TICK-MARKS, AXES, AND LABELS: THE EFFECTS OF ADDING CONTEXT TO AUDITORY GRAPHS

Daniel R. Smith

Sonification Lab, School of Psychology
Georgia Institute of Technology,
274 5th Street
Atlanta, GA 30332-0170
daniel.robert.smith@us.army.mil

Bruce N. Walker

Sonification Lab, School of Psychology
Georgia Institute of Technology,
274 5th Street
Atlanta, GA 30332-0170
bruce.walker@psych.gatech.edu

ABSTRACT

As the use of sonification expands, researchers and designers continue to employ techniques for adding context (such as tick marks, axes, or labels) whose benefit remains unquantified. This study examined the effect of several such techniques on the perceivability of an auditory graph.

In Block 1, participants listened to a simple auditory graph, which had no added context (such as tick marks, axes, and labels), and answered trend analysis and point estimation questions about the information presented by the graph. In Block 2, participants repeated the process but the graph was augmented by 1 of 6 types of added context.

The data revealed differences in perceivability between conditions for both trend analysis and the point estimation task, and an explainable ordering of error levels based on the amount and type of information provided by a particular contextual setting.

1. INTRODUCTION

1.1. Sonification

The most common tools and techniques used in data analysis and information display are almost entirely visual. This can be problematic for a number of reasons [1]. Among them is the realization that while system complexity and user information requirements are on the rise, requirements for smaller, more mobile devices continue to limit and further shrink already overcrowded, single modality, vision-based displays.

Sonification, it is hoped, will provide alternatives and flexibility for future information displays, but researchers have only begun to develop a base of empirically grounded knowledge for use in design and implementation of such displays. Walker has pointed out the need for appropriate mappings, polarities, and scalings [1]. In addition to these requirements, it is becoming increasingly accepted among researchers and designers that the addition of context is beneficial to the comprehension of sonifications [2-4]. That is, the addition of the auditory equivalent of axes, tick marks, labels, and so on, should lead to improved performance. But little has been done to examine how much benefit is actually realized by the addition of a given contextual cue, relative to another.

1.2. Perceivability

Sonifications are somewhat analogous to visual graphs in that they are an alternative means of communicating, or otherwise helping a perceiver comprehend, or "visualize" quantities and variations within a given data set [1].

Although they are analogous to visual graphs, common terminology used in visual display does not translate well to sonification, especially where one is interested in the ability of a user to accurately extract information from a display. We might term it "usability", but a display could be usable and still not be designed in a way that facilitates the *accurate* extraction of information. "Readability" seems to fill the requirement for a visual display, but one doesn't exactly read a sonification. Therefore, in the absence of commonly accepted terminology in the field, we have settled on "perceivability". In the context of this study, *perceivability* is defined as the ability of a user to accurately extract information from a given sonification.

Despite such differences, efforts to improve perceivability of sonifications continue to attempt to capitalize on existing parallels between auditory and visual displays. One such parallel is the use of context.

1.3. Context

Intentional context refers to the purposeful addition of information to a display. In visual information display, additional, useful information (such as axes and tick marks) increases readability and aids perception by enabling more effective top-down processing [5, 6]. If one is able to view labeled tick marks along the axis of a display, one is better able to judge the magnitudes and data dimensions represented in the graph or chart [6].

Consider then, the effect on readability if the display were devoid of all intentional contexts. In visual line graphs, for example, the line itself provides some context, but only the incidental context inherent in the observation that some data points are further to the right, or above or below the data points to the left [7]. This inherent context might enable an observer to execute rudimentary trend analysis of the data (eg. Is the line generally rising or falling?), but the accurate extraction of a specific value is impossible. Unfortunately, many auditory graphs, even if making use of optimal mappings, polarities, and scalings, approximate this impoverished amount and type of context [1].

It is important to note that there is a limit to how much intentional context should be added to a display. There is an optimal amount of context. After that, the addition of context fails to provide useful additional information. Instead it interferes with, clutters, or distracts from the extraction of more useful information [8]. In addition, the inclusion of a contextual cue might provide information that is useful in one task (such as trend analysis), but serve as no more than clutter for another task (such as point estimation).

Thus far little has been done to explore how the principles of context apply to sonification. As a result, lacking empirically supported design principles and guidelines, the employment of context—and thus the performance of sonifications—have been inconsistent [9].

Assuming findings show that the addition of intentional context improves perceivability, there are countless types of context we might add (auditory equivalents to x-axis context, y-axis context, thresholds, labels, or other types of intentional context). The most effective types of context might be analogous to those of visual displays. On the other hand, sound might lend itself to an entirely different type of context not available in other modalities.

A common method for adding x-axis context to a sonification is the addition of a series of clicks or percussive sounds [2-4]. But, is there more than one way to add context to an x-axis? If so, it seems unlikely that the same techniques could be used to add context to the y-axis. In addition, it remains to be seen how one most effectively sonifies a label.

The purpose of the present study was not to advocate a specific design technique or the addition of a specific set of contextual cues. Rather, this was an initial attempt to quantify the benefit (if any) of adding some contextual cues, relative to and in combination with other cues. We anticipated that added intentional context would enhance perceivability so long as it introduced new and useful information, and did not interfere with, clutter, or distract from more useful information. As in a visual display, if added intentional context fails in these respects, it will cease to improve, and will ultimately degrade the perceivability of the sonification.

2. METHOD

2.1. Participants and Apparatus

Fifty-two Georigia Tech undergraduate students (42 men and 10 women) completed the point estimation task, and 63 (48 men and 15 women) completed the trend analysis task. Students participated for course credit. Participants provided demographic details about age, sex, handedness, and number of years of musical training, and all reported normal or corrected to normal vision and hearing. Instructions and visual stimuli appeared on a 17-in. Apple Macintosh G4 display. Auditory stimuli were presented via Sony MDR-7506 headphones. The experiment ran in Netscape Navigator v.4.77 on Macintosh OS 9.2.

2.2. Stimuli

Participants listened to an auditory graph representing the variation in price of a single, unidentified stock over a 10-hour trading day (from 8 am to 6 pm). In the display, price (in dollars) was mapped to frequency using the preferred positive polarity, and the preferred scaling (expressed as a logarithmic slope) for dollars to frequency for sighted listeners: .9256 [10].

Therefore, the amount of increase in frequency representing a given increase in number of dollars was modeled by the equation:

$$Y \;\square\; X^{(.9256)} \qquad (1)$$

where Y represents the number of dollars and X represents frequency in Hz. Time (over the 10-hour trading day) was mapped against time in the display, where each hour of the trading day was represented by 1 second in the display.

2.3. Design

In Block 1, all participants tested under the control condition for a total of 22 trials. Then, in Block 2, they completed an additional 22 trials under one of six experimental conditions.

2.3.1. *Variables*

The method of measurement for the dependant variable, *perceivability*, varied depending upon the task. For the point estimation questions, perceivability was operationalized as the mean absolute error (in dollars) with which the participants in a given group, under a given condition, reported the queried data value represented in the display. For the trend analysis questions, it was represented by the mean proportion of questions answered correctly by the participants in a given group, under a given condition. The manipulated independent variables included the amount and type of intentional context added.

2.3.2. *Conditions*

In Block 1, participants in all groups tested under the control condition. Under the control condition participants experienced the graph without added intentional context.

In Block 2, Group 1 experienced the graph again under the same control condition. Group 2 experienced the graph with the addition of x-axis context. This context was created by the insertion of audible clicks in the data. Each click represented the passing of one hour in the 10-hour trading day (1 click/sec). Group 3 experienced the graph with the addition of y-axis context, which was created via the addition of a beeping reference tone. This pitch of this reference tone represented the opening price of the stock. Group 4 also experienced the graph with the addition of y-axis context, but this time it was created via the addition of a dynamic beeping reference tone. When the price of the stock was rising the pitch of the beeping reference tone corresponded to the highest price of the day ($84). When the price was falling, the pitch of the beeping reference tone changed to the lowest price of the day ($10). The applicable values were made known to the participants as necessary in both Groups 3 and 4. Group 5 experienced the graph with the combination of x-axis context (clicks), with y-axis context (the opening-price reference tone from Group 3). Finally, Group 6's graph combined x-axis context (clicks) with y-axis context (the dynamic max/min-price reference tone from Group 4).

It is important to note that in every condition, participants were given the initial price of the stock at the opening of the trading day.

2.4. Procedure

Upon arrival participants were randomly assigned to one of the six groups. To start Block 1, listeners received an explanation of the nature of the display and instructions pertaining to the completion of the task. On each of the 11 trend analysis trials in Block 1, participants were asked to report whether the price of the stock was rising, staying the same, or falling at a randomly selected time of day. On each of the 11 point estimation trials, participants where asked to estimate the price of the stock at a randomly selected time of day. Participants were able to listen to the graph as many times as required to answer each question. No accuracy feedback was provided.

In Block 2, listeners received an explanation of the nature of the new display and instructions pertaining to the completion of the task. Participants again responded on each of 22 trials, but in this block, the graph contained contextual cues as described by one of the six experimental groups above.

3. RESULTS

3.1. Point Estimation

The response from each point estimation trial was compared to the correct response to yield the absolute value of the error for that trial. From these, each listener's average absolute error was calculated for each block. Then a two-way, univariate ANOVA was conducted by experimental group and block with mean absolute error as the dependent measure (see descriptive statistics in Table 1.). Three outlying observations were deleted from Block 1 and one from Block 2 because their studentized residual exceeded 1.96.

The ANOVA found a significant main effect of block, demonstrating that listeners answered with smaller absolute errors in Block 2 than they did in Block 1 ($F(1,88) = 9.226$, $p=.003$). There was a significant main effect of group ($F(5,88) = 6.566$, $p<.001$), and a marginally significant block by group interaction. ($F(5,88) = 2.156$, $p=.066$) (See Figure 1.).

A one-way ANOVA completed on only the Block 1 data showed no difference in mean absolute error between the six groups. This is as expected because in Block 1 they all test under the control condition. The same analysis conducted on Block 2 found significant differences between group means ($F(5,45) = 9.420$, $p<.001$). Individual comparisons taken between the groups using Fisher's LSD revealed that Groups 4 and 6 had significantly lower absolute errors relative to the control group. Mean differences were 5.378 (SE=1.535, $p<.001$), and 4.616 (SE=1.585, $p=.006$) respectively.

Lastly, contrasts taken between each group's mean in Block 1 and the same group's mean in Block 2 revealed that only Group 4 and 6 realized a significant decrease in error in Block 2 compared with the same group's performance in Block 1 ($F(1,15) = 28.401$, $p<.001$ and $F(1,10) = 7.779$, $p=.019$ respectively).

Table 1. *Mean Absolute Error.*

Block	Group	Mean Error ($)	Std. Deviation ($)	N
1	1	12.461	3.189	7
	2	11.718	3.588	10
	3	14.409	3.663	10
	4	11.878	2.625	9
	5	12.720	4.218	8
	6	11.636	3.449	5
	Total	12.558	3.461	49
2	1	11.798	3.277	7
	2	10.200	2.942	10
	3	14.227	3.500	10
	4	6.420	1.280	8
	5	12.958	3.622	9
	6	7.181	2.113	7
	Total	10.688	4.025	51

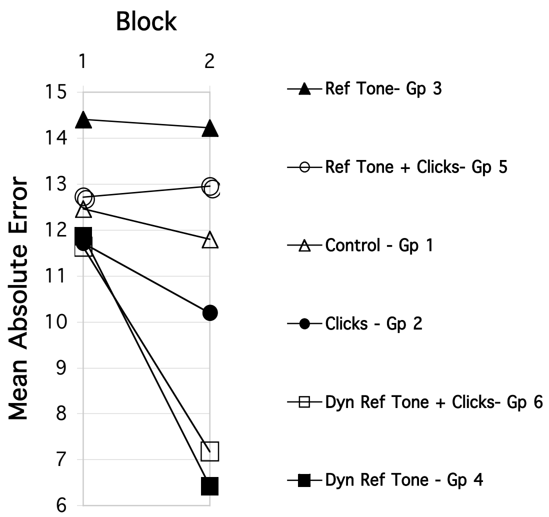

Figure 1. *Mean Absolute Error.*

3.2. Trend Analysis

The response from each trend analysis trial was compared to the correct response to yield the proportion of questions answered correctly by each participant in Block 1 as well as in Block 2. Then a two-way, univariate ANOVA was conducted by experimental group and block with proportion correct as the dependent measure (descriptive statistics at Table 2.). Four outlying observations were deleted from Block 1, and one from Block 2, because their studentized residual exceeded 1.96.

The ANOVA found a significant main effect for block demonstrating that listeners answered with a greater proportion of correct answers in Block 2 than in Block 1 ($F(1,109) = 13.780$, $p<.001$), but neither the main effect of group nor the interaction reached significance (see Figure 2.).

Still, exploratory individual contrasts were taken between each group's mean in Block 1 and the same group's mean in Block 2. This revealed that only Groups 4 and 6 realized a significantly increased average proportion correct in Block 2 compared with the same group's performance in Block 1 ($F(1,18) = 6.330$, $p=.022$ and $F(1,16) = 7.023$, $p=.017$ respectively).

Table 2. *Mean Portion Correct.*

Block	Group	Mean Portion Correct	Std. Deviation	N
1	1	.606	.213	12
	2	.707	.168	9
	3	.686	.208	9
	4	.636	.135	10
	5	.694	.183	11
	6	.590	.168	8
	Total	.653	.180	59
2	1	.681	.137	12
	2	.800	.071	10
	3	.818	.128	8
	4	.781	.122	10
	5	.704	.145	12
	6	.772	.123	10
	Total	.753	.130	62

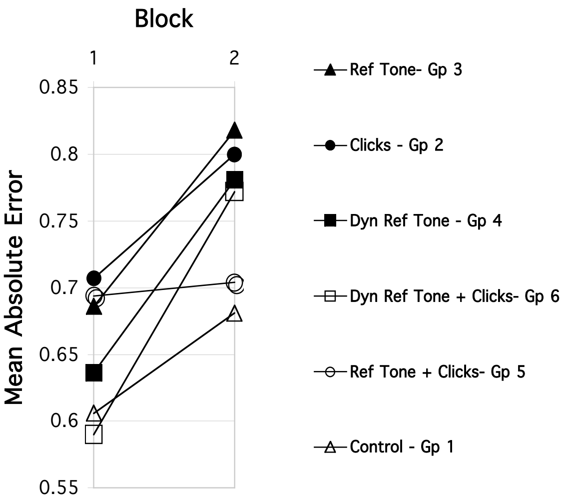

Figure 2. *Mean Portion Correct.*

4. DISCUSSION & CONCLUSIONS

4.1. Point Estimation

The analysis of the data from the point estimation task found that Groups 4 and 6 showed a significant reduction in absolute error over their performance in Block 1. In addition, Groups 4 and 6 had significantly smaller absolute errors in Block 2 than either the control or other groups. This is evidence that the y-axis context information provided by the dynamic reference tones assisted the participants with the required perceptual and cognitive tasks. Some of this reduction in error should be attributed to training, practice, and testing effects but our finding that both groups showed a significantly smaller error than the control group, is evidence that the additional reduction in absolute error was a result of the intentional context provided by the dynamic reference tones.

Accurate point estimation of queried values requires users to execute several perceptual and cognitive tasks, including the simultaneous execution of an interval division task and a magnitude estimation task. First, the listener must estimate what point in the duration of the graph corresponds to the queried time of day. Next, the listener must estimate the price of the stock, as represented by the difference in the pitch perceived during that part of the display, relative to the pitch perceived at the onset of the display.

Under the control condition, the opening price of the stock is the only added context, but it is not the only context present. Also present is the incidental context provided by the sound itself: the duration of the sound, and the variation of the frequency. When given the opening price and asked to report the price at a given time (noon, for example); the subject must listen to the entire graph, recall the pitch he or she perceived at approximately half the duration (noon time), compare it to the pitch perceived at the very onset of the graph (the opening of trading), estimate the change in price represented by the difference between the noon-time pitch relative to the opening pitch, and add or subtract that change in price to the opening price of the stock.

One might assume–perhaps with good reason—that this set of tasks, unassisted by context, will be extremely difficult and result in a relatively low level of accuracy. The addition of useful information (context) could assist users with these tasks and may result in substantially improved perceivability of the graph.

For example, the addition of y-axis context could assist the listener in the magnitude estimation task by helping to judge the magnitude and direction of change of the price of the stock relative to either the opening price of the day or to the maximum and minimum prices of the day. If the user is given a reference tone representing the opening price, the listener is still required to recall the pitch he perceived at approximately half the duration of the sonification, but he is now able to judge the noon-time pitch, relative to a sound he is hearing right now instead of against a sound which, under the control condition, he was forced to maintain in working memory. If the user is provided with a dynamic reference tone that changes to represent either the maximum or minimum prices of the day, not only is the user relieved of the working memory task, but the reference tones serve as reinforcement of the preferred scaling.

The addition of x-axis context (clicks) should eliminate the interval division task, and assist the listener in the magnitude estimation task. To return to the above example, the listener will no longer have to estimate what part of the graph represents noon. Knowing the trading day starts at 8 am, the listener is free to focus attention on perceiving the pitch in immediate temporal proximity of the fourth click. This should also provide some assistance in the magnitude estimation task. Since the listener is no longer required to listen to the entire graph, she is no longer required to recall the pitch perceived at half the duration of the graph. Upon perception of the fourth click, the listener is immediately free to begin a comparison of the pitch perceived at that moment, to the pitch perceived at the onset of the sonification. This should make it easier to estimate the price represented by the noon sound, relative to that of the opening sound. If a listener is provided with the addition of both types of context in combination, it should eliminate the interval division task, and also provide the several forms of assistance in the magnitude estimation task as described above.

For these reasons, a particularly interesting, and somewhat surprising finding was that the addition of x-axis context (clicks) did not produce a significant reduction of error. We hypothesized that x-axis context would be valuable because the added context updates the listener's scaling from 1 hour = 1 hour, to 1 hour = 1 second. Considering the difficulty of the interval division task this could be of significant assistance to the listener. Although one might conclude that the addition of the clicks did not provide useful information, it is also possible (in this case) that the addition of the clicks did not provide *additional* information. To be specific, in our auditory graph, changes in price were represented by sounds whose frequency varied discretely on the hour and half hour. Given the knowledge that it was a ten hour trading day, it is likely that participants counted 20 tones, realized that there were two tones per hour and thus, using the discrete variation of the frequencies in the data itself, they had already extracted the information that the clicks were meant to provide.

Lastly, the findings showed that the addition of the single reference tone, whose pitch represented the opening price of the stock, did not result in a significant reduction in error. The reference tone may have helped with certain cognitive tasks (as described above), but it may not have provided enough useful, additional information to make a significant impact on perceiveability.

Although unobserved in this study, we still hypothesize that the continued addition of context could possibly interfere with, clutter, or distract from either the initially added type of context, or from the data itself. If this is the case, mean error would increase from block 1 to Block 2.

Finally, it is important to note that the efficacy of any decrease in error can only be evaluated relative to the magnitudes and variations in the data being displayed. For example, Group 4's decrease in mean absolute error (from \$11.87 to \$6.42) is nearly a 50% decrease in error. Clearly one could argue that it improved performance, but was that improvement practically relevant? Did it make it a good graph? Did it even make it a better graph?

The relative "goodness" of such a level of performance could depend on any number of factors that would be evaluated in proper user and task analyses. In our study, the opening price of the stock was \$50, it fluctuated during the day between a maximum of \$84 and a minimum of \$10, with the average price of the day being \$50. Under the control condition (no context), the average absolute error for Group 4 was about 20% of the average price of the stock. With the addition of the dynamic reference tones, the average absolute error dropped to about 12% of the price of the stock. Replication and applied study is required to evaluate the usability of such a graph. This approach is essential to continued progress in the field, and should be applied to each innovation and new idea, as well as to evaluate the usability of the resulting sonifications.

4.2. Trend Analysis

The analysis of the data from the trend analysis task found that only Groups 4 and 6 showed a significant increase in proportion correct over their performance in Block 1. However Groups 4 and 6 did not show significantly greater proportions correct in Block 2 relative to the control or other groups. Given these results, we tentatively conclude that the y-axis context provided by the dynamic reference tones assisted the participants with the required perceptual and cognitive tasks. But it is also possible, given our results, that the improvement resulted from training, practice, and testing effects.

Accurate identifications of trends in price at particular times of day require participants to execute a reduced number of the same perceptual and cognitive tasks required for point estimation. Under the control condition, when asked to describe the trend at a given time (noon, for example): the subject must listen to the entire graph and recall the *variations* in pitch perceived at approximately half the duration (noon time).

One would assume that this set of tasks should be easier than the set required for point estimation, but we hypothesized that intentional context should still assist users and result in improved perceivability. In addition, given that the task demands are different, it is likely that certain contexts will affect perceivability differently for trend analysis tasks than they do for point estimation.

For example, the addition of y-axis context may no longer assist the listener. If the user is given a single reference tone representing the opening price, they are provided with little or no information pertinent to the task. The reference tone, which may have been useful in the point estimation task, may now only be clutter resulting in more errors. But if the user is provided with a dynamic reference tone that changes to represent either the maximum or minimum prices of the day dependent upon whether the price is rising or falling, then the trend analysis task is essentially automated. The distinction between two tones should be less perceptually difficult than the analysis of the pitch variations over a queried time interval. Thus, the dynamic reference tone should assist the user and increase perceivability.

The addition of x-axis context should eliminate the interval division task relieving the user of the requirements to listen to the entire graph and to *recall* the variations in pitch proximal to the queried time of day. This should make it easier to identify trends. If a listener is provided with the addition of both types of context in combination, the clicks should help and the reference tone should either degrade or assist as described above.

As in the point estimation task, our finding that the clicks did not result in improved perceivability was unexpected but explainable, given the discrete variation of the sounds representing price.

4.3. Conclusions

The results of the current study support the theoretical position that the addition of useful information enhances the perceivability of auditory graphs. If future results are in accordance with the alternative outcome, our conclusion may be qualified by evidence to showing that, at some point, adding more "chart junk" fails to provide new, useful information or it interferes with, clutters, or distracts from more useful information [5].

Again, this study has not attempted to test whether these particular contextual cues are "the best". Rather, cues were chosen that seem to be used more and more often in sonification, and the performance benefit they bring to two sonification tasks has been assessed. Hopefully, these findings and further research will help us understand how to design and employ sonifications in ways that enhance system performance, give users more flexibility, and improve system interfaces wherever sonification is appropriate. Ongoing research should continue to push this frontier by: investigating the relative benefits of different techniques for the creation of context; defining the point where adding more "chart junk" ceases to produce practically relevant improvements in performance; examining the effects of training on perceivability; and otherwise exploring alternatives for improved design and implementation of sonifications.

References

1. Walker, B.N., *Magnitude estimation of conceptual data dimensions for use in sonification.* Journal of Experimental Psychology: Applied, in press.
2. Bonebright, T.L., et al. *Testing the effectiveness of sonified graphs for education: A programmatic research project*. in *The Proceedings of the International Conference on Auditory Display*. 2001. Espoo, Finland.
3. Flowers, J.H., D.C. Buhman, and K.D. Turnage, *Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples.* Human Factors, 1997. 39(3)(September): p. 341-351.
4. Gardner, J.A., R. Lundquist, and S. Sahyun. *TRIANGLE: A practical application of non-speech audio for imparting information.* in *The Proceedings of the International Conference on Auditory Display*. 1996. San Francisco, CA.
5. Tufte, E.R., *Envisioning information*. 1990, Cheshire, Connecticut: Graphics Press.
6. Bertin, J., *Semiology of graphics*. 1983, Madison, Wisconsin: The University of Wisconsin Press.
7. Walker, B.N., *Auditory display and earcons*. 1994, Rice University: Houston, TX.
8. Tufte, E.R., *The visual display of quantitative information*. 2nd ed. 2001, Cheshire, Conn.: Graphics Press. 197.
9. Edworthy, J., S. Loxley, and I. Dennis, *Improving auditory warning design: Relationship between warning sound parameters and percieved urgency.* Human Factors, 1991. 33(2): p. 205-231.
10. Walker, B.N. and D.M. Lane. *Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners*. in *The Proceedings of the International Conference on Auditory Display*. 2001. Espoo, Finland.