within-subjects variable. In the first part of the experiment, there was a total of 210 trials. The trials were grouped in blocks of 10, with the order of the conditions randomized based on the Latin Square design. The stimuli were the same as in the first experiment with the difference that in this experiment only 10 values of the axis were used: 3, 8, 13, 18, 23, 28, 33, 38, 43, and 48.

The second part of the experiment used sounds only from the redundant condition intermixed with “conflicting” sounds (40 conflicting sounds out of 100 trials). The subjects did not receive any feedback during any of these trials. The “conflicting” sounds used both dimensions, so they were apparently similar to the sounds in the redundant condition. However, for these sounds the pitch and the timing coded different values of the axis. For example, if pitch coded the value 13 then timing represented the value 23 on the axis, thus conveying conflicting information.

Four sounds were designed based on four values 13 and 23, 18 and 28: pitch coding 13 - timing 23 (P13-T23), pitch 23 - timing 13 (P23-T13), pitch 18 - timing 28 (P18-T28), and pitch 28 - timing 18 (P28-T18). Similarly to Experiment 1, the subjects listened to the target sound by clicking a button on the screen. Their task was to indicate the value that the sound represented on a visually presented axis of 51 units.

3.2. Results

The order of the sound conditions had no effect, so the data were collapsed across this variable. The first 30 trials (ten for each condition) and data from two subjects were removed from the analysis based on the same criterion as in Experiment 1.

3.2.1. Part 1. Absolute Differences

The absolute values of the differences between the choices and the original values were computed for each subject and the descriptive statistics for these differences are shown in Table 4.

Table 4. *Mean absolute differences for the three conditions, standard errors in parenthesis*

Group of trials	N	M (SE)
Pitch	73	5.22 (0.19)
Redundant	73	3.66 (0.13)
Temporal	73	3.32 (0.13)

A within-subjects analysis of variance showed a significant effect of condition, $F(2,140) = 66.18$, $p < 0.001$. Consistent with the findings from Experiment 1, subjects were more accurate on the trials in the temporal condition than in either the pitch condition, $F(1,70) = 121.40$, $p < 0.001$, or the redundant condition, $F(1,70) = 10.93$, $p = 0.0015$.

3.2.2. Part 2. Conflicting Trials

The mean choice for the conflicting trials P13-T23 and P23-T13 were close to 18, the midpoint of the two values coded by pitch and timing. Similarly the mean for P18-T28 was 24.40, which is close to being halfway between 18 and 28. However, for P28-T18 the mean choice of 26.69 was much closer to the value coded by pitch than to the value coded by timing. The finding of choices intermediate between the two values indicates that both dimensions were used.

To investigate whether subjects were influenced by both dimensions on the same trial, we re-coded the choices based on how close they were to the value conveyed by pitch, timing or to the middle. That is, for sound P13-T23 if a subject chose a value less than 15, the subject was judged to have used pitch for that specific trial; if the choice was between 15 and 20 was judged to have used both dimensions; and if it was larger then 20, the subject was judged to have used timing in the task. The dependent variable calculated for analysis was the proportion of trials on which a subject was judged to have used both dimensions. For comparison purposes, this variable was also computed for the single dimension trials by pooling the data from the Pitch 13 and Timing 23 responses. The same variable was computed for other stimulus pairs in like manner. Higher scores on this variable on the conflict trials than on the combined data from the single dimension trials are indicative of using a different strategy on the conflicting trials, that is, using both dimensions rather than only one on at least some of the trials.

Although the mean proportion of using both dimensions was generally higher on the conflicting trials than on the pooled trials, the difference was significant only for the sound P23-T13, $t(35) = 3.59$, $p < 0.001$). For the other three conflicting sounds subjects used predominantly one of the dimensions. This is also revealed by the flatness and slight bimodality of the distribution of choices.

3.2.3. *Musical Training*

Based on sample size considerations, the subjects were categorized into three groups: (1) those with no musical training, N = 32; (2) those with one to three years of musical training, N = 20, and (3) those with more than three years of musical training, N = 21. There was a significant effect of musical training, $F(2,70) = 8.95$, $p < 0.001$. The Tukey's HSD test (using the 0.01 level) revealed that the group with more than three years of musical training did significantly better on the task than the other two groups for all three conditions with the exception of the temporal condition for which the difference was significant only between the group with more than three years of musical training and those without any musical training.

4. DISCUSSION

We investigated the limits of representing absolute values with sounds. The levels of accuracy as determined by Experiment 1 can give guidelines depending on the details to be converted to auditory information. The approach taken in identifying the mappable difference in pitch, timing, and the redundant representation for one numerical value could be applied to more complex auditory displays as well. Knowing the average accuracy of identifying absolute values based on sounds, sound displays could be designed with a given accuracy range for a large population, and as a result information could be presented with a lower error rate.

The task used in these experiments is simple compared to real sonification applications and involves mapping sounds to visual displays. Naturally, further research is needed to test the generalizability of these findings. Moreover, because reference sounds were always provided, subjects could have been judging intervals rather than absolute values.

In both experiments subjects were more accurate on the temporal condition than either the redundant or pitch conditions. The redundant representation was worse than the temporal representation, but better than mapping by pitch.

The results from the conflicting trials in Experiment 2 suggest that subjects use mainly one dimension to indicate the value represented by the sound, but the other dimension influences them as well. This suggests that the pitch and temporal aspects of sounds are separable dimensions [8]. It is plausible that if the two dimensions used in a redundant way were Garner integral the redundant usage of the two dimensions might lead to redundancy gain not only in a classic Garner sorting task but in a mapping task as well. Two dimensions are Garner integral if selective attention is not possible, and there is facilitation in correlated-dimension tasks; whereas two dimensions are Garner separable if selective attention is possible, and there is no facilitation in correlated dimension tasks [9]. There is no research yet investigating the effects of Garner integrality in sonification. Therefore, in future experiments we plan to look at how two dimensions, namely pitch and duration, relate to each other in a classic Garner sorting task. In addition, we will also examine the relation between these two dimensions using the same paradigm as in the presented experiments, that is, representing numerical values with pitch and duration as single dimensions and also in a redundant fashion.

In Experiment 2 we found that subjects who had more than three years of musical training did better on the pitch and redundant trials than those with three or less years of musical training. The relation between musical training and temporal representation was weaker than for the other two conditions suggesting that this representation might be more suitable for wider populations with and without musical training considering also that it led to the highest performance among all three conditions.

ACKNOWLEDGEMENTS

We thank Gregory Kramer for his comments on an earlier version of this paper. This research was supported by NSF Grant IIS-9906818.

5. REFERENCES

[1] G. Kramer, "An Introduction to Auditory Displays," in *Auditory display: Sonification, audification and auditory interfaces. Proceedings of the First International Conference on Auditory Displays (ICAD) 1992*, G. Kramer, Ed. Reading, MA: Addison-Wesley, 1994, pp. 1-77.

[2] "Sonification Report: Status of the Field and Research Agenda," 1997.

[3] 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.

[4] J. H. Flowers, L. E. Whitwer, D. C. Grafel, and C. A. Kotan, "Sonification Of Daily Weather Records: Issues Of Perception, Attention And Memory In Design Choices," presented at International Conference on Auditory Display, Helsinki, Finland, 2001.

[5] T. L. Bonebright, M. A. Nees, T. T. Connerley, and G. R. McCain, "Testing the Effectiveness of Sonified Graphs for Education: A Programmatic Research Project," presented at Proceedings of the 7th International Conference on Auditory Display, Helsinki, Finland, 2001.

[6] S. C. Peres and D. M. Lane, "Sonification of Statistical Graphs," presented at International Conference on Auditory Display, Boston, MA, 2003.

[7] 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 Display, Palo Alto, CA, USA, 1996.

[8] C. Palmer and C. L. Krumhansl, "Independent Temporal and Pitch Structures in Determination of Musical Phrases," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, pp. 116-126, 1987.

[9] J. R. Pomerantz, "Global and Local Precedence: Selective Attention in Form and Motion Perception," *Journal of Experimental Psychology: General*, vol. 112, pp. 516-540, 1983.