

Effects of Auditory Context Cues and Training on Performance of a Point Estimation Sonification Task

DANIEL R. SMITH and BRUCE N. WALKER*

Georgia Institute of Technology, USA

SUMMARY

Research on auditory graph interpretation has investigated mappings, scalings, and polarities, as well as the addition of some contextual design features like clicks. However, little has been done to quantify the benefits of those or other design features, or to investigate training in specific sonification tasks such as point estimation. In Study 1, 160 undergraduates listened to auditory graphs and estimated exact dollar values at specific points in the graph. X-axis context (clicks) improved performance by aiding time calibration. Y-axis context (reference tones) that provided scaling cues improved performance. In Study 2, listeners performed similar tasks with continuous as opposed to discrete stimuli, and with or without training. X-axis clicks did not help in this case. Training improved performance overall. When there was no training, y-axis context improved performance to the level of trained listeners, suggesting the effects of training and y-axis were comparable, but not additive. Copyright © 2005 John Wiley & Sons, Ltd.

Displays of quantitative information are often an essential part of human machine systems. In addition, they find utility in numerous circumstances such as education, management, and the analysis and interpretation of scientific results (Johannsen, 2002). Currently, the most common tools and techniques for such purposes are almost entirely visual. This can be problematic for a number of reasons. First, visual displays fail to exploit the superior ability of the human auditory system to recognize temporal changes and patterns (Bregman, 1990; Flowers & Hauer, 1995; Flowers, Buhman, & Turnage, 1997; Kramer et al., 1999; McAdams & Bigand, 1993; Moore, 1997). Second, in practical work environments the operator is often unable to *look* at, or unable to *see*, a visual display. The visual system might be busy with another task (Fitch & Kramer, 1994; Wickens & Liu, 1988), or the perceiver might be visually impaired, either physically or as a result of environmental factors such as smoke or line of sight (Walker & Lindsay, in press; Fitch & Kramer, 1994; Kramer et al., 1999; Walker, 2002; Wickens, Gordon, & Liu, 1998). Third, auditory and voice modalities have been shown to be most compatible when systems require the processing or input of verbal-categorical information (Salvendy, 1997; Wickens & Liu, 1988; Wickens, Sandry, & Vidulich, 1983). Other features of auditory perception that suggest sound as an effective data representation technique include our ability to monitor and process multiple auditory data sets (parallel listening) (Fitch & Kramer, 1994), and our ability for rapid auditory detection, especially in high stress

*Correspondence to: Bruce N. Walker, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170, USA. E-mail: bruce.walker@psych.gatech.edu

environments (Kramer et al., 1999; Moore, 1997). In light of these issues, it is important to understand, both from a practical and a theoretical perspective, how to design auditory displays for optimal comprehension and utility. We present a theory-based task decomposition of the auditory display interpretation task, and use that to motivate methods for improving performance.

Sonification uses non-speech audio to convey quantitative information (Kramer et al., 1999; Walker & Lane, 2001). It is characterized by the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation (Kramer et al., 1999). As the use of sonification expands, there is an increasing need to understand how to enhance performance to levels that enable its use in more complex systems and displays (Johannsen, 2002; Marila, 2002; Neuhoff & Wayand, 2002; Zotkin, Duraiswami, & Davis, 2002). However, relatively little work has been directed at systematic evaluation of specific design features alone or in conjunction with training. An *auditory graph* is a specific class of sonification that includes design features intended to provide context and assist the user in a more accurate perception of the data relations. That is, additional sounds are included to serve as auditory equivalents of the axes, tick marks, and labels found in traditional (visual) graphs. Such auditory context cues should lead to more accurate perceptions of the data being displayed (Walker, 2002). Unfortunately, given the lack of empirically supported design principles and guidelines, the employment, design, and performance of auditory graphs (and indeed all sonifications) have been inconsistent (Edworthy, Loxley, & Dennis, 1991; Kramer et al., 1999). Exactly how best to design and implement auditory displays of quantitative information is still very much an open question. It also remains to be understood what perceptual or cognitive mechanisms underlie any approaches that are successful. In the present paper we discuss one design approach (adding context cues) and one behavioural approach (training the listeners) to improve the match between designer intentions and listener interpretations of auditory graphs.

MAPPINGS, SCALINGS, AND POLARITIES IN SONIFICATION

In sonification (and auditory graphs) the available display dimensions are sound parameters such as frequency (pitch), amplitude (loudness), timbre, and tempo, among others (Carlile, 2002; Walker & Kramer, 1996). An important issue is the best *mapping* of data values to the sound dimensions. Walker (2002) has pointed out that different sound dimensions are better for representing certain data types. *Polarity* refers to how the data dimension and the display dimension co-vary. If a data dimension (e.g. temperature) increases, a positive polarity dictates an increase in the assigned display dimension (e.g. increasing pitch). A negative polarity would dictate a decrease in the assigned display dimension. *Scaling* refers to how much change in a data dimension is represented by a given change in the display dimension. Scaling is often expressed as the slope of the magnitude estimation plot relating changes in a physical display dimension to perceptions of a data dimension. It is important to note that since it involves a mixture of perceptual and cognitive processing, the 'true' or 'best' scaling value for representing data with sound can depend on the exact type of data and display dimensions in use (e.g. B. N. Walker, submitted; Consistency of magnitude estimations with conceptual data dimensions used for sonification; Edworthy et al., 1991; Edworthy, Hellier, & Hards, 1995; Hellier, Edworthy, Weedon, Walters, & Adams, 2002; Walker, 2002). This means that there will be different scaling factors for, say, dollars, temperature, and urgency, when mapping onto pitch. The stability of these scaling factors is also somewhat questionable

(see e.g. B. N. Walker, submitted; Consistency of magnitude estimations with conceptual data dimensions used for sonification; Hellier, Wright, Edworthy, & Newstead, 2000), and they must be used with the understanding that the listener may not immediately have in mind the exact scaling factor that the display designer used in creating the display. For a more detailed discussion of these issues see B. N. Walker (submitted; Consistency of magnitude estimations with conceptual data dimensions used for sonification). The summary message is that the use of the most preferred parameters should, overall, lead to better performance with a sonification or auditory display (Neuhoff & Wayand, 2002; Walker, 2002), however there remains a need to help the listener adjust her or his mental representations or expectations in order to more closely match the data-to-display relations intended by the display designer.

CONTEXT AND TRAINING IN SONIFICATION TASKS

Context refers to the purposeful addition of non-signal information to a display. In visual displays, additional information such as axes and tick marks can increase readability and aid perception by enabling more effective top-down processing (Bertin, 1983; Tufte, 1990). If one is able to view labelled tick marks along the axis of a graph, for example, one is better able to judge the data dimensions and magnitudes being displayed (Bertin, 1983). A visual graph without context cues (e.g. no axes) provides no way to estimate the value at any point. The contour of the line provides some incidental context, which might allow an observer to perform a *trend analysis* (rising versus falling), but the accurate extraction of a specific value (i.e. a *point estimation* task) is impossible.

Unfortunately, many auditory graphs, even if they attempt to make use of optimal mappings, polarities, and scalings, still employ this impoverished amount and type of context (e.g. Brown & Brewster, 2003; Flowers et al., 1997; Flowers & Grafel, 2002; Flowers & Hauer, 1995; Neuhoff, Kramer, & Wayand, 2002; Walker, 2002). Commonly used 'tone graphs', in which the pitch of a tone simply changes with the data, do not have the sophistication required for point estimation (though they may still be suitable for trend analysis, Flowers & Grafel, 2002). In the few cases where some kind of context cues have been added to auditory graphs, it has typically been through the addition of timing cues, in the form of a series of clicks playing along with the data like a metronome (Bonebright, Nees, Connerley, & McCain, 2001; Flowers et al., 1997; Gardner, Lundquist, & Sahyun, 1996). This approach has not been studied for its effectiveness, nor has there been a theoretical explanation presented for how it should help a listener interpret an auditory display. Further, it has not been discussed how, if at all, the timing information imparted by clicks would benefit the point estimation task, since it does not provide any scaling cues such as those provided by the y-axis in a visual graph. Other approaches to adding scaling cues need to be identified through a task decomposition, and their utility investigated.

In addition to understanding how to develop a display that helps the perceptual and cognitive tasks of the listener, it should be possible to improve performance in graph interpretation tasks via training, given the information processing demands and task characteristics for completing a point estimation task. Consider a simple tone graph of stock market data where price is mapped to pitch such that an increase in pitch represents a corresponding increase in price (analogous to changes along the y-axis). Time of day in the stock market is mapped to time passage in the auditory graph. As discussed, such an impoverished display is actually representative of many auditory graphs in the literature, including some very recent studies (e.g. Brown & Brewster, 2003; Neuhoff et al., 2002).

Even some recent auditory graph design recommendations have surprisingly made no mention of auditory context (e.g. Brown, Brewster, Ramloll, Burton, & Riedel, 2003). In any case, determining the value of a specific data point requires the listener to execute several perceptual, cognitive, and working memory tasks. Listeners must: (1) listen to the entire graph; (2) perform an interval division task to determine the part of the sound duration corresponding to the queried data point; (3) recall the pitches perceived both at the queried time and at the onset of the graph; (4) compare one pitch to the other and estimate the change in price represented by the difference (a magnitude estimation task, Walker, 2002); and lastly, (5) recall the value of the initial data point, add or subtract the perceived change in price, and report the value. This task analysis corresponds to the established Mixed Arithmetic-Perceptual model of comprehension in visual graphs (Gillan & Lewis, 1994). Clearly, training should be able to assist the listener in learning to accomplish this complex procedure. There is considerable research indicating that performance on perceptual tasks is susceptible to improvement with training (e.g. Ahissar, 1999, 2001; Aiken & Lau, 1967; Annett, 1966; Annett & Patterson, 1967; Cuddy, 1968; Lunn, 1948; Prather, Berry, & Bermudez, 1972; Proctor & Dutta, 1995; Sidley, Winograd, & Bedarf, 1965; Swets, Millman, Fletcher, & Green, 1962), including perceptual learning specific to audition (Cuddy, 1968). Cuddy's results indicated not only that training could alter the performance of untrained subjects, but also that the focus of the training programme was an important factor in the amount of improvement attained (Cuddy, 1968; Heller & Auerbach, 1972).

Although auditory skills can be trained, it remains to be determined if listeners can be trained to integrate and use these skills to interpret an auditory graph more effectively. If they can, then questions follow as to how much and what type of training is most beneficial. Finally, it may be the case that the amount and type of training—or even the overall efficacy of a training programme—could depend on the contextual design features chosen for the display. For this reason, the present investigation explicitly studied auditory context, intended to provide additional and useful information to the listener, as well as training focused on the cognitive and perceptual tasks and skills hypothesized to be operative in the perception and use of auditory graphs.

STUDY 1

The first experiment focused exclusively on adding sounds to an auditory graph to provide context to either the x-axis (i.e. timing) or y-axis (i.e. scaling) in order to investigate the following specific research questions: (1) How does the addition of different forms of x-axis context affect performance in a point estimation task? (2) How does the addition of different forms of y-axis context affect performance in a point estimation task? and (3) Is there an interaction between the design features such that the benefits of adding one type of context depends on the presence or absence of the other? To answer these questions Study 1 employed a pretest-posttest experimental design to investigate the effects of adding two types of x-axis context and three types of y-axis context.

Participants and apparatus

A total of 160 undergraduate students at the Georgia Institute of Technology participated for course credit (97 males, 63 females; mean age = 20.4 years, $SD = 3.0$). All reported

normal or corrected to normal vision and hearing. The experiment was written in JavaScript, and run in Netscape Navigator v.4.77 on Mac OS 9.2. Instructions and visual stimuli appeared on a 17-in. (43.2 cm) Apple Macintosh studio display set to a resolution of 1024×768 pixels. Auditory stimuli were presented at a comfortable listening level (approximately 60 dB SPL) via Sony MDR-7506 headphones.

Baseline auditory graph stimuli

Participants listened to an auditory graph representing the variation in price of a single, unidentified stock over a 10-h trading day (from 8 am to 6 pm). The data (stock price) fluctuated discretely (two values per hour) and followed a pseudo-sinusoidal pattern, climbing from the opening price (\$50) to the maximum price of the day (\$84), then falling to the lowest price of the day (\$10) before returning to a closing price equal to the opening price. In the display, price (in dollars) was mapped to the frequency of a tone using a positive polarity, with the scaling factor for dollars and frequency for sighted listeners (0.9256) determined previously (Walker & Lane, 2001). Therefore, changes in the number of dollars (\$) were represented by changes in frequency (f) as $\$ = cf^{(0.9256)}$. The constant c was chosen to be 0.0765 so that the frequency of the tone for the starting value (\$50) would equal 1100 Hz. The frequency of the tones ranged, therefore, from 200–1900 Hz. Time over the 10-h trading day was mapped against time in the display, such that each hour of the trading day was represented by 1 s in the display, thus there were a total of twenty 0.5-s tones (one for each half hour in the day). The amplitude of each separate tone was ramped up and down over 40 ms to avoid sudden onsets or offsets. The tones were played sequentially, with no temporal separation between them, for a total graph length of 10 s.

Added x-axis and y-axis context sounds

In the pretest all participants heard only the baseline auditory graph stimuli, without added context. In the posttest, however, auditory context was added, as determined by the factorial combination x-axis context cues (no clicks, clicks) and y-axis context cues (no y-axis cues, static, dynamic) as described in Table 1. Note that Condition 1 involved no added context, as a replication of the baseline condition from the pretest (i.e. a control condition).

The x-axis context was created by the addition of a series of brief (10 ms) broadband clicks. Each represented the passing of 1 h in the day (1 click/s), and therefore played in time with every odd data point. Y-axis context was created via the addition of a beeping reference tone. The reference tone was produced as a triangle wave to be distinct from the sine wave tones representing the data. In the static y-axis context, the pitch of the reference tone always represented the opening price of the stock, and beeped at that starting pitch throughout the entire graph. In the dynamic y-axis context, the pitch of the reference tone represented the maximum or minimum price of the stock for the day, as follows. When the price of the stock was rising (such as at the beginning of the trial) the pitch of the reference tone corresponded to the highest price of the day (\$84). When the price was falling, the pitch of the reference tone changed to match the lowest price of the day (\$10). Thus, the dynamic tone provided a scaling cue for reference to maximum or minimum data values, as well as providing secondary information about the direction of stock price movement. The context sounds were in time with the data sounds, in that there were no temporal offsets to their starting times.

Table 1. Study 1 experimental conditions, numbers of participants, and mean RMS errors (adjusted and unadjusted)

	Y-axis context cue		
	No added y-axis context	Static reference tone	Dynamic reference tone
No added x-axis context	Condition 1 (<i>n</i> = 26) 19.44 (1.26) 21.66 (19.30) [†]	Condition 3 (<i>n</i> = 26) 18.89 (1.25) 18.23 (3.76) [†]	Condition 4 (<i>n</i> = 24) 13.13 (1.31) 10.80 (4.92) [†]
X-axis clicks	Condition 2 (<i>n</i> = 29) 16.74 (1.19) 15.02 (6.27) [†]	Condition 5 (<i>n</i> = 30) 16.02 (1.17) 17.97 (11.69) [†]	Condition 6 (<i>n</i> = 25) 11.14 (1.28) 11.41 (5.61) [†]

Note: Mean adjusted RMS error is shown in bold, with standard error of the estimate in parentheses. Unadjusted RMS error is shown below that, with standard deviation in parentheses, and marked with ([†]). In every condition, participants were given the initial price of the stock at the opening of the trading day (\$50).

Task and procedure

Before the pretest, listeners received a short explanation of the display and task, including the opening stock price. The task in both the pretest and posttest was to listen to the auditory graph and estimate the exact price of the stock at a randomly selected time of day (once for each of 11 times, ranging from 8 am to 6 pm). On each trial participants were reminded of the opening price, then saw a question asking for the price at a given hour (e.g. 10 am), then heard the entire auditory graph at least once, before typing an answer. Participants were able to listen to the graph as many times as required to answer the question. No accuracy feedback was provided. Trials were self-paced: participants clicked buttons to play the graph and then to advance to the next trial. The dependent variable, performance, was operationalized as the root mean squared (RMS) error (in dollars) with which a participant reported the 11 queried data values represented in the display. RMS error is a measure of dispersion of points from a set of known or fixed points, and is mathematically the spatial (two-dimensional in this case) equivalent of standard deviation.

In the pretest the auditory graph consisted of only the data sounds; there were no extra context sounds. Following the pretest and a brief rest, listeners were randomly assigned to one of the six listening conditions for the posttest. The task was the same as for the pretest (11 trials of estimating the price of the stock at different times of the day), but was completed with (in most cases) an augmented or enhanced auditory graph, described above. Instructions explained the purpose and meaning of the new sounds in the graphs (depending on the condition), including, if appropriate, the fact that the y-axis context sounds corresponded to the maximum and minimum prices (\$84 and \$10).

Hypotheses

We predicted that auditory graphs with added context sounds would lead to better performance than auditory graphs without context (i.e. the baseline condition). This improvement should come from added x-axis context, y-axis context and both in combination. Specifically, the addition of x-axis context (clicks) should eliminate the

interval division task, and assist the listener in the magnitude estimation task. When the listener is asked to report the price of the stock at noon she should 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 (noon). Responses should be more in phase with the actual data series. Since the listener is no longer required to listen to the entire graph, she is also 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 graph (or to y-axis context cues, if present). This should make it easier to estimate the price represented by the noon sound, relative to that of the opening sound, leading to better point estimation.

Y-axis context should help the listener 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. Consider again the trial where the listener reports the stock price at noon. Given a tone representing the opening price (i.e. static reference tone), the listener is still required to recall the pitch she perceived at approximately half the duration of the auditory graph, but is now able to judge the noon-time pitch relative to a sound she is hearing right now, instead of against a sound which, in the baseline condition, she would be forced to maintain in working memory. If the user is provided with a dynamic y-axis 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, the reference tone can also theoretically serve to reinforce the intended scaling. Thus, we predicted that both forms of y-axis context should lead to better performance (i.e. lower RMS errors), and, more specifically, that the dynamic tone should be better than the static tone.

It is difficult to predict the nature of any interaction. It is possible that the x-axis and y-axis context effects would be simply additive. Alternatively, a synergy could result from the doubly enhanced listening experience. On the other hand, so many additional sounds could be difficult to parse, leading to a confusing auditory graph. This would be analogous to visual clutter or chartjunk (see Tufte, 1990), and could result in reduced performance.

Results

Individual RMS error scores on the posttest were subjected to a two-way, univariate, analysis of covariance (ANCOVA), with type of x-axis and y-axis context as the between-subjects factors, and pretest score as the covariate. The pretest provides a measure of the participant's perceptual skills, listening experience, and any pre-established mental models of how sounds might vary, how data values like stock prices can vary, and how changes in a sound might be used to represent changes in data. Preliminary checks were conducted to ensure that there was no violation of the assumptions of the ANCOVA (see e.g. Huitema, 1980; Maxwell & Delaney, 1990; Porter & Raudenbush, 1987). There was a strong effect of pretest score on posttest score, with a partial eta squared of 0.61, indicating that the variability in the pretest scores accounted for more than 60% of the variability of the posttest RMS errors, making it an effective covariate.

The ANCOVA-adjusted group means for RMS error on the posttest are listed in Table 1 and plotted in Figure 1. There was a significant main effect of type of x-axis context, which reflects that, overall, listeners answered with smaller error if they were provided with x-axis context in the form of clicks (RMS error = 14.63), than if they were not (RMS

Study 1: Effects of Context

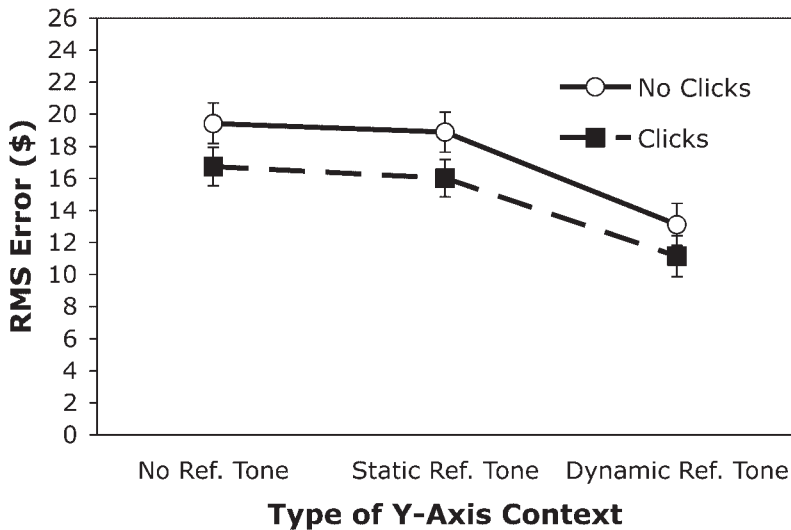


Figure 1. Effects of x-axis and y-axis context on RMS error (in dollars) in Study 1. Error bars represent one standard error above and below the mean. The addition of clicks and a dynamic reference tone each improved performance

error = 17.16), $F(1, 153) = 6.21$, $p = 0.014$, Cohen's $f = 0.20$. Note that Cohen's f of 0.10 typically indicates a small effect size; 0.25 indicates a medium effect; and 0.40 indicates a large effect.

There was also a significant main effect of type of y-axis context, $F(2, 153) = 13.27$, $p < 0.001$, Cohen's $f = 0.42$, indicating that RMS error differed between the types of y-axis context (see Figure 1). Simple planned contrasts between each type of y-axis context and the baseline condition (i.e. no y-axis context; RMS error = 18.09) revealed that the static (opening-price) reference tone (RMS error = 17.46) did not yield a significant improvement in performance, $p = 0.602$, whereas the dynamic (min/max-price) reference tone did (RMS error = 12.14; $p = 0.001$). There was no interaction of x-axis and y-axis context, $F(2, 153) = 0.068$, $p = 0.934$.

Discussion

First, the results of Study 1 supported our hypothesis that the addition of x-axis context in the form of clicks can improve performance on point estimation with an auditory graph. Second, the addition of dynamic y-axis context improved performance over the baseline condition, whereas the addition of static y-axis context did not lead to a significant improvement. And third, since there was no interaction between the two types of context, it seems that in this case the presence of both x and y-axis context led to neither a synergistic improvement, nor a decrement due to, for example, clutter.

Possible functional explanations for context

The discussion of context that accompanied the hypotheses considered the benefits of adding certain types of context, in terms of the aggregate RMS scores within a group. It is

also interesting to examine how these effects may arise at the individual level. One key precept is that context enhances performance only to the extent that it introduces new and useful information, and to the extent that it does not interfere with, clutter, or distract from more useful information. The response patterns in Figure 2 show how such an addition of context to the x-axis can result in better performance. In the figure, the unadjusted value estimates from each participant are traced with grey lines. The solid black line represents the correct data values. Note that the lines are continuous and smoothed in the figure for legibility, but the data values and responses were, of course, 11 separate data points for each participant. For clarity in the following discussion, the unadjusted responses of a single representative participant have been highlighted in the top-most pair of graphs in Figure 2 (filled triangles and dashed line).

First, as discussed previously, one would expect that the context provided by the addition of x-axis clicks would assist the user with the interval division task. Therefore, although a participant's y-axis scaling might remain unchanged, his or her responses would shift on the x-axis, and performance would improve to reflect the participant's better perception of the x-axis scaling chosen for the display. In the top-left panel of Figure 2 the listener perceives the changing price trends, but is out of phase with the intended display values. In the top-right panel this same participant's perception of the price changes is essentially unchanged in terms of magnitude (y-axis scaling), but the added x-axis context updates his x-axis scaling such that it nearly mirrors that intended by the display designer. The clicks help to bring the listener's responses more into phase with the actual auditory graph. These kinds of qualitative changes in responding are present to various degrees in performance traces throughout the data set, and are quantified by the statistically significant improvement observed in the aggregate scores of participants who were provided with the clicks.

Although the significant main effect for added clicks was expected, we questioned why there was not a larger improvement observed here. Mean RMS error, collapsing across types of y-axis context, only improved \$2.53 (from \$17.16 to \$14.63, a small-to-medium effect, $f = 0.20$). We had considered that updating the listener's x-axis scaling should result in a larger effect on performance. However, it is also possible that adding the clicks was not adding as much *useful* information as anticipated. That is, the changing price over the 10-h trading day was represented by a sound whose frequency varied discretely on the hour and half hour (i.e. each half-second in the graph). Listeners, given the knowledge that it was a 10-h trading day, likely counted 20 tones, realized that there were two tones per hour and thus, using the discrete variation of the frequencies in the data itself, had already extracted the information that the clicks were meant to provide. Thus, the baseline display is actually not entirely devoid of x-axis context. The clicks had the effect of providing an auditory version of major and minor tick-marks to the graph, which merely enhanced the x-axis context already present in the data sounds. Therefore, it is likely that we found a modest decrease in RMS error for added x-axis context because the amount of useful information added by the clicks was less than expected.

By imposing a y-axis scale (information not provided by the static reference tone), the dynamic reference tone provided more information for completing the perceptual and cognitive tasks involved in point estimation. The static reference tone did not help with the scaling, so it should not lead to as large a performance gain. The results with the y-axis context provide further evidence to support the principle that added context enhances performance so long as it introduces new and useful information and does not interfere with more useful information. The middle and bottom pairs of graphs in Figure 2 illustrate

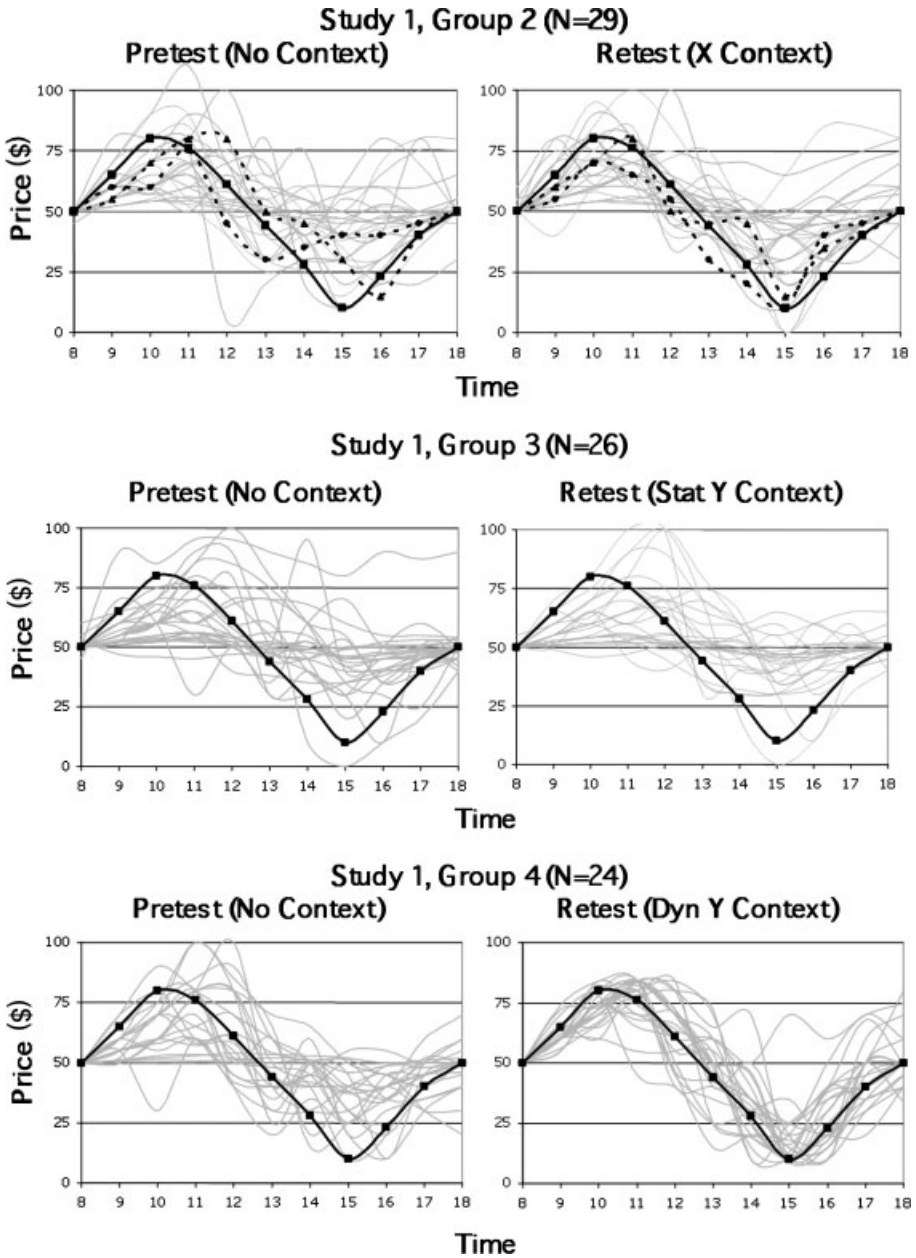


Figure 2. The correct displayed values (squares on solid black line) viewed with individual participant response traces from Study 1 (in grey). Responses from the pretest (left panels), and posttest (right panels) illustrate typical changes in performance observed between the discrete data mapping alone, and the discrete mapping plus added context of different types. The top pair of panels shows the effect of adding x-axis context (clicks). Participant 59 is highlighted with triangles and dashed line. The middle and bottom pairs of panels illustrate typical changes in performance with the addition of static and dynamic y-axis context, respectively

the effects of added y-axis context. The pretest performance traces for individuals in Conditions 3 and 4 (middle and bottom-left panels of Figure 2) show that pretest y-axis scalings are often significantly different from that intended by the display designer. Again, this is not surprising, given that participants have been given no specific instruction in how to interpret the auditory graphs. The responses in the posttest (middle and bottom-right panels, Figure 2) reflect the correct data values more accurately. Let us consider more closely the addition of the static reference tone, shown in the middle pair of graphs in Figure 2. The overall trends are more in line with the actual data patterns, but the individual y-axis scalings are generally unchanged. This makes sense because the context provided by the static reference tone helps with some working memory tasks and may help the listener judge the direction of change more readily, but does nothing to reinforce the designer-defined scaling and therefore fails to help the listener better judge magnitude or rate of change. Although it may have helped with certain cognitive tasks, it may not have provided *enough* useful, additional information. Alternatively, consider the addition of the dynamic reference tone, shown for Condition 4 in the bottom panels of Figure 2. Posttest responses again reflect the intended sinusoidal pattern more accurately than in the pretest, but also show that the y-axis scalings often become nearly exactly that chosen for the display. The context provided by the dynamic reference tone helps with the working memory tasks and also reinforces the chosen scaling, thereby helping the listener better judge the magnitude, direction, and rate of change in the price of the stock. For this reason, the posttest responses of participants who were provided with the dynamic reference tone reflect a relatively dramatic decrease in RMS error.

To summarize the results of Study 1, the addition of contextual design features can result in improvements in performance in a point estimation sonification task as the timing and scaling of the listener's responses is brought more in line with the intended design of the display. This is in keeping with the proposed task decomposition. However, not all context helped: x-axis clicks and the dynamic (min/max-price) reference tone each improved performance while the static opening-price reference tone did not. Finally, the results here indicated that the effects of x-axis context do not seem to depend on the presence or absence of y-axis context, at least as they were implemented here, suggesting that the interval division and magnitude estimation tasks are somewhat independent.

STUDY 2

In Study 1, the discussion of the effects of x-axis context brought up the issue of the data being displayed in separate, discrete tones. As discussed at the outset, many auditory displays are created using separate sounds for each discrete data point, so the use of that display design in Study 1 was perfectly reasonable. However, there are also many situations in data sonification where a changing data source drives changes in the pitch or loudness of a continuous sound. This is analogous to the use of a series of points versus a continuous line to represent changing data in a visual graph. It is important, then, to consider how performance with such a continuous graph might compare to performance with a discretely varying graph, as in Study 1.

Another important question is whether training affects performance. Instructing listeners how to perform the task, and how to take advantage of any context that is provided, should help the listeners to interpret the auditory display as intended by the

designers. In particular, training the listeners to divide the task into component subtasks, and providing training on those subtasks, should lead to improved performance (Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999; Cuddy, 1968).

To investigate these questions, the data interpretation task and the general experimental methods in Study 2 were held identical to Study 1. However, Study 2 differed in two specific ways. First, instead of mapping price to a sound whose frequency varied discretely, changing dollar values were mapped to a sound whose frequency varied smoothly and continuously through the data. Second, Study 2 incorporated the additional manipulation of training as an independent variable, thus employing a pretest-training-posttest experimental design to investigate the effects of x-axis context, y-axis context, and training.

Procedure

One hundred and fifty undergraduate students participated for course credit. Of these, 99 were students at the Georgia Institute of Technology (44 males, 55 females; mean age 20.3 years), and 51 were students at the United States Military Academy (USMA) at West Point (38 males, 13 females; mean age 19.8). None had participated in Study 1. Screening for normal vision and hearing were identical to Study 1. Also, Study 2 apparatus and software were identical to that used in Study 1, with the exception of the added training programmes (described below), which were created in Macromedia Director 8.5 Shock-wave Studio.

As in Study 1, the task required participants to listen to an auditory graph representing the variation in price of a single, unidentified stock over a 10-h trading day (from 8 am to 6 pm). The display was identical to that used before, with the important exception that the pitch variations of the baseline display sound were continuous as opposed to discrete. The pure-tone frequency started at the value representing the opening price, then smoothly glided up and down through the 20 subsequent data values. The pitch transitions lasted half a second, for the same total graph length as in Study 1. There were no discontinuities or other amplitude modulations in this baseline data display to distinguish when the graph was at a particular time of day.

As before, in the pretest the participants answered 11 point estimation questions about the information presented by an auditory graph that was devoid of any added context cues. Next, listeners were randomly assigned to experimental conditions (described below) and completed a training session or filler task. After this training (or filler task) participants received a short explanation of their new display, and then performed the posttest under the appropriate experimental condition of added context.

Conditions: x and y-axis context and training

The experimental conditions were defined by the factorial combination of type of x-axis context, type of y-axis context, and presence or absence of training (see Table 2). X-axis context was again provided by the addition of a stream of audible clicks identical to those in Study 1. Y-axis context was created via the addition of a dynamic (max/min price) beeping reference tone, identical to that used in Study 1. To constrain the size of this study, the static reference tone (ineffective in Study 1) was not used here.

In the No Training conditions (Conditions 1–4), participants completed a filler task consisting of three reading comprehension passages and accompanying reading compre-

Table 2. Study 2 experimental conditions, numbers of participants, and mean RMS errors (adjusted and unadjusted)

	No training (filler task)		Training	
	No added y-axis context	Dynamic reference tone	No added y-axis context	Dynamic reference tone
No added x-axis context	Condition 1 (<i>n</i> = 17) 24.49 (3.28) 23.62 (28.71) [†]	Condition 3 (<i>n</i> = 18) 12.63 (3.21) 15.23 (6.17) [†]	Condition 5 (<i>n</i> = 19) 8.37 (3.12) 11.03 (4.91) [†]	Condition 7 (<i>n</i> = 20) 11.66 (3.3) 10.40 (4.67) [†]
X-axis clicks	Condition 2 (<i>n</i> = 18) 20.54 (3.19) 21.96 (34.05) [†]	Condition 4 (<i>n</i> = 19) 12.90 (3.11) 11.03 (5.34) [†]	Condition 6 (<i>n</i> = 20) 12.70 (3.03) 11.27 (10.55) [†]	Condition 8 (<i>n</i> = 19) 9.27 (3.11) 8.28 (2.92) [†]

Note: Mean adjusted RMS error is shown in bold, with standard error of the estimate in parentheses. Unadjusted RMS error is shown below that, with standard deviation in parentheses, and marked with (†). In every condition, participants are given the initial price of the stock at the opening of the trading day (\$50).

hension questions. The reading comprehension filler task was judged as sufficiently similar in cognitive and sensory resource demands because it was of similar duration (25–30 min dependent upon the participant), it required attention and comprehension of materials presented, and both the training programmes and the filler task were interactive in nature (both required input and provided feedback). The primary difference between the filler task and the training programmes was that the filler task provided no training, or practice in the perception or interpretation of auditory graphs.

For the conditions involving training, a different training programme was created for each type of added context present in the posttest display. The training programmes were designed in accordance with established principles of learning and skill acquisition and broke the point estimation task into its component supporting tasks (Adams, 1987; Proctor & Dutta, 1995; Quinones & Ehrenstein, 1996; Salvendy, 1997; Smith, Best, Stubbs, Archibald, & Roberson-Nay, 2002). The nature of each supporting task was explained, and strategies practised for the completion of the task in the presence or absence of each of the applicable design features for that group. Feedback was provided as participants performed subtasks and completed point estimation tasks as appropriate, during the training.

For Condition 5 (training, but no x or y context) the training programme did not discuss the use of any context cues, focusing only on teaching the perceptual and cognitive skills used in the interval division and magnitude estimation tasks. These included perception of time and pitch, the relation of that perception to an imposed scaling, and finally the integration of these skills in a point estimation task. In addition to the information discussed for Condition 5, the training for Condition 6 (training plus x-axis context) included training on the nature and use of the added clicks to assist the user in the interval division task. The training for Condition 7 (training plus y-axis context) included the information from Condition 5, as well as training on the nature and use of the dynamic reference tone to assist the listener in the magnitude estimation task. Finally, the training programme for Condition 8 incorporated all the training described above.

Hypotheses

We predicted that there would again be a significant effect of adding dynamic y-axis context. We also hypothesized that there would be a significant effect of adding x-axis clicks, especially in this case where the continuous changes in the sounds representing the data do not provide as much temporal context as the discrete pitch steps in the Study 1 auditory graphs. Without clicks it should be much harder to keep track of the hour in the day, primarily making the interval division task more difficult. The listener would simply have to estimate the time passed in the day without any reference, which would have the effect of adding variability to the magnitude estimation task. The results of Study 1 suggest no interaction between x and y-axis context, however the continuous stimulus and the addition of training could result in a different pattern of results. We further hypothesized that trained participants would perform better than untrained participants, since understanding the subtasks involved is likely very helpful, as shown in other domains. We also hypothesized that there may be an interaction between training and the presence of context (x-axis, y-axis, or both), since it may be that training is more effective with certain kinds of subtasks. The details of this interaction, if any, could not be predicted.

Results

Individual scores on posttest RMS error were calculated as for Study 1 and subjected to a three-way univariate ANCOVA, using type of x and y-axis context, and level of training as the between-subjects factors. The effect of the covariate (pretest score) was moderate, with a partial eta squared of 0.31. The ANCOVA-adjusted group means are listed in Table 2. There was a significant main effect of training (see Figure 3), reflecting the fact that, overall, listeners answered with larger RMS errors if they received no training in the interpretation of the auditory graph (mean RMS error = 17.64; solid line of Figure 3), than if they did receive training (mean RMS error = 10.50; dashed line), $F(1, 141) = 10.41$, $p = 0.002$, Cohen's $f = 0.26$.

There was no significant main effect for the inclusion of x-axis context, $F(1, 141) = 0.038$, $p = 0.846$, Cohen's $f = 0.02$. There was, however, a significant main effect of type of y-axis context, demonstrating that listeners answered with smaller RMS errors if they had the dynamic y-axis reference tone than if they did not (mean RMS error = 11.61 and 16.53, respectively), $F(1, 141) = 4.92$, $p = 0.028$, Cohen's $f = 0.18$. Finally, training and y-axis context interacted (see Figure 3), $F(1, 141) = 4.77$, $p = 0.031$, $f = 0.31$, reflecting that for groups not receiving training, adding y-axis context helped, whereas for groups receiving training, y-axis context did not help.

Discussion

The results of Study 2 confirm the hypotheses of significant effects of training and y-axis context, and produced a significant interaction between amount of training and at least one type of added context (y-axis). However, the lack of a significant effect of adding x-axis context was unexpected—especially in light of the significant effect found in Study 1—so we begin our discussion there. Based on both the results of Study 1 and a power analysis (with 150 participants, and assuming even just a small effect size, estimated power was 0.90), we felt confident we would be able to detect any effects that were truly present. We expected that in both studies any new and useful information would produce improvements

Study 2: Effects of Training and Context

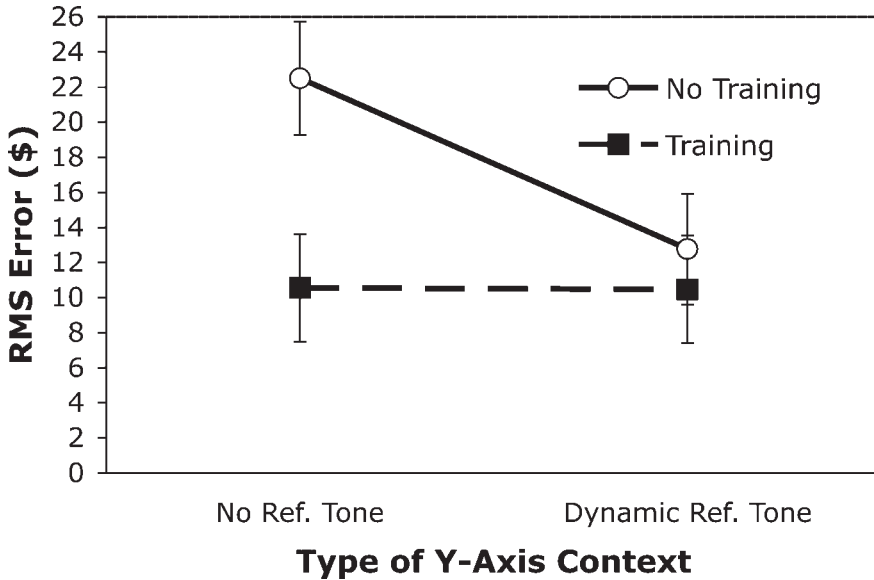


Figure 3. Effects of training and y-axis context on RMS error (in dollars) in Study 2. Note that there was no effect of adding x-axis context in Study 2. Error bars represent one standard error above and below the mean

in performance (Tufté, 2001; Walker, 2002). Thus, the lack of an effect of adding x-axis context in Study 2 seems inconsistent with the findings of Study 1. In the first study the discrete nature of the sound mapping already provided some incidental x-axis context, as defined in the beginning of this paper. Therefore, adding the clicks in Study 1 was not so much adding x-axis context, as it was adding *more* x-axis context, creating the effect of major and minor tick marks. In Study 2 the continuous nature of the sound mapping provided no x-axis context; thus the added clicks were the only x-axis context present, creating the effect of major tick marks only. One would expect the addition of clicks where no x-axis context previously existed to add more new and useful information than the addition of clicks where the discrete sound mapping already provides some x-axis context. This would presumably lead to an even larger effect size than obtained in Study 1. However, these results, viewed together with those of Study 1, seem to indicate that adding x-axis context where none exists does not necessarily result in improvements in performance. On the other hand, adding more x-axis context where some already exists can be helpful, so long as it has not cluttered or detracted from the perception of more useful information (Tufté, 2001; Walker, 2002). Perhaps creating the effect of major and minor tick marks (where before only major tick marks existed) adds a greater increment of usefulness over the addition of major tick marks only. It may be that adding clicks may only make a significant impact when the sound mapping itself already provides some x-axis context.

The discussion thus far has reflected an information processing perspective, but it is also possible to consider these results from a more psychoacoustical perspective. That is, in

changing from a discrete mapping to a continuously changing pitch, either the nature of the underlying dimension, or the way it is perceived, may have changed. For example, in Study 1 the sounds representing the data and the clicks, with similar temporal patterns and onsets, may have formed one integrated auditory stream. On the other hand, in Study 2 the continuous data sound and the clicks may have been dissimilar enough to form separate auditory streams (see Bregman, 1990; see also McNally & Handel, 1977, where dissimilar sounds formed different streams). Integrating and comparing the contents of two auditory streams, as would then be the case in Study 2, may have greater attentional demands than in Study 1, making the information in the clicks harder to use. Macken and colleagues (Macken, Tremblay, Alford, & Jones, 1999; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003) also discuss the possibilities of attention and memory disruptions resulting in or from different streams, which could impact performance here. It has also been suggested that a discretely changing pitch may act as a metathetic dimension, whereas a smoothly changing pitch may be closer to a prothetic dimension (for more, see Stevens, 1975, p. 58). While it remains exactly unclear how this would affect the utility of x-axis context, it is certainly at least worth noting that the listening experience may be qualitatively different in the discrete versus continuous mapping conditions.

It seems that, for the time being, the exact mechanisms underlying the effects of adding x-axis context invite further investigation. Regardless, it again points out the need for empirically tested design principles and guidelines to take the place of quasi-heuristics and intuitive design. The actual functioning, and therefore the practical utility of a somewhat widely accepted design technique (clicks) is apparently not as clear as expected (see also Walker & Kramer, 1996).

Although it is clear that training itself is beneficial to performance, it is difficult to discern exactly how training plays its role. Apart from sharpening sensory skills, training and the knowledge of results acquired via 'practice with feedback' both contribute to the user's contextual understanding of the display for subsequent trials. Practice with feedback (as in the training provided here) imparts knowledge of both x and y-axis scaling, which is not available to the untrained participant. As a result, in the untrained condition the effects of added y-axis context are evident (compare to Study 1), since context is the only means by which scaling information is communicated. In the trained conditions though, we find relatively little differences between contextual settings. It is sometimes difficult to distinguish the positive results of training in required skills, from the contextual scaling information imparted by knowledge of results gained in practice, from whatever effects might have been obtained by the addition of context. For example, the dynamic reference tone in Study 2 (Figure 4, bottom left) improved performance by not only helping with the working memory task, but also by reinforcing the preferred y-axis scaling. However, when training is provided (both right rows of Figure 4) there seems to be no effect of added contextual features beyond that of training alone. This may be because training and context provide partially redundant information—or it may be that the particular training programmes in this study were not effective enough in enhancing participants' abilities to use the context provided. Floor effects may also play a role here, in that RMS error may already have been reduced as much as possible in this experimental paradigm.

Overall, the results of Study 2 corroborate earlier findings pertaining to the performance effects of added context and the theory describing it. Furthermore, they clarify earlier findings by providing empirical evidence that: (1) training programmes focused on the operative cognitive and perceptual skills used in the perception of a sonification are an effective means of increasing performance; (2) that the effectiveness of added context

depends on the level of training given to the user; and (3) that the types of context chosen for the display may interact with the effectiveness of training in several of the specific subtasks required in a point estimation task.

GENERAL DISCUSSION

Comparing the findings across both studies, the results support the conclusion that the addition of useful information enhances human performance with auditory graphs. Further, the results suggest that context is not the only means by which to add useful information: training is also important in improving performance in point estimation sonification tasks.

Examining the group means from both studies (see Figure 5) shows a general decreasing trend in RMS error (i.e. better performance) from left to right as each succeeding condition

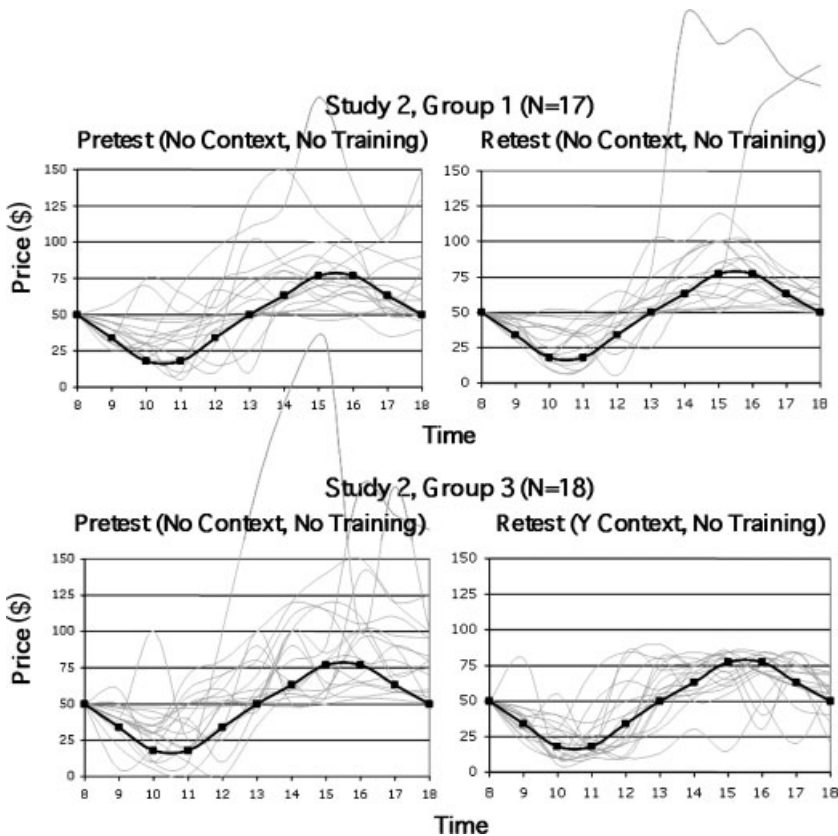


Figure 4. The correct display values from Study 2 (squares) viewed with participant responses from the pretest (left panels), and posttest (right panels) illustrate effects on performance due to adding various types of context and training. The top left pair of graphs shows the control group (No Context/No Training). The bottom left pair of graphs shows the effect of y-axis context for Group 3 (Y Context/No Training). The top right pair of graphs shows the effect of training in Group 5 (Training/No Context), and the bottom right pair of graphs shows the effects of both context and training observed in Group 8 (Training/X&Y Context)

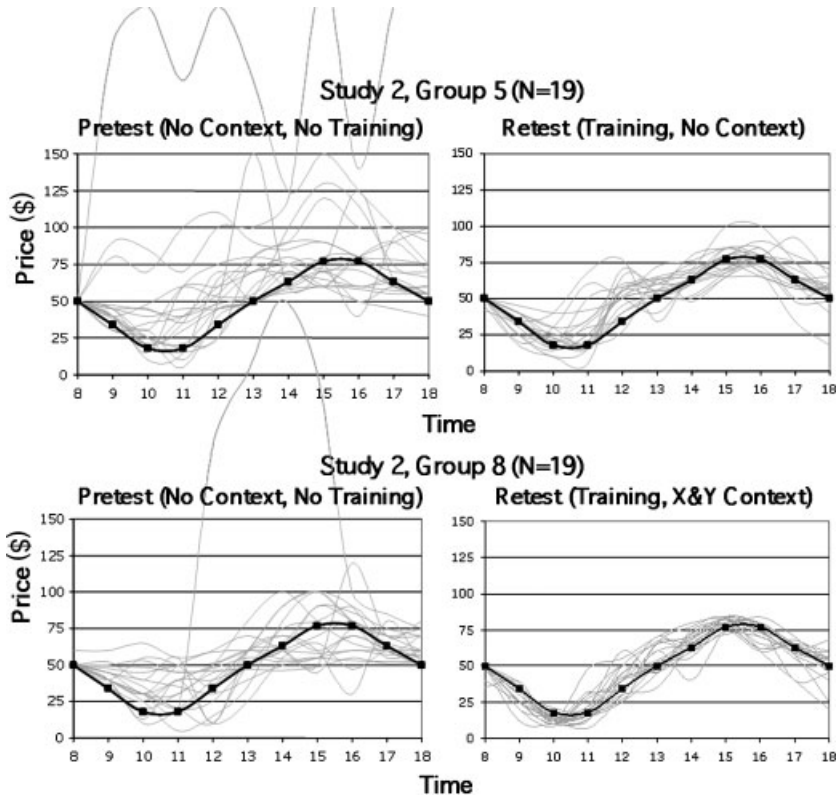


Figure 4. Continued

provides users with what should be greater amounts of new and useful information—and in the cases of the trained groups, as they are also provided with increased skill in the required cognitive and perceptual tasks. Note that this is not intended as a rigorous meta-analysis, but rather as a unifying discussion across these two studies, considering that they were very comparable in terms of tasks, conditions, and participants. We should also point out that the no-training conditions in Study 2 were essentially replications of Study 1, but with a continuous as opposed to discrete sound mapping.

We can consider the consistencies across the two studies, first. Beginning with the x-axis context, the first point to make is that the group with the highest mean error was in Study 2, Condition 1, which had no training or context cues at all (denoted *Baseline* in the right half of Figure 5). The continuous data set they heard even deprived them of the incidental x-axis context imparted by the discrete tones, as previously discussed. Next, compare Study 2, Condition 2 (*Clicks*), which had intentional x-axis context in the form of clicks, and Study 1, Condition 1 (*Baseline*), which had incidental x-axis context due to the discrete data mapping. Both of these conditions benefited from contextual settings that created the effect of ‘major tickmarks’ on the x-axis, and both showed numerically lower means than the *Baseline* condition in Study 2. Further, there was not much difference between them, suggesting that the effects of incidental and intentional context are of a similar magnitude. This point is further supported by comparing Study 1, Condition 4 (*Dynamic Y*) to Study 2, Condition 4 (*Clicks + Dynamic Y*). Here, again, the effect of

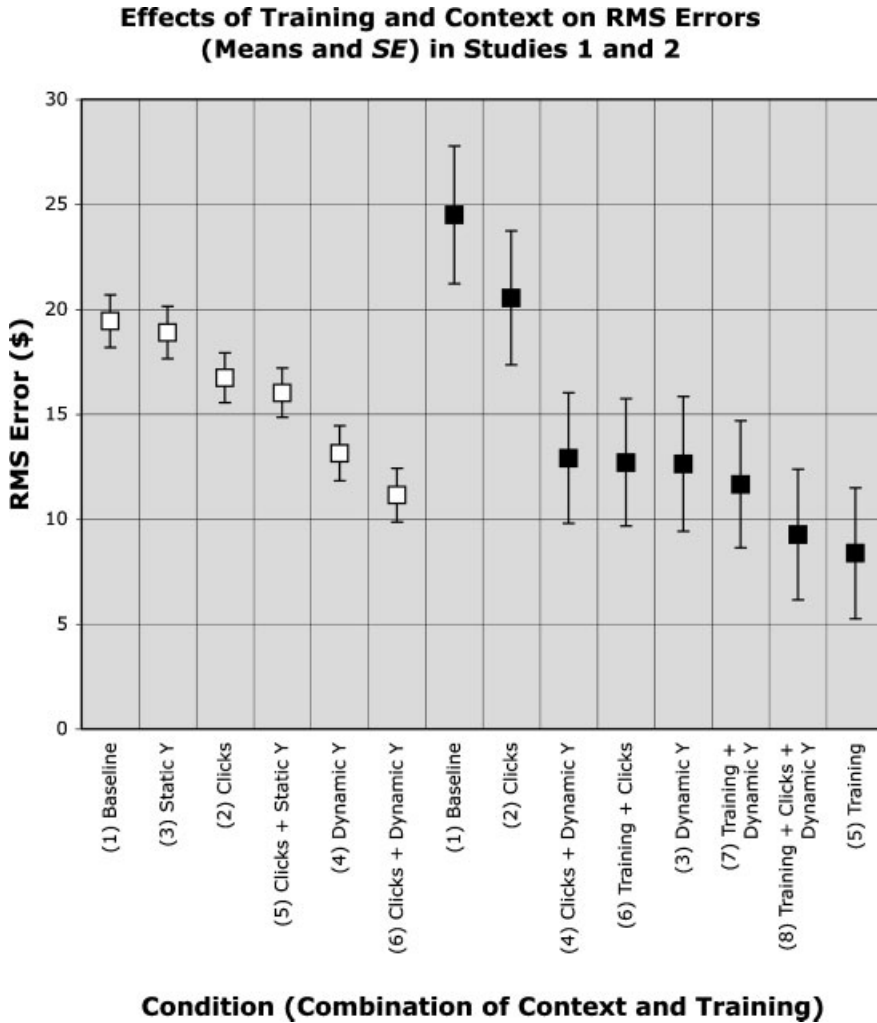


Figure 5. Mean RMS error for each study in order of error from greatest to least (left to right). Study 1 results are to the left, in open squares. Study 2 results are to the right in filled squares. Labels indicate condition number in parentheses, and the type of context and training (where appropriate) in the condition. Error bars represent standard errors. See text for detailed explanation

‘major tickmarks’, whether created by discrete data points or by a series of clicks, seem to have similar effects, at least in the present studies.

Next, adding y-axis context seemed to help (or at least not hurt) in both studies. Comparing Study 1, Condition 1 (*Baseline*) and Study 1, Condition 3 (*Static Y*), there is a slight numerical improvement in the means (a lessening of RMS error) resulting from the addition of static y-axis context. Comparing Study 1, Condition 1 (*Baseline*) to Study 1, Condition 4 (*Dynamic Y*) there is again an improvement when y-axis context is added, in this case in the form of dynamic max/min reference tones. This same pattern is evident when comparing Study 2, Condition 2 (*Clicks*) with Study 2, Condition 4 (*Clicks + Dynamic Y*). Thus, for both studies, the involvement of some form of context is beneficial. We do note that this is a little more complicated when the effects of training are considered

in Study 2, but those effects are not available in Study 1, so a direct comparison across studies is not possible.

There are also some notable differences between the two studies. Most notable is the significant beneficial effect of adding x-axis context in Study 1, but the lack of a similar effect in Study 2. For example, consider in Study 1, the difference between Conditions 1 (*Baseline*) and 2 (*Clicks*), as opposed to the lack of such a difference in Study 2, between Conditions 3 (*Dynamic Y*) and 4 (*Clicks + Dynamic Y*), which has already been discussed.

In addition to an examination across studies, another interesting question to consider is when and how the continued addition of context might begin to interfere with, clutter, or distract from more useful contextual information or from the data itself. The idea of clutter or 'chartjunk' is not new (Tufte, 1990), but it has yet to be conclusively demonstrated that the same principles will apply to auditory graphs. Looking at the categories representing the addition of training in Figure 5 (*Training*, compared to *Training + Clicks*, *Training + Dynamic Y*, and *Training + Clicks + Dynamic Y*), one might conclude that the information added and the skills acquired in training alone were sufficient to yield the observed improvement in performance—and that additional context in those conditions caused no additional benefit in performance because it was redundant (or merely cluttered the display). This may be true, but it is also possible that floor effects merely hid the effects of the context. It may be that given a more complicated and unpredictable data set, the effects of context combined with training would be more readily observed.

Finally, it is important to note that the practical importance of any decrease in error can only be evaluated relative to the actual magnitudes and variations in the data being displayed. For example, Study 2, Condition 5 (*Training*) showed the lowest overall mean RMS error of \$8.37. Compared to the *Baseline* group (RMS error of \$24.49), the training yielded an improvement of more than \$16 on average. Considered in another way, in both studies of this project the opening price of the stock was \$50, fluctuating during the day between a maximum of \$84 and a minimum of \$10 (with the average price of the day being \$50). Under the *Baseline* condition from Study 2 (no training, no incidental or added context), the error was almost 50% of the average price of the stock. With the addition of training the error dropped to about 16% of the average price of the stock. Regardless of the fact that this represents a sizable effect statistically ($f = 0.26$), the practical relevance of such a level of performance could still depend on any number of factors that would be evaluated in proper user and task analyses. To truly evaluate the usability *and utility* of such a graph requires replication and a more applied examination of specific users, tasks, and systems. Such an approach is essential to best facilitate progress and practical application of continuing innovations in this area. This approach should be applied to each innovation and new idea, as well as to evaluate the usability of the resulting displays. In this way we can best facilitate the practical applications which follow this type of research.

Representing data with sound for visually impaired students or scientists will be more effective if context cues are part of the sound design. Software to create such auditory graphs (e.g. The Sonification Sandbox, Walker & Cothran, 2003) needs to include methods for adding context, just as a visual graphing programme needs to include axes, tick marks, and so on. Further, design guidelines need to stress the importance of context, and at least survey the possible methods for implementing it. Context cues, and training in how to interpret the sonifications that employ them, will allow users of all sonifications to perform point estimation tasks, where before they were really only able to perform trend analysis tasks. The interpretation of complex stock market data (Nesbitt & Barrass, 2002), planetary seismology (Dombois, 2002), and the complex behaviour of Markov chain

Monte Carlo simulations (Hermann, Hansen, & Ritter, 2001) represent just a few of the wide range of recent sonifications and auditory graphs that could benefit from the improved point estimation abilities that appropriate auditory context cues can provide.

Finally, the results reported here have important implications for continued research in alternative information displays in general, and not just auditory displays. The purpose of this research is not to advocate any specific type of added context or training programme. Rather, it begins to quantify factors affecting performance, which in turn will lead to better design, implementation, and deployment auditory graphs and other alternative displays in ways that enhance system performance, give users more flexibility, and allow the improvement of system interfaces wherever such displays are required. The goal of all of these, of course, is a repertoire of effective and acceptable ways to go beyond the bounds of a visual display, for the benefit of all potential interface users.

ACKNOWLEDGEMENT

This research was completed as part of the Masters thesis entitled, 'Effects of Training and Context on Human Performance in a Point Estimation Sonification Task' by Daniel R. Smith, School of Psychology, Georgia Institute of Technology. Dan Smith is now on the faculty of the US MA at West Point. Parts of this research have been presented at International Conference on Auditory Display (ICAD), 2002 and at the Annual Meeting of the Human Factors and Ergonomics Society (HFES), 2004.

REFERENCES

- Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*, 101(1), 41–74.
- Ahissar, M. (1999). Perceptual learning. *Current Directions in Psychological Science*, 8(4), 124–128.
- Ahissar, M. (2001). Perceptual training: a tool for both modifying the brain and exploring it. *Proceedings of the National Academy of Sciences*, 98(21), 11842–11843.
- Aiken, E. G., & Lau, A. W. (1967). Response prompting and response confirmation: a review of recent literature. *Psychological Bulletin*, 68, 330–341.
- Annett, J. (1966). Training for perceptual skills. *Ergonomics*, 9(6), 459–468.
- Annett, J., & Patterson, L. (1967). Training for auditory detection. *Acta Psychologica*, 27, 420–426.
- Arno, P., Capelle, C., Wanet-Defalque, M. C., Catalan-Ahumada, M., & Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception*, 28(8), 1013–1029.
- Bertin, J. (1983). *Semiology of graphics* (W. J. Berg, Trans.). Madison, Wisconsin: The University of Wisconsin Press.
- Bonebright, T. L., Nees, M. A., Connerley, T. T., & McCain, G. R. (2001). Testing the effectiveness of sonified graphs for education: a programmatic research project. *Proceedings of the International Conference on Auditory Display* (pp. 62–66), Espoo, Finland.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. (Vol. xiii). Cambridge, MA: MIT Press.
- Brown, L. M., & Brewster, S. A. (2003). Drawing by Ear: Interpreting Sonified Line Graphs. *Proceedings of the International Conference on Auditory Display ICAD2003* (pp. 152–156), Boston, MA.
- Brown, L. M., Brewster, S. A., Ramloll, R., Burton, M., & Riedel, B. (2003). Design Guidelines for Audio Presentation of Graphs and Tables. *Proceedings of the International Conference on Auditory Display ICAD2003* (pp. 284–287), Boston, MA.
- Carlile, S. (2002). An overview of auditory dimensions that can be used to represent information. *Proceedings of the International Conference on Auditory Display* (pp. 3–8), Kyoto, Japan.
- Cuddy, L. L. (1968). Practice effects in the absolute judgement of pitch. *Journal of the Acoustical Society of America*, 43(5), 1069–1076.

- Dombois, F. (2002). Auditory seismology—On free oscillations, focal mechanisms, explosions, and synthetic seismograms. *Proceedings of the 8th International Conference on Auditory Display* (pp. 27–30), Kyoto, Japan.
- Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors*, 33(2), 205–231.
- Edworthy, J., Hellier, E. J., & Hards, R. (1995). The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics*, 38(11), 2341–2361.
- Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: superiority of an auditory over a visual display in a complex, multivariate system. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 307–326). Reading, MA: Addison-Wesley.
- Flowers, J. H., & Grafel, D. C. (2002). Perception of sonified daily weather records. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, Baltimore, Maryland.
- Flowers, J. H., & Hauer, T. A. (1995). Musical versus visual graphs: cross-modal equivalence in perception of time series data. *Human Factors*, 37(3), 553–569.
- Flowers, J. H., Buhman, D. C., & Turnage, K. D. (1997). Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors*, 39(3), 341–351.
- Gardner, J. A., Lundquist, R., & Sahyun, S. (1996). TRIANGLE: a practical application of non-speech audio for imparting information. *Proceedings of the International Conference on Auditory Display* (pp. 59–60), San Francisco, CA.
- Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: I. Linear regression modeling. *Human Factors*, 36(3), 419–440.
- Heller, M. A., & Auerbach, C. (1972). Practice effects in the absolute judgement of frequency. *Psychonomic Science*, 26(4), 222–224.
- Hellier, E., Wright, D. B., Edworthy, J., & Newstead, S. (2000). On the stability of the arousal strength of warning signal words. *Applied Cognitive Psychology*, 14, 577–592.
- Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: semantics versus acoustics. *Human Factors*, 44(1), 1–17.
- Hermann, T., Hansen, M. H., & Ritter, H. (2001). Sonification of Markov chain Monte Carlo simulations. *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)* (pp. 208–216), Espoo, Finland.
- Huitema, B. E. (1980). *Analysis of covariance and alternatives*. New York: John Wiley & Sons.
- Johannsen, G. (2002). Auditory display of directions and states for mobile systems. *Proceedings of the International Conference on Auditory Display* (pp. 98–103), Kyoto, Japan.
- Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J. G., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W. T., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M., Típei, S. (1999). The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: The International Community for Auditory Display (ICAD).
- Lunn, J. H. (1948). Chick sexing. *American Scientist*, 36, 280–287.
- Macken, W. J., Tremblay, S., Alford, D., & Jones, D. M. (1999). Attentional selectivity in short-term memory: similarity of process, not similarity of content, determines disruption. *International Journal of Psychology*, 34(5/6), 322–327.
- Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 43–51.
- Marila, J. (2002). Experimental comparison of complex and simple sounds in menu and hierarchy sonification. *Proceedings of the International Conference on Auditory Display* (pp. 104–108), Kyoto, Japan.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: a model comparison perspective*. Belmont, CA: Wadsworth Pub. Co.
- McAdams, S., & Bigand, E. (1993). *Thinking in sound: the cognitive psychology of human audition*. Oxford: Oxford University Press.
- McNally, K. A., & Handel, S. (1977). Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 3(3), 451–460.

- Moore, B. C. J. (1997). *An introduction to the psychology of hearing* (4th ed.). San Diego, CA: Academic Press.
- Nesbitt, K. V., & Barrass, S. (2002). Evaluation of a multimodal sonification and visualization of depth of market stock data. *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)* (pp. 233–238), Kyoto, Japan.
- Neuhoff, J. G., & Wayand, J. (2002). Pitch change, sonification, and musical expertise: which way is up? *Proceedings of the International Conference on Auditory Display* (pp. 351–356), Kyoto, Japan.
- Neuhoff, J. G., Kramer, G., & Wayand, J. (2002). Pitch and loudness interact in auditory displays: can the data get lost in the map? *Journal of Experimental Psychology: Applied*, 8(1), 17–25.
- Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: its model and use in psychological research. *Journal of Counseling Psychology*, 34(4), 383–392.
- Prather, D. C., Berry, G. A., & Bermudez, J. M. (1972). The effect of prompting and feedback on performance during learning, stress, and transfer of a perceptual skill. *Proceedings of the Annual Convention of the American Psychological Association*, 7(2), 643–644.
- Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage Publications.
- Quinones, M. A., & Ehrenstein, A. (1996). *Training for a rapidly changing workplace: Applications of psychological research*. Washington, DC: American Psychological Association.
- Salvendy, G. (1997). *Handbook of human factors and ergonomics* (2nd ed.). New York: Wiley.
- Sidley, N. A., Winograd, E., & Bedarf, E. W. (1965). Learning to identify complex sounds: prompting versus confirmation. *Journal of the Acoustical Society of America*, 38, 1050–1052.
- Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge: the role of graphs and tables in hard and soft psychology. *American Psychologist*, 57(10), 749–761.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- Swets, J. A., Millman, S. H., Fletcher, W. E., & Green, D. M. (1962). Learning to identify nonverbal sounds: an application of a computer as a teaching machine. *Journal of the Acoustical Society of America*, 34(7), 928–935.
- Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.
- Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, 8, 211–221.
- Walker, B. N., & Cothran, J. T. (2003). Sonification Sandbox: a graphical toolkit for auditory graphs. *Proceedings of the Ninth International Conference on Auditory Display ICAD2003* (pp. 161–163), Boston, MA.
- Walker, B. N., & Kramer, G. (1996). Mappings and metaphors in auditory displays: an experimental assessment. *Proceedings of the International Conference on Auditory Display* (pp. 71–74), Palo Alto, CA.
- Walker, B. N., & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: a comparison of visually impaired and sighted listeners. *Proceedings of the International Conference on Auditory Display* (pp. 90–94), Espoo, Finland.
- Walker, B. N., & Lindsay, J. (In press). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors*.
- Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. *Human Factors*, 30(5), 599–616.
- Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25(2), 227–248.
- Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An introduction to human factors engineering*. New York: Longman.
- Zotkin, D. N., Duraiswami, R., & Davis, L. S. (2002). Customizable auditory displays. *Proceedings of the International Conference on Auditory Display* (pp. 167–176), Kyoto, Japan.

Copyright of Applied Cognitive Psychology is the property of John Wiley & Sons Inc.. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.