Running head: TRAINING AND CONTEXT IN SONIFICATION

Effects of Training and Context on Human Performance in a Point Estimation Sonification Task

A Thesis Presented

To The Academic Faculty

by

Daniel R. Smith

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Psychology

Georgia Institute of Technology

April 2003

Abstract

This study extends research in sonification, auditory perception, and skill acquisition to examine ways to improve human performance with auditory graphs. Recent research has investigated use of mappings, scalings, and polarities (Neuhoff & Wayand, 2002; Walker, 2002), as well as the addition of contextual design features (Bonebright, Nees, Connerley, & McCain, 2001; Flowers, Buhman, & Turnage, 1997; Flowers & Hauer, 1995; Gardner, Lundquist, & Sahyun, 1996), but little has been done to quantify and compare the performance effects of such features, or to investigate performance effects of training in specific sonification tasks such as point estimation. In Study 1 of this thesis, the performance effects of adding several contextual design features were measured and compared. An analysis of the data revealed that contextual design features, such as x-axis clicks (in conjunction with a discretely varying sound data mapping), and a dynamic y-axis reference tone (reinforcing the correct y-axis scaling) both improved performance. While a static y-axis reference tone (representing the y-axis value of the initial data point) did not. Study 2 expanded on those findings by investigating the performance effects of added contextual design features under different conditions - and in conjunction with training. An analysis of those data revealed that when the data to be displayed was mapped to a continuously varying sound, adding the dynamic y-axis reference tone still improved performance - as did training - but adding the x- axis clicks did not. Finally, an interaction was discovered between training and the dynamic y-axis reference tone.

Table of Contents

Effects of Training and Context on Human Performance in a

Point Estimation Sonification Task

Displays of quantitative information are often an essential part of human machine systems. In addition, they find utility in numerous circumstances including, but not limited to, education, management, and the analysis and interpretation of scientific results (L. D. Smith, Best, Stubbs, Archibald, & Roberson-Nay, 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 et al., 1997; Flowers & Hauer, 1995; Kramer et al., 1999; McAdams & Bigand, 1993; Moore, 1997). Second, often the perceiver is either unable to *look* at, or unable to *see*, a visual display. The visual system might be busy with another task (Cohen, 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 (Cohen, 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 environments (Kramer et al., 1999; Moore, 1997). Last, and perhaps most salient, 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. In light of these issues, researchers and designers are seeking better, more flexible ways to display information to the user and have begun to develop a base of knowledge for use in design and implementation of displays using modalities other than vision.

Sonification

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). It is hoped that sonification will provide alternatives and flexibility for design of future information displays, but as the use of sonification expands, researchers and designers continue to struggle to increase performance of the user-display subsystem to levels that enable the use of sonification in more complex systems and displays (Johannsen, 2002; Marila, 2002; Neuhoff & Wayand, 2002; Zotkin, Duraiswami, & Davis, 2002). Thus far, research has focused on exploring new applications, and on ways to better design the sonifications themselves. But among the many options being considered, little research has been directed at qualitative evaluation of the performance effects of using specific contextual design features alone - or in conjunction with training. The current study extends research in sonification, auditory perception, and skill acquisition to examine these issues in relation to one specific type of sonification: auditory graphs.

Auditory Graphs

An auditory graph is one type of sonification. An auditory graph is not only a sonification of numerical information, but they also include design features intended to provide context and assist the user in a more accurate perception of the data relations. A sonification is analogous to a visualization as an alternative means of communicating, or otherwise helping a perceiver comprehend, or "visualize" quantities and variations within a given data set (Walker, 2002). But given the added contextual features (e.g., common features include auditory tick-marks, axes, labels, etc.), auditory graphs become analogous to visual graphs, prompting theorists to predict better, more accurate perceptions of the data being displayed (Walker, 2002). It is important to note that although displays in more than three dimensions can be problematic in most visual paradigms, dynamic multidimensional auditory graphs are capable of complexity not easily accessible to vision based displays (Kramer, 1993). Still, 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). For these reasons, despite the advantages of the auditory modality and the medium, the potential of well designed and implemented auditory graphs, as an alternative solution for information display is difficult to quantify. Further, exactly how to best design and implement such displays is still in question.

Design and Implementation of Auditory Graphs: What Do We Know?

While very little research has explicitly studied auditory graphs, there are findings from the study of sonification, more generally. Progressing from early applications such as warning sounds, recent endeavors sonify more dynamic, multidimensional events and processes. In one such effort, researchers examined use of auditory icons in a complex, cooperative task. Participants controlled a bottling factory software simulation with, and without auditory icons. The results suggested that the presence of the auditory icons added to participants' perception of the plant, as well as to their collaboration (Gaver, Smith, & O'Shea, 1991). In another such effort, auditory displays were studied as a means for monitoring and controlling the operation of a computer simulated crystal factory (Walker & Kramer, 1996). Studies such as these illuminated the need to develop a set of principles to be used to guide the representation of data dimensions, such as temperature, through sound display dimensions, such as pitch.

Mappings, Scalings, and Polarities

This representation of quantitative conceptual data dimensions through purposeful and meaningful co-variation of chosen sound parameters is known as *mapping*. In sonification, the available display dimensions are sound parameters such as frequency (pitch), amplitude (volume), timbre, and tempo, among others (Carlile, 2002; Walker & Kramer, 1996).

Given a specific mapping, sonification design then requires the selection of polarities and scalings. *Polarity* refers to how the data dimension and the display dimension co-vary. In other words, if a data dimension (e.g., temperature) increases, a *positive polarity* would dictate that such a change be represented by a corresponding increase in the assigned display dimension (e.g., increasing pitch). A *negative polarity* would dictate that such a change be represented by a corresponding decrease in the assigned display dimension. Lastly, *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 display dimension to data dimension magnitude estimation plot (Walker, 2002).

With the realization that these factors were central to design, sonification researchers incorporated magnitude estimation studies to search for optimal mappings, and the best polarities and scalings for those mappings. An important finding was that users prefer some mappings, scalings, and polarities over others, leading to the conclusion that the use of the most preferred mappings, scalings, and polarities should lead to better performance of the human-display subsystem (Neuhoff & Wayand, 2002; Walker, 2002).

Researchers have been seeking other ways to improve performance as well. Especially, seeking to capitalize on important parallels existing between auditory and visual displays. One such parallel involves the use of intentional context. Other possibilities include the investigation of individual differences, perceptual learning, and the application of training (Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999; Johannsen, 2002; Neuhoff & Wayand, 2002; D. R. Smith & Walker, 2002).

Context and Training In Sonification Tasks

Sonification research has built upon on the full literature relating to the physiological, perceptual, and cognitive aspects of auditory perception (Kramer et al., 1999; Moore, 1997). However, there is little research pertaining to the acquisition and application of the perceptual and cognitive skills employed in sonification tasks. In addition, it is hypothesized that the perceptual and cognitive skills in question vary according to the type of sonification task being executed, and according to the contextual design features provided by the display (D. R. Smith & Walker, 2002).

Intentional context.

Intentional context (hereafter referred to as simply “context”) refers to the purposeful addition of non-signal 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 (Bertin, 1983; Tufte, 1990). 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 (Bertin, 1983).

Consider then, the effect on readability if the display were devoid of all contexts. In visual line graphs, for example, the line of data itself provides some context, but only the incidental context inherent in the observation that some data points are farther to the right, and above or below the data points to the left (Walker, 1994). This incidental context might enable an observer to execute a rudimentary *trend analysis* of the data (e.g., Is the line generally rising or falling?), but the accurate extraction of a specific value (i.e., a *point estimation task*) is impossible. Unfortunately, many auditory graphs, even if making use of optimal mappings, polarities, and scalings, employ this impoverished amount and type of context (D. R. Smith & Walker, 2002; Walker, 2002).

It is important to note that there is a limit to how much context should be added to a display. Presumably, there is actually an optimal amount of context. After some point, the addition of context fails to provide useful additional information. Instead, it interferes with, clutters, or distracts from the extraction of more useful information (Marila, 2002; D. R. Smith & Walker, 2002; Tufte, 2001). 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 auditory graphs. Assuming findings show that the addition of context improves performance, there are countless types of context one might add (auditory equivalents to x-axis context, y-axis context, thresholds, labels, or other types of 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 appropriate or applicable in other modalities.

A common method for adding x-axis context to an auditory graph is the addition of a series of clicks or percussive sounds (Bonebright et al., 2001; Flowers et al., 1997; Gardner et al., 1996). But there are undoubtedly many ways to add context to an axis. However, 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.

Smith and Walker (2002) made an initial attempt to quantify and compare the benefits of adding some commonly used auditory contextual cues. Using an auditory graph of simple stock price data, they measured the performance effects of adding different contextual cues relative to and in combination with other cues in both trend analysis and point estimation sonification tasks. Their findings provide evidence that added context enhances performance so long as it introduces new and useful information and does not interfere with, clutter, or distract from more useful information.

The results also led them to conclude that some contextual features seemed to improve performance more than others. For example, the addition of a dynamic reference tone (reinforcing the preferred y-axis scaling) seemed to improve performance, whereas a static reference tone (representing the y-value of the initial data point) did not. Also, the addition of x-axis context (such as percussive clicks that reinforce the preferred x-axis time scaling) did not produce a significant improvement in performance (D. R. Smith & Walker, 2002).

In that study, it was unclear exactly why certain design features improved performance more than others. For example, it was unclear as to why x-axis context failed to produce improved performance when the addition of y-axis context did. It was hypothesized that x-axis context would be valuable. The clicks, heard once a second in the 10-second display, were meant to update the listener's x-axis scaling from 1 hour = 1 hour, to 1 hour = 1 second. However, it is possible that the x-axis context, while adding useful information, was not adding *additional* information. To be specific, the changing price (over the 10-hour trading day) was represented by a sound whose frequency varied discretely on the hour and half hour (e.g, on the second and half second in the graph). Listeners, given the knowledge that it was a 10-hour 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 (D. R. Smith & Walker, 2002). Thus, questions remain about the performance effects of adding context.

Training.

When considering the design of potential training programs for sonification, there are several questions about how different types of added context may, or may not, affect the performance benefit of a given training program, and inversely, how a training program might affect the performance benefits of different types of added context.

Training has been defined as attempting to change individuals in a manner that is consistent with task requirements. It is also defined as a way of applying principles of human learning and skill acquisition. Both definitions imply a focus on individual capabilities, task characteristics, and information processing demands (Quinones & Ehrenstein, 1996). For example, consider the information processing demands and task characteristics for completing a point estimation task using an auditory graph. First, the perceptual and cognitive tasks involved in the perception of an auditory graph vary greatly from those involved in the perception of a visual graph. They also vary greatly in accordance with the amount and type of contextual cues added to the display. Specific techniques and features used in the creation of context vary as well. Therefore, the results of a detailed task analysis - and thus the creation of a focused training program - would most likely depend on the type of context provided to the user.

For the purposes of this analysis, consider a simple auditory graph of stock market data where price is mapped to frequency such that an increase in frequency represents a corresponding increase in price (y-axis); and time is mapped to time such that 1 second of sound represents the price variation over 1 hour in real time (x-axis). As described, sounds varying in pitch over 10 seconds would represent the variations in the price of a stock over a 10-hour trading day. The only added context is the initial stock price, told to the listener. Further contextual cues are provided incidentally by the sound itself, such as the duration of the sound and the specific variations of the sound parameter mapped to the data (in this case it is frequency).

Given such a graph, determining the value of a given 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 (D. R. Smith & Walker, 2002).

One might assume, perhaps with good reason, that this set of tasks will be extremely difficult and result in relatively poor performance of the human-display subsystem. One might also assume that the cognitive demands of the listening task might outweigh the gains realized by the use of modalities other than vision. Neither question can be fully answered without empirical evidence pertaining to the efficacy of training, acquired skill proficiency, and context in such a complex and difficult set of tasks.

The purpose of training is to acquire skill proficiency (Adams, 1987). The three primary characteristics of skill, as defined in psychological literature, are that: (1) skill encompasses a wide variety of behavior, most of which are complex; (2) skill proficiency is not innate and must be learned; and (3) goal attainment requires combinations of perceptual, cognitive, and motor behaviors – each with different weights of importance (Adams, 1987). Some other characteristics include the observations that (4) skills develop in response to some demand; (5) proficiency is said to have been acquired when the trained behavior is integrated and well organized; and (6) as proficiency is acquired, perceptual and cognitive demands are reduced, allowing improved performance and freeing limited mental resources for other activities (Adams, 1987; Annett, 1991; Fitts, 1964; Proctor & Dutta, 1995).

The investigation of training (and its relation to the performance effects of added context) seems intuitively promising because of the relative inexperience of the general population in the use of sonification. People tend to be familiar with visual graphs and the visual display of quantitative information, but may have never heard of a sonification, let alone experienced one (Kramer et al., 1999). Further, relative to visual displays, charts, or graphs, very few of us have any experience, much less training, in the skills required in the accurate perception of auditory graphs, or in the skills required for the efficient use of sonified contextual cues (if present). In addition (as mentioned above) the perceptual and cognitive processes involved differ greatly relative to those involved in reading a visual graph. Given these facts, it is no wonder the tasks seem overwhelming.

Yet with training, it is certainly possible to become skilled at unfamiliar, perceptually difficult, complex, and/or subjective tasks. The skills involved in wine tasting, for example, are difficult to master. Yet, with training humans are capable of surprising feats of perceptual skill and accuracy in this arena (Proctor & Dutta, 1995). Recognizing ripe melons and classifying day-old chicks by sex are additional examples of perceptually difficult and subjective tasks.

Tasks such as these may be even more difficult because they require discrimination between perceptual stimuli that are virtually identical. In fact, the sex organs of male and female day-old chicks appear exactly identical, even to poultry farmers. But enhanced ability to discriminate between and classify stimuli based on perceivable properties is the hallmark of perceptual skill; and after three to eight weeks of training, expert “chick sexers” can classify up to 1000 chicks an hour with over 98% accuracy (Lunn, 1948).

Admittedly, both tasks described above only require simple dichotomous responses (e.g., good or bad; male or female). But such a response might be all that is required from an auditory graph. The user may only be required to recognize the sound of a good or bad state of system parameters in relation to others; or merely identify if there is a need (or not) to take the time to examine a more detailed visual or multi-modal display. Simple auditory alarms might suffice in some dichotomous response situations, but they rarely facilitate data visualization, or human situational awareness any more than a written table or spoken digits (L. D. Smith et al., 2002). Such alternatives rarely allow for the checking and monitoring of system parameters before alarm conditions are met - nor do they impart knowledge of the multivariate temporal trends that may have led to the alarm condition. Lastly, and perhaps most prohibitive, even contemporary computing systems lack the situational awareness/understanding required to reliably determine situations where a certain set of conditions requires a response (or alarm), versus situations when the exact same data relations are an expected, acceptable, and normal part of system operation. Often such analysis requires situational awareness currently only possible through human data visualization - anything less risks the possibility of becoming more of an unreliable distraction or annoyance, than an informative and effective display. In contrast, given training in certain cognitive and perceptual skills, an auditory graph might fit where other solutions fail.

Perceptual learning is defined as an increase in the ability to extract information from the environment as a result of experience, training, or practice with the stimulation coming from it (Gibson, 1969). As the above cases illustrate, it seems apparent that training can result in perceptual learning and consistent and reproducible improvements in performance, even in the most difficult, complex, and ambiguous circumstances. Further, such improvements can result from perceptual learning in tasks involving such diverse modalities as touch, vision, and taste. Finally, research also shows that training can result in perceptual learning specific to audition as well. For example, Cuddy (1968) found that students trained in music judge pitch more accurately than untrained listeners. Results indicate not only that training could alter the performance of untrained subjects, but that different training programs result in different amounts of improvement (Cuddy, 1968; Heller & Auerbach, 1972). Even in complex tasks similar to interpreting auditory graphs, results support the existence of training benefits (such as those found in the recognition of visual patterns via a vision-to-audition sensory-substitution system) (Arno et al., 1999).

Although research exists to show that auditory perceptual skills can be trained, little research has been done to examine if people can be trained to integrate and use these skills to better perceive an auditory graph. If they can, then how much and what type of training is most effective? Finally, it may be the case that the amount and type of training - or even the overall efficacy of a training program - could depend on the contextual design features chosen for the display.

Rationale for the Present Investigation

To date, most research efforts in sonification have been directed at investigating new applications and achieving higher levels of performance. Some research has been done to evaluate the effectiveness of specific contextual design features and innovations (D. R. Smith & Walker, 2002). But many questions remain, such as: what are the performance effects of added x-axis context in auditory graphs, and how might the *implementation* of auditory graphs affect their performance - specifically, how might training programs affect performance, and how might design features, such as added context, affect the efficacy of a given training program.

It was already evident that the assignment of mappings, polarities, and scalings are important distinctions in design (Walker, 2002). But the quantitative measurement and comparison of the specific performance effects of contextual design features - alone, in combination, and in conjunction with training - is the next logical step in the effort to verify factors governing sonification performance, describe their effects, and apply this knowledge in the innovation of potential solutions for improving performance.

For this reason, the present investigation explicitly tested contextual design features, which provide additional and useful information. This research also examined the implementation of training focused on the cognitive and perceptual tasks and skills hypothesized to be operative in the perception and use of auditory graphs and sonified contextual features expected to result in significant differences in human performance in a point estimation task.

Overview of Method

This thesis consists of two studies. Study 1 incorporated and extended Smith and Walker (2002), adding participants and making use of more sophisticated and appropriate statistical procedures (a two-way Analysis of Covariance (ANCOVA)), to clarify earlier findings. In addition, the more sophisticated analysis provided a direct statistical test of the potential interaction between added x and y-axis context (an analysis not possible given the procedures chosen for the earlier study). Finally, unlike Smith and Walker (2002), Study 1 focused entirely on the point estimation task and did not evaluate performance effects on a trend analysis task.

Study 2 built on Study 1 in two ways. First, instead of mapping price to a sound where frequency varies discretely, Study 2 mapped it to a sound whose frequency varies continuously. Measurement and comparison of performance under this condition provided data to address the question of the lack of an x-axis performance effect in Smith and Walker (2002). That is, did the added clicks not provide *useful* information, or did they merely not provide *additional* information? Second, Study 2 incorporated the additional manipulation of *training* as an independent variable. Measurement and comparison of performance under different conditions (defined by the combination of added x and/or y-axis context, with two levels of training) provided data to address three important issues. First, it would replicate or refute earlier findings concerning the performance effects of adding dynamic y-axis reference tone - and it would do so under different conditions (e.g., a continuous sound mapping, as opposed to the discrete sound mapping used by Smith and Walker (2002)). Second, it would provide evidence for or against the efficacy of training programs for improving human performance in point estimation sonification tasks. Third, it provided data allowing the evaluation of potential interactions between different contextual features and amount of training.

Study 1

Study 1 made use of the same experimental task, methods, and procedures employed by Smith and Walker (2002). As proposed, it incorporated the point estimation data already collected by Smith and Walker (2002), but it also expanded upon it through the inclusion of additional participants run under identical experimental conditions.

Participants and Apparatus

In all, a total of 160 undergraduate students from the Georgia Institute of Technology participated for course credit. All participants took part in an informed consent dialogue with the researcher, read and signed informed consent forms, and provided demographic details about age, sex, handedness, and number of years of musical training (see Table 1 below). All reported normal or corrected to normal vision and hearing.

Instructions and visual stimuli appeared on a 17-in. (43.2 cm) Apple Macintosh studio display set to a resolution of 1024 x 768 pixels, and typically viewed from a distance of 24 in. (61 cm). Auditory stimuli were presented via Sony MDR-7506 headphones, adjusted for fit and comfort. The experiment was written in JavaScript, and run in Netscape Navigator v.4.77 on Macintosh OS 9.2. The amplitude and frequency of stimuli were of sufficient and limited ranges so as to ensure the hearing protection of the participants.

Table 1

Study 1 Participant Demographics

		Handedness		Age		Years Of Music Experience	
	N	Left	Right	Mean	SD	Mean	SD
Men	97	10 (10%)	87 (90%)	20.47	2.42	3.76	4.32
Women	63	8 (13%)	55 (87%)	20.38	3.81	5.14	6.02
Total	160	18 (11%)	142 (89%)	20.44	3.03	4.31	5.09

Design and Procedure

Study 1 employed a pretest-retest experimental design to investigate the effects of two between-subjects independent variables. During the pretest, all listeners received a short explanation of the nature of the display and task. Then they performed under the control condition for a total of 11 point estimation trials. On each of the 11 trials, participants were asked to listen to the 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). Participants were able to listen to the graph as many times as required to answer each question. No accuracy feedback was provided. After random assignment, listeners received a short explanation of the nature of their new display, and then completed the retest (another 11 trials) under one of six experimental conditions (defined below). Upon completion of the retest, participants were thoroughly debriefed, thanked, and released. (See task view screen-shots at Appendix B).

Variables.

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. Individual scores were calculated by subtracting each observed response from the correct (or expected) response, squaring the resulting value, then taking the mean of the 11 squared error terms, and finally the square root of that mean to yield each individual's RMS error score. The process was repeated to calculate RMS error scores for the retest in the same way as for the pretest.

The experimental conditions were defined by the combination of independent variables (two types of x-axis, and three types of y-axis context) as illustrated in Table 2 below.

Conditions.

In the pretest, all participants performed under the control condition, which was devoid of added context, relying entirely on the intuitions of the novice participant.

In the retest, 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, created via the addition of a beeping reference tone. The pitch of this reference tone was static, constantly representing the opening price of the stock. Group 4 also experienced the graph with added 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 reference tone changed to match 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 static (opening-price) reference tone from Group 3). Finally, the graph for Group 6 combined x-axis context (clicks) with y-axis context (the dynamic (min/max-price) reference tone from Group 4).

Table 2

Study 1 Experimental Conditions

	No added y-axis context	Static ref. tone (representing opening price of stock)	Dynamic ref. tone (reinforcing y-axis pitch to dollars scaling)
No added x-axis context	Group 1 (Control)	Group 3	Group 4
Clicks (reinforcing x-axis time scaling)	Group 2	Group 5	Group 6

Note. In every condition, participants were given the initial price of the stock at the opening of the trading day ($50).

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 (Walker & Lane, 2001).

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

$$Y \alpha X^{(.9256)}$$

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

Results

Individual RMS error scores were calculated as they were in Study 1and subjected to the following statistical analysis. First, a two-way, univariate, ANCOVA was conducted on retest scores, using type of x and y-axis context as the between-subjects independent variables, and pretest score as the covariate.

The results of the ANCOVA are presented in Figure 1 (see descriptive statistics at Table 3). There was a significant main effect of type of x-axis context, which reflects that listeners answered with smaller RMS errors if they were provided with x-axis context in the form of clicks ($F(1,160) = 6.211$, $p = .014$). There was also a significant main effect of type of y-axis context ($F(2,160) = 13.270$, $p < .001$); but no x by y-axis context interaction. Simple planned contrasts between each type of y-axis context and the control group revealed that although the static (opening-price) reference tone did not yield a significant decrease in RMS error, the dynamic (min/max-price) reference tone did (p<.001).

Table 3

Study 1 Descriptive Statistics: RMS Error

Type of x-axis context	Type of y-axis context	Mean	SD	N
No added clicks	No added Ref. Tone	21.65	19.30	26
	Static Ref.Tone	18.23	3.76	26
	Dynamic Ref. Tone	10.80	4.92	24
	Total	17.06	12.52	76
Clicks (reinforcing x-axis time scaling)	No added Ref. Tone	15.02	6.27	29
	Static Ref.Tone	17.97	11.69	30
	Dynamic Ref. Tone	11.41	5.61	25
	Total	14.99	8.78	84
Total	No added Ref. Tone	18.16	14.28	55
	Static Ref.Tone	18.09	8.86	56
	Dynamic Ref. Tone	11.11	5.24	49
	Total	15.98	10.73	160

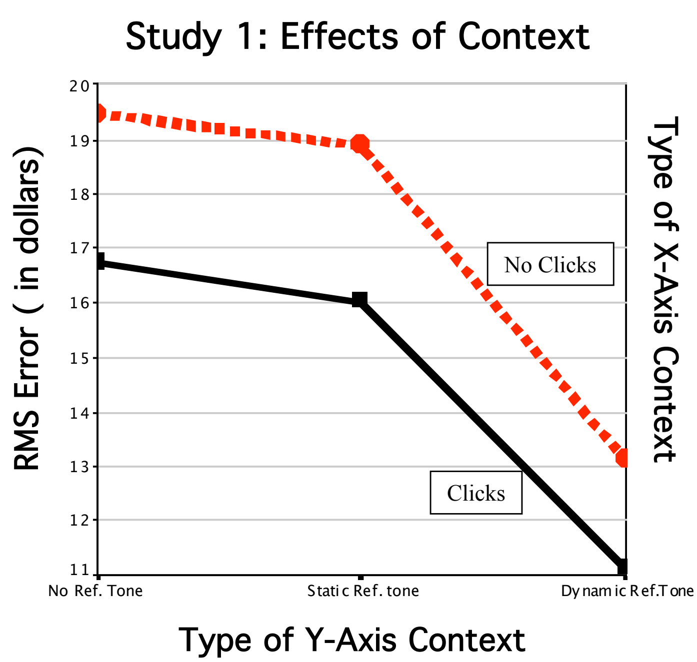


Figure 1. Study 1 effects of context on RMS error (in dollars).

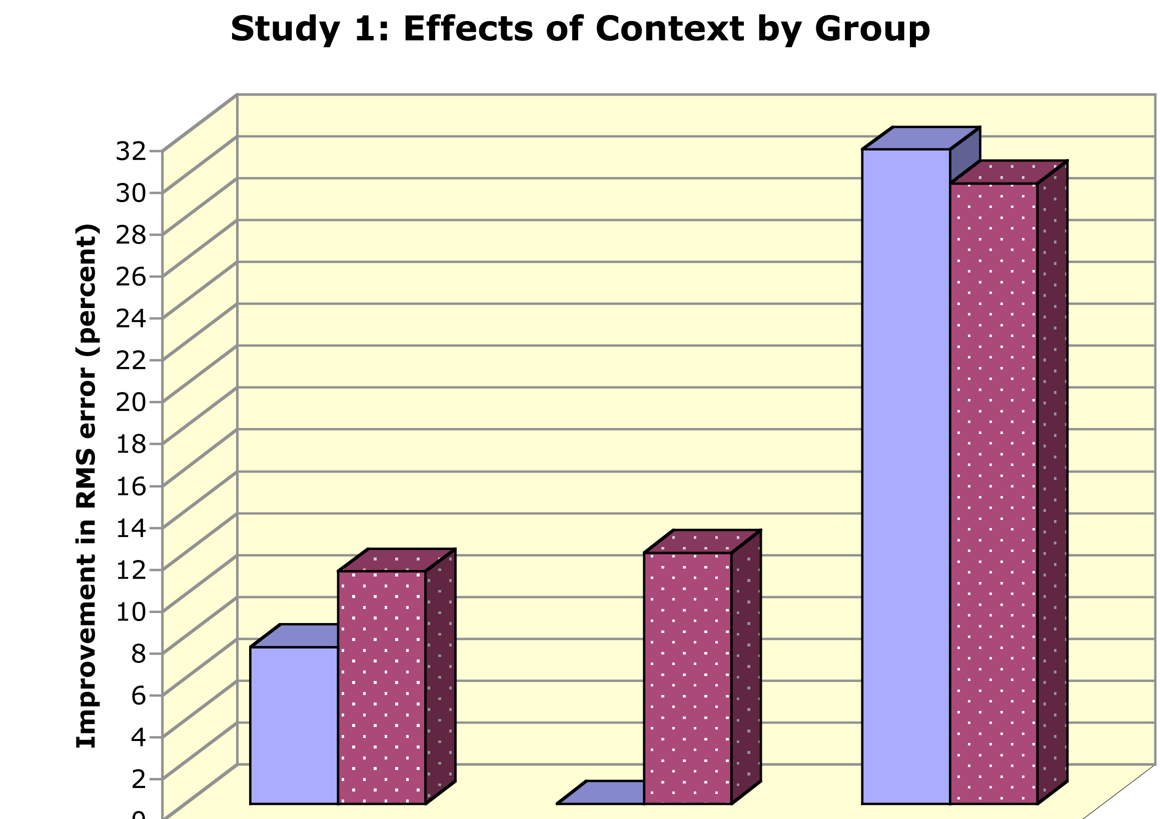
Next, in order to gain a different perspective on the data and a better understanding of the amount of improvement within and between groups, the percent improvement in RMS error from the pretest to the retest was calculated for each participant. Percent improvement was calculated by subtracting the retest score from the pretest score - and then dividing the result by the pretest score. These data were subjected to a univariate Analysis of Variance (ANOVA) with group membership as the between-subjects independent variable.

The results of this analysis are presented in Table 4 and Figure 4. There was a significant main effect of group membership, demonstrating that improvement in listener performance differed significantly depending on the contextual setting provided in the display ($F(5,160) = 5.20$, $p < .001$). The analysis revealed an explainable ordering of accuracy levels based on the contextual setting of each group (see Figure 2). In addition, simple planned contrasts between each group and the control group revealed that only Group 4 and 6 realized a significantly greater percent improvement in RMS error over the control group ($p = .003$ and $p = .005$ respectively). (For a full discussion the assumptions from each procedure see Appendix A.)

Table 4

Study 1 Descriptive Statistics by Group: Percent Improvement In RMS Error

Group	Type of context	Mean	SD	N
1	No context	7.50	18.15	26
2	Clicks	11.13	25.15	29
3	Static (Opening Price) Ref. Tone	-0.48	23.15	26
4	Dynamic (Max/Min Price) Ref. Tone	31.28	26.18	24
5	Clicks + Static (Opening Price) Ref. Tone	12.01	32.38	30
6	Clicks + Dynamic (Max/Min Price) Ref. Tone	29.66	37.46	25
Total		14.74	29.60	160



	No Ref Tone	Stat Ref Tone	Dyn Ref Tone
No Clicks	7.5	-0.48	31.28
Clicks	11.13	12.01	29.66

Figure 2. Study 1 effects of context on percent improvement in RMS error by group.

Discussion

The results of Study 1 largely replicate and clarify the theory and findings of Smith and Walker (2002). As in the earlier study, there were significant differences between groups and an explainable ordering of accuracy levels based on the differential benefits realized by each group's particular contextual setting.

To review, 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 (see Training, page 11). It was hypothesized that the addition of useful information (context) could assist users with these tasks and result in substantially improved performance.

For example, suppose the listener is asked to report the price of the stock at noon. 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 auditory graph, 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 tone also serve to reinforce the intended 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 (noon). This should also provide some assistance in the magnitude estimation task. Since the listener is no longer required to listen to the entire graph, he 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 graph. 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. Keeping this theory in mind (and the results of Smith and Walker (2002) (see Intentional Context, page 9), several predictions about the results of Study 1 were made.

First, with the understanding that the discrete sound mapping already provides x-axis context incidentally, a significant effect of adding clicks (more x-axis context) was not expected in Study 1. Nor was a significant interaction between added x and y-axis context. Surprisingly, although there was no interaction found, there was a significant effect of added x-axis context. This unexpected finding could be attributed to several factors. First, the findings of this study represent significantly added power from more participants and a more sophisticated statistical analysis than Smith and Walker (2002). In addition, the significant effect of the clicks calls into question if the context provided by the added clicks was *entirely* redundant (as hypothesized by Smith and Walker (2002) to explain the earlier lack of a significant result) – or if it somehow provided additional context to the x-axis after all. Given the significant decrease in RMS error for added x-axis context found in this study, it is possible that since the tones vary discretely on the hour and half hour, and the clicks occur only on the hour, that the added clicks had the effect of providing an auditory version of major and minor tick-marks to the graph; thus providing new and useful information, and therefore having a significant effect on performance (see Figure 3 below).

The typical examples shown were chosen because each illustrates a theorized effect of added context to the x-axis. For example, (as shown in the top left panel of Figure 3) subject 59's y-axis scaling already approximated that chosen for the display, but comparing his responses in the pretest to those in the retest (shown in the top right panel of Figure 3), one notices a small but distinct shift of the curve to the left. With the context provided by the x-axis

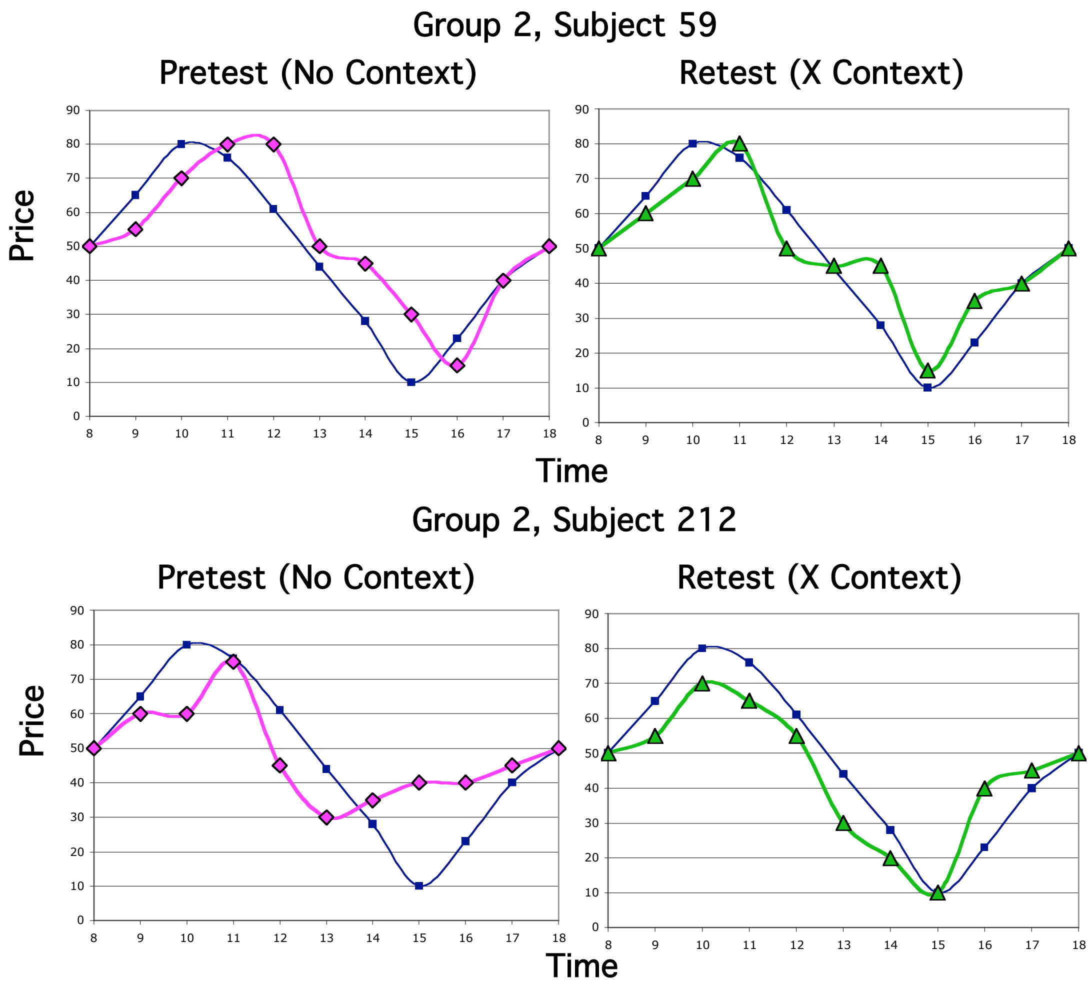


Figure3. The correct displayed values (blue squares) viewed with specific participant responses from the pretest (pink diamonds, left panels), and retest (green triangles, right panels) illustrate typical changes in performance observed between the discrete data mapping alone, and the discrete mapping plus the added context of x-axis clicks (right panels).

clicks, subject 59's y-axis scaling remains unchanged, but his responses still better approximate the correct values from the display (resulting in a decrease in RMS error from 11.14 to 8.10).

Alternatively, subject 212's pretest responses (shown in the bottom left panel of Figure 3) reveal that his individual y-axis scaling was not as close to that chosen for the display. But comparing his responses in the pretest to those in the retest (shown in the bottom right panel of

Figure 3), one notices not only that his responses better reflect the trends being displayed; but also that his y-axis scaling better approximates that chosen for the display. This illustrates the theorized effect that if added context helps the participant with the interval division task, cognitive resources (such as attention and working memory resources) would be freed to focus on the magnitude estimation task. Thus, although subject 212 is given no added y-axis context or scaling information, his performance shows not only an improvement in the interval division task, but also in the magnitude estimation task. Therefore his responses better approximate the actual correct values from the display (resulting in a decrease in RMS error from 13.95 to 9.20). These effects, although small, are found throughout the group and resulted in the statistically significant improvement for the group.

Next, we consider the significant main effect of added y-axis context. To clarify this finding, simple contrasts revealed significant effects of the dynamic (min/max-price) reference tone, but none for the static (opening-price) reference tone. These effects were verified by the analysis of percent improvement in RMS error, where only Group 4 and 6 showed significant improvements relative to the control group. These results confirm earlier findings pertaining to the performance effects of both types of y-axis context design features when employed with a discretely varying data sound. Moreover, this outcome provides further evidence to support the theory that added context enhances performance so long as it introduces new and useful information and does not interfere with, clutter, or distract from more useful information (D. R. Smith & Walker, 2002). By imposing a y-axis scale (information not provided by the static (opening-price) reference tone), the dynamic (min/max-price) reference tone provided more information, and therefore provided more assistance with the required perceptual and cognitive tasks involved in point estimation (see Figure 4 below).

The typical examples shown were chosen because each illustrates a theorized effect of each type of added y-axis context. For example, subject 5's individual y-axis scaling (shown

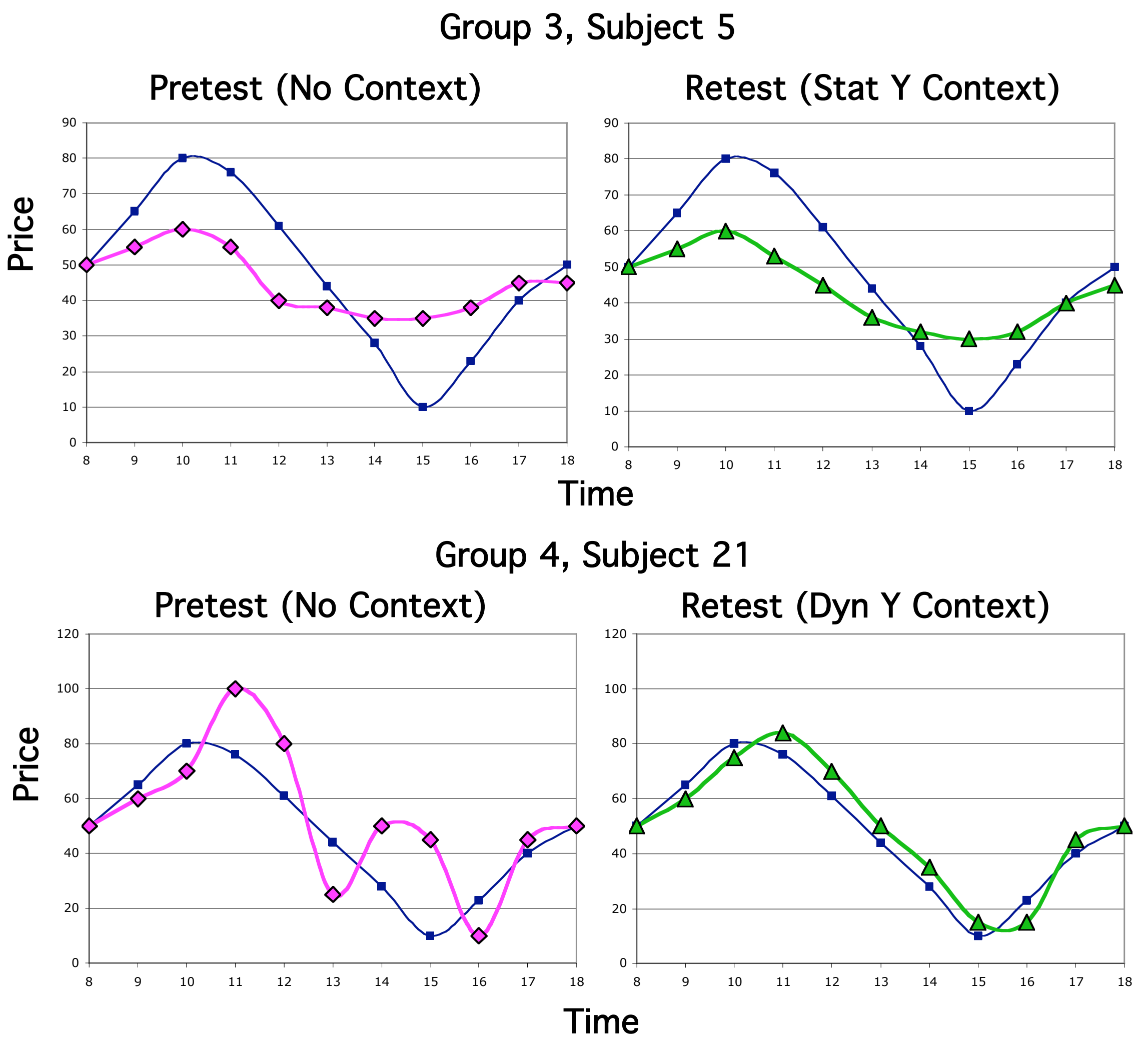


Figure4. The correct displayed values (blue squares) viewed with specific participant responses from the pretest (pink diamonds, left panels), and retest (green triangles, right panels) illustrate typical changes in performance observed between the discrete data mapping alone, and the discrete mapping plus added y-axis context (right panels).

in his pretest, top left panel, Figure 4) is significantly different from that chosen for the display. Then, when compared with his responses in the retest (top right panel, Figure 4), one notices that his responses change, vary more smoothly and reflect the displayed trend more accurately, but his individual scaling is unchanged. This makes sense because although the context provided by the static (opening-price) reference tone helps with some working memory tasks (as detailed above) and may help the listener judge the direction of change more readily, it does nothing to reinforce the chosen 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. For this reason, like many in his group, although subject 5's retest responses reflect a decrease in RMS error (from 14.67 to 13.04) the effect is not strong enough to cause a statistically significant shift in the mean.

Alternatively, subject 21's pretest responses (bottom left panel, Figure 4) also reveal that his individual y-axis scaling differs from that chosen for the display. However, comparing his responses in the pretest to those in the retest (bottom right panel, Figure 4), one notices not only that his responses become smoother and reflect the displayed trend more accurately, but also that his y-axis scaling becomes nearly exactly that chosen for the display. This is expected because the context provided by the dynamic (min/max-price) reference tone not only helps with the working memory tasks, it 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, subject 21's retest responses reflect a relatively dramatic decrease in RMS error (from 17.39 to 5.98). This larger effect, found throughout the group, resulted in the statistically significant finding for the dynamic reference tone.

Study 2

Study 2 employed the same task, and methods and procedures similar to Study 1, but varied in design to include a previously uninvestigated independent variable and a different stimulus set to investigate the following specific research questions: (1) How does the addition of contextual cues (including x-axis cues) affect performance when the data are represented via a continuously varying sound, as opposed to a series of discrete tones? (2) How does training affect performance in a point estimation task? and (3) Is there an interaction between the design features present in an auditory graph and amount of training such that the relative benefits of specific design features depend on the level of training provided; or that the efficacy of a training program depends on the design features chosen?

Participants and Apparatus

One hundred and fifty undergraduate students from the Georgia Institute of Technology and the United States Military Academy (USMA) participated for course credit (see Table 5). Administrative informed consent procedures and collection of demographic details were identical to Study 1.

Table 5

Study 2 Participant Demographics

			Handedness		Age		Years Of Music Experience	
		N	Left	Right	Mean	SD	Mean	SD
GT	Men	44	6 (14%)	38 (86%)	20.61	1.94	3.48	4.23
	Women	55	3 (5%)	52 (95%)	20.09	1.59	4.11	4.37
	Total	99	9 (9%)	90 (91%)	20.32	1.77	3.83	4.30
USMA	Men	38	3 (8%)	35 (92%)	19.97	3.89	3.8	3.58
	Women	13	2 (15%)	11 (85%)	19.08	1.04	6.27	3.35
	Total	51	5 (10%)	46 (90%)	19.75	3.41	4.43	3.65
Total		150	14 (9%)	136 (91%)	20.13	2.45	4.03	4.09

All reported normal or corrected to normal vision and hearing. Also, Study 2 apparatus and software were identical to that used in Study 1, with the exception of the added training programs, which were created and viewed in Macromedia Director 8.5 Shockwave Studio.

Design and Procedure

This study also employed a pretest-retest experimental design to investigate the effects of three between-subjects independent variables. As in Study 1, in the pretest untrained participants answered 11 point estimation questions about the information presented by an auditory graph. After random assignment, half the participants took part in a training phase and half did not. Next, listeners retested in one of eight experimental conditions (defined below).

Variables.

The dependent variable, RMS error, was identical to Study 1. The experimental conditions, however, were defined by the combination of two types of x-axis context, two types of y-axis context, and two levels of training (see Table 6 below).

Conditions.

In the pretest, all participants performed under the control condition, which was devoid of added context, relying entirely on the intuitions of the novice participant.

After the training phase, the participants retested. In the retest, Groups 1 and 5 experienced the graph under the same condition of no context as they did in the pretest. Groups 2 and 6 experienced the graph with the addition of x-axis context. This context was created by the insertion of audible clicks identical to those in Study 1. Groups 3 and 7 experienced the graph with the addition of y-axis context. This context was created via the addition of a dynamic (min/max-price) beeping reference tone, identical to that used for Groups 4 and 6 of Study 1.

Finally, Groups 4 and 8 experienced the graph with the combination of x-axis context (clicks), and y-axis context (the dynamic (min/max-price) reference tone).

Table 6

Study 2 Experimental Conditions

	No training (Filler Task)		Training	
	No added y-axis context	Dynamic ref. tone (reinforcing y-axis pitch to dollars scaling)	No added y-axis context	Dynamic ref. tone (reinforcing y-axis pitch to dollars scaling)
No added x-axis context	Group 1 (Control)	Group 3	Group 5	Group 7
Clicks (reinforcing x-axis time scaling)	Group 2	Group 4	Group 6	Group 8

Note. In every condition, participants are given the initial price of the stock at the opening of the trading day ($50).

Although there were only two levels of training (training and no training), this study required four different training programs - each differed in accordance with the type of added context present in the display. In the No Training condition, Groups 1 through 4 may have benefited somewhat from the familiarization they received as a result of participation in the pretest, but they received no formal training program whatsoever during the Training Phase. Instead, they completed a filler task consisting of three reading comprehension passages and accompanying reading comprehension 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 minutes dependent upon the participant), it required attention and comprehension of materials presented, and both the training programs and the filler task were interactive in nature (both required input and provided feedback). The primary difference between the filler task and the training programs was that the filler task provided no information, training, or practice in the perception or interpretation of auditory graphs.

In contrast, Groups 5 through 8 took part in training programs. Each program was 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; D. R. Smith & Walker, 2002). The nature of each supporting task was explained, and strategies practiced for the completion of the task in the presence or absence of each of the applicable design features for that group.

For Groups 5 through 8, the Training Phase varied as follows. Due to the fact that the auditory graph for Group 5 would have no added context, the training program for Group 5 focused mainly on the perceptual and cognitive skills used in interval division and magnitude estimation. These include 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 that discussed for Group 5, the training program for Group 6 included training on the nature and use of the added clicks to assist the user in the interval division task. In addition to the training discussed for Group 5, the training program for Group 7 included training on the nature and use of the dynamic reference tone to assist the listener in the magnitude estimation task. Finally, the training program for Group 8 incorporated all the training described above. (See Appendix B for screen-shots illustrating and contrasting the training programs and filler task.)

After this training phase, all groups started the retest by receiving a short explanation of their new display. Next, each group experienced the graph under the appropriate experimental condition of added context. Then, upon completion of the retest, participants were thoroughly debriefed, thanked, and released. (See task view screen-shots at Appendix B).

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). The display was identical to that used in Study 1, with the exception that the pitch variations of the display sound were continuous as opposed to discrete.

Results

Individual scores on RMS error were calculated as for Study 1 and subjected to the following statistical analysis. First, a three-way, univariate, ANCOVA was conducted on retest scores, using Type of x and y-axis context, and Level of training as the between-subjects independent variables; and pretest score as the covariate.

The results of the analysis are presented in Table 7 and Figure 5. There was a significant main effect of training, reflecting that listeners answered with smaller RMS errors if they received training ($F(1,150) = 10.405$, $p = .002$). There was also a significant main effect of type of y-axis context, demonstrating that listeners answered with smaller RMS errors if they had the dynamic (min/max-price) y-axis reference tone ($F(1,150) = 4.92$, $p = .028$). Finally, there was a significant training by y-axis context interaction reflecting that the effect of the contextual design feature was dependant upon level of training ($F(1,150) = 4.774$, $p = .031$). This interaction, combined with post-hoc comparisons between each group that received training and each of the other groups receiving training that were not significant, demonstrate that when groups received training, adding y-axis context did not produce significant shifts in the means. Where, if groups did not receive training, adding y-axis context did (see Figure 5 below).

Table 7

Study 2 Descriptive Statistics: RMS Error

Level of training	Type of x-axis context	Type of y-axis context	Mean	SD	N
No training (filler task)	No added clicks	No added Ref. Tone	23.62	28.71	17
		Dynamic Ref. Tone	15.23	6.17	18
		Total	19.30	20.61	35
	Clicks (reinforcing x-axis time scaling)	No added Ref. Tone	21.96	34.05	18
		Dynamic Ref. Tone	11.03	5.34	19
		Total	16.35	24.34	37
	Total	No added Ref. Tone	22.77	11.12	35
		Dynamic Ref. Tone	13.07	6.06	37
		Total	17.78	22.50	72
Training	No added clicks	No added Ref. Tone	11.03	4.94	19
		Dynamic Ref. Tone	10.40	4.67	20
		Total	10.71	4.74	39
	Clicks (reinforcing x-axis time scaling)	No added Ref. Tone	11.27	10.55	20
		Dynamic Ref. Tone	8.28	2.92	19
		Total	9.81	7.87	39
	Total	No added Ref. Tone	11.15	8.19	39
		Dynamic Ref. Tone	9.37	4.01	39
		Total	10.26	6.47	78
Total	No added clicks	No added Ref. Tone	16.98	20.73	36
		Dynamic Ref. Tone	12.69	5.89	38
		Total	14.77	15.11	74
	Clicks (reinforcing x-axis time scaling)	No added Ref. Tone	16.33	24.88	38
		Dynamic Ref. Tone	9.65	4.47	38
		Total	12.99	18.07	76
	Total	No added Ref. Tone	16.64	22.80	74
		Dynamic Ref. Tone	11.17	5.41	76
		Total	13.87	16.64	150

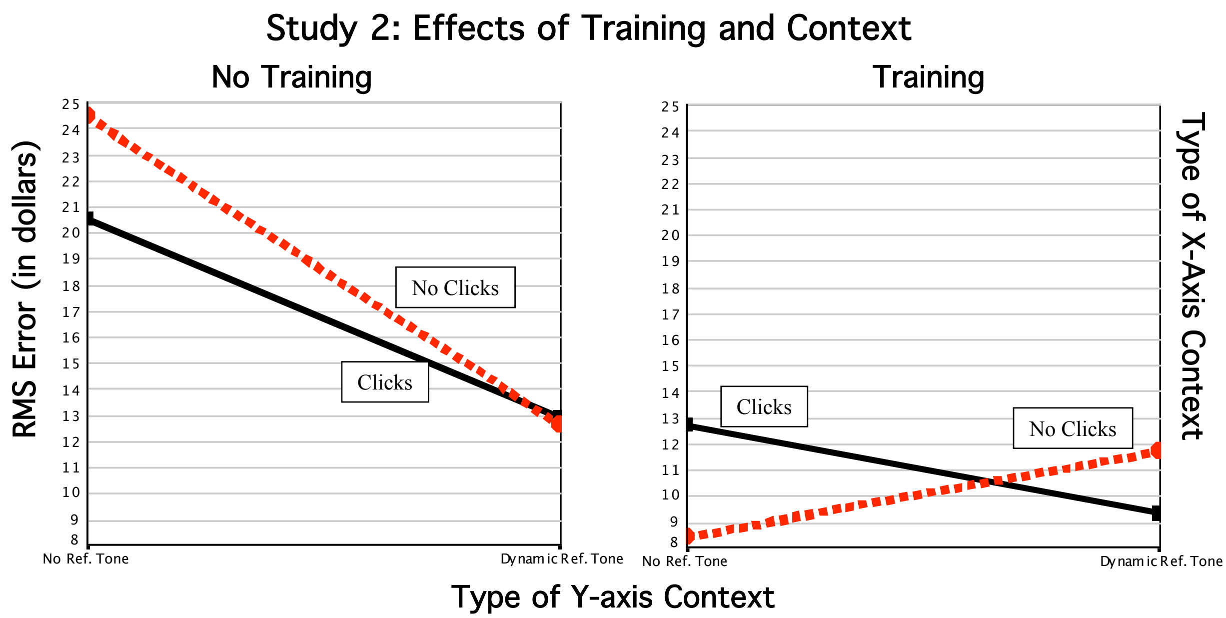


Figure 5. Study 2 effects of training and context on RMS error (in dollars).

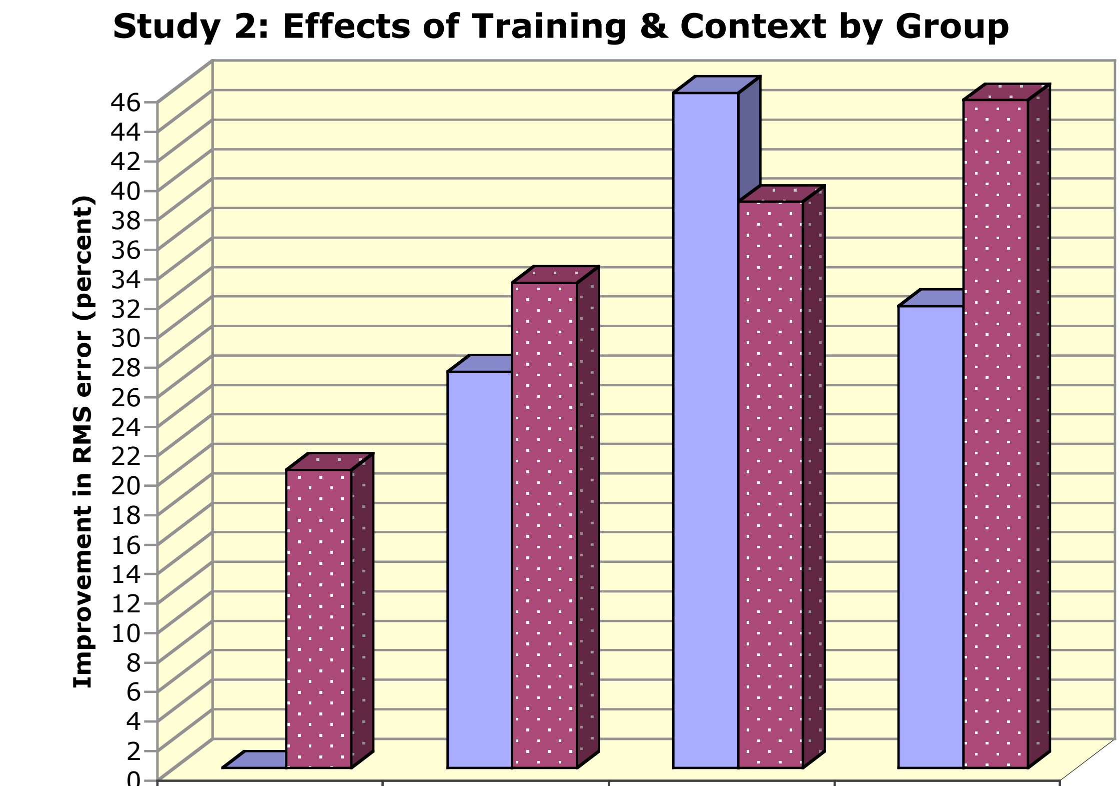
Next, in order to facilitate comparison with Study 1, and gain a better understanding of the amount of improvement within and between groups, the percent improvement in RMS error from the pretest to the retest was calculated for each participant. These data were subjected to a univariate ANOVA with group membership as the between-subjects independent variable (see Table 8 and Figure 6).

There was a significant main effect of group membership, demonstrating that improvement in listener performance differed significantly depending on the level of training and the contextual setting provided in the display ($F(7,150) = 2.96$, $p = .006$). The analysis revealed an ordering of accuracy levels based on the level of training and contextual setting of each group. In addition, simple planned contrasts between each group and the control group revealed that of the groups that did not receive training, only Groups 3 and 4 realized a significantly greater percent improvement in RMS error over the control group (p=.031 and p=.008 respectively). But of the groups that did receive training, every group realized a significantly greater improvement over the control group (Group 5: $p < .001$; Group 6: $p = .002$; Group 7: $p = .010$; and Group 8: $p < .001$). Again, post-hoc comparisons between each group that received training and each of the other groups receiving training, were not significant, demonstrating that when groups received training, adding context did not produce significant shifts in the means (see Figure 6 below). (For a full discussion the assumptions from each procedure see Appendix A.)

Table 8

Study 2 Descriptive Statistics by Group: Percent Improvement In RMS Error

Group	Training	Context	Mean	SD	N
1	No Training	No context	-0.29	54.78	17
2	No Training	Clicks	20.21	22.75	18
3	No Training	Dynamic (Max/Min Price) Ref. Tone	26.87	59.39	18
4	No Training	Clicks + Dynamic (Max/Min Price) Ref. Tone	32.90	37.82	19
5	Training	No context	45.81	22.65	19
6	Training	Clicks	38.41	28.00	20
7	Training	Dynamic (Max/Min Price) Ref. Tone	31.35	29.87	20
8	Training	Clicks + Dynamic (Max/Min Price) Ref. Tone	45.31	23.91	19
Total			30.63	38.61	150



	No Ref Tone	Dyn Ref Tone	Tng/No Ref Tone	Tng/Dyn Ref Tone
No Clicks	-0.29	26.87	45.81	31.35
Clicks	20.21	32.9	38.41	45.31

Figure6. Study 2 effects of training and context on percent improvement RMS error by group.

Discussion

The results of Study 2 confirm theoretical predictions expecting significant main effects of at least two of the three independent variables (training and y-axis context); as well as a significant interaction between amount of training and at least one type of added context (y-axis).

The lack of a significant effect of adding x-axis context, however, is especially noteworthy here - especially in light of the significant effect found in Study 1. It appears that in Study 2, the effect of adding x-axis context is inconsistent. This is difficult to explain. Theoretically, we would expect that anything adding new and useful information would produce improvements in performance (D. R. Smith & Walker, 2002; Tufte, 2001; Walker, 2002). Why then would added x-axis context seem to work/help as expected in Study 1, and also in Study 2 (but only when there was no reference tone or training), and then have unpredictable and inconsistent effects when training and/or the reference tone is added (see right panel, Figure 5, page 36)?

Several factors cloud this issue. First, it is important to note that adding the clicks in this study (Study 2), was not the same as adding them in Study 1. In Study 1, the discrete nature of the sound mapping already provided some incidental x-axis context. Therefore, adding the clicks in Study 1 wasn't so much adding x-axis context, as it was adding *more* x-axis context (creating the effect of major and minor tick marks). In this study (Study 2), the continuous nature of the sound mapping provided no x-axis context; and 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. But these results, viewed together with those of Study 1, seem to say 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 does.

It may be that the effect of adding *more* x-axis context is not the same as adding it where none formerly existed. Theoretically, we would expect positive results wherever new and useful information has been added, and not cluttered or detracted from the perception of more useful information (D. R. Smith & Walker, 2002; Tufte, 2001; Walker, 2002). It is possible that adding clicks where no context exists, although a seemingly larger increment of added information may not be a larger increment of added usefulness. Perhaps creating the effect of major and minor tick marks (where before only major tick marks exist) adds a greater increment of usefulness over the addition of major tick marks only (where previously no context exists). Perhaps the clicks themselves are not enough to assist the listener in some conditions. Added clicks may only make a significant impact when the sound mapping itself already provides some x-axis context. It is also possible that the inconsistent effects of adding clicks in the presence of training could be explained by clutter. Perhaps such context is added useful information to an untrained participant, yet merely clutter to a trained user. It seems that, for the time being, the question concerning the effects of x-axis context will remain unanswered. Regardless, it again points out the need for empirically tested design principles and guidelines to take the place of quasi-heuristics and intuitive design. Yet again, the effect of a rather widely accepted design technique is not clearly what one might expect (see also Walker and Kramer (1996)).

To understand how training influences performance and the effects of added context it is important to contrast the left and right panels of Figure 5, page 36 (No Training and Training, respectively). Although it is clear that training itself is beneficial to performance it is difficult to discern how these variables affect each other. 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 especially, imparts knowledge of both x and y-axis scaling. This information is not available to the untrained participant. As a result, in the untrained condition, we witness rather clearly the familiar effects of added y-axis context. This is because the context is the only means by which scaling information is communicated. In the trained conditions though, we find relatively little differences between contextual settings; and what differences do exist are difficult to explain. This is partially because of floor effects, but also due to the fact that where training is involved, it is difficult to distinguish the positive results of training in required skills, from the contextual information imparted by knowledge of results gained in practice, from whatever effects might have been obtained by the addition of context.

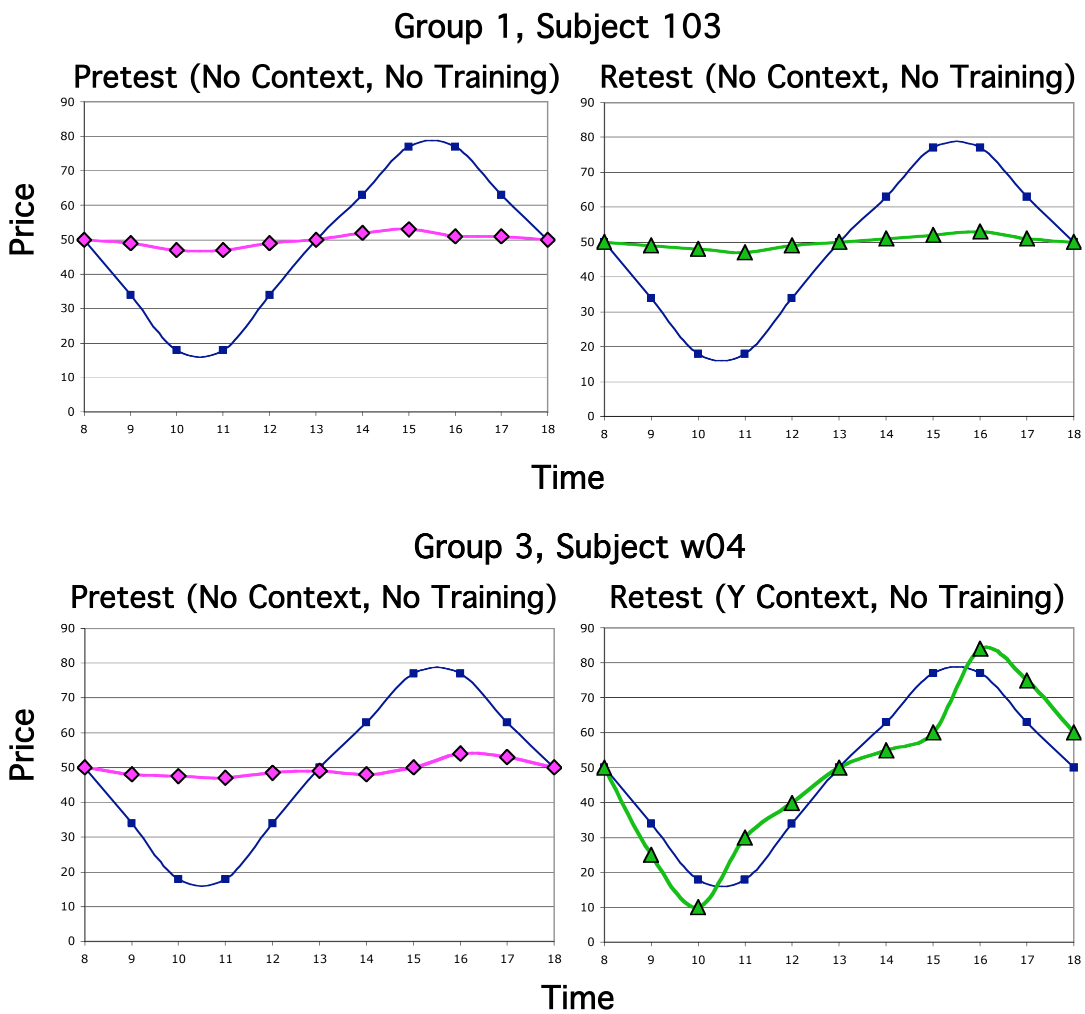


Figure7. The correct display values (blue squares) viewed with specific participant responses from the pretest (pink diamonds, left panels), and retest (green triangles, right panels) illustrate and contrast the typical performance changes observed in the control group (No Context/No Training), versus that observed in Group 3 (Y Context/No Training).

For example, in Figure 7 one can easily see the effect typically found with the addition of y-axis context. As we saw in Study 1, the dynamic (min/max-price) reference tone (bottom right panel) results in improved performance by not only helping with the working memory task, but by reinforcing the preferred y-axis scaling.

However, when training is added there seems to be no effect of added contextual features beyond that of training alone (see Figure 8 below).

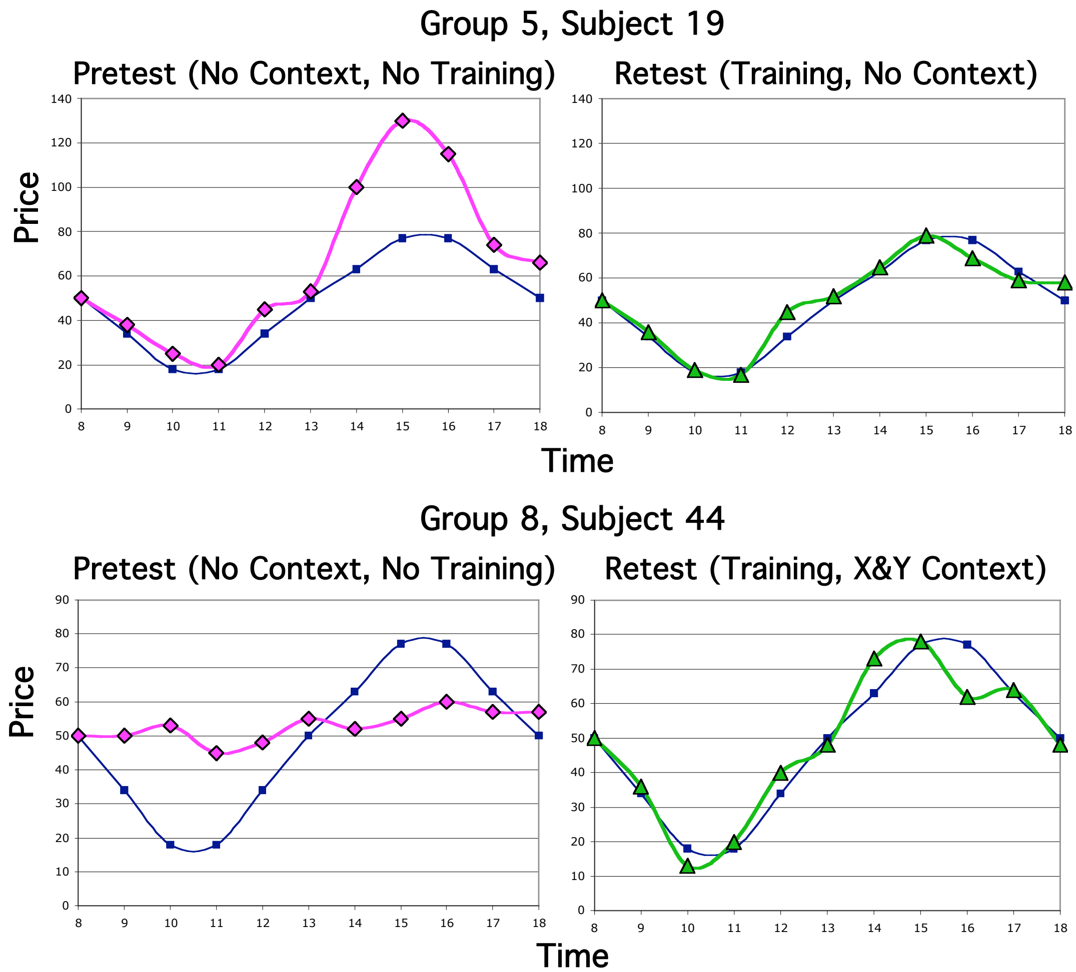


Figure8. The correct display values (blue squares) viewed with specific participants responses from the pretest (pink diamonds, left panels), and retest (green triangles, right panels) illustrate and contrast the typical changes in performance observed in Group 5 (Training/No Context) versus that observed in Group 8 (Training/X&Y Context).

In either condition (top or bottom), the retest (shown in the right panels) reveals similar levels of performance regardless of the presence of context. Viewing these performance graphs (each typical of its group), one is apt to conclude that both context and training produce the same performance effects and ultimately yield the same level of performance. Furthermore, it seems that although context affects the performance of untrained users, it does not produce significant differences in the performance of trained users. This may be because training and context provide partially redundant information - or it may be that the particular training programs in this study were not effective enough in enhancing participants' abilities to use the context provided.

To corroborate these findings, the simple planned contrasts revealed significant performance differences between specified groups and the control group. These differences are explainable in terms of the amount of information provided by a particular contextual setting, and in terms of the increased skill proficiency of trained participants over untrained participants in the perceptual and cognitive tasks involved in the execution of a point estimation task. This is witnessed by the fact that in the untrained condition only Groups 3 and 4 (both having y-axis context) showed differences from the control group. Whereas, all the trained groups showed significant differences from control, but no significant differences from any other trained group.

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 programs focused on the operative cognitive and perceptual skills used in the perception of a sonification are an effective means of increasing performance of the human-sonification subsystem; (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 alter the effectiveness of training in several of the specific tasks required in the performance of a point estimation task.

Conclusions and Further Research

In addition to the individual analysis above, it is also helpful to combine, compare and contrast results across the two studies. Taken together, the results support 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 information, or improve performance. Training is also an important factor in performance in point estimation sonification tasks.

An examination of the group means from both studies (interleaved in the bottom of Figure 9) reveals a steadily decreasing trend in error (and smaller standard deviations in error) from left to right as each succeeding condition provides users with 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.

Looking at the categories, one notices several distinct consistencies. The group with the highest error was the control group, Group 1 from Study 2 (2-1 None). This group had no training or context; and the continuous data mapping deprived them of even incidental context. Next, notice that Group 1 of Study 1 (1-1 Disc X), which had incidental x-axis context derived from the discrete data mapping in Study 1, and Group 2 of Study 2 (2-2 Clicks) have lower (and visibly similar) means than '2-1 None'. It is important to note discrepancies between the standard deviations of these two groups, but one might also conclude that the incidental x-axis context provided by the discrete data mapping, and the intentional x-axis context provided by the clicks (both creating the effect of major tick marks alone) had visibly similar effects. This same consistency is again observed between Group 4 of Study 1 (1-4 Disc X/Dyn Y) and Group 4 of Study 2 (2-4 Clicks/Dyn Y).

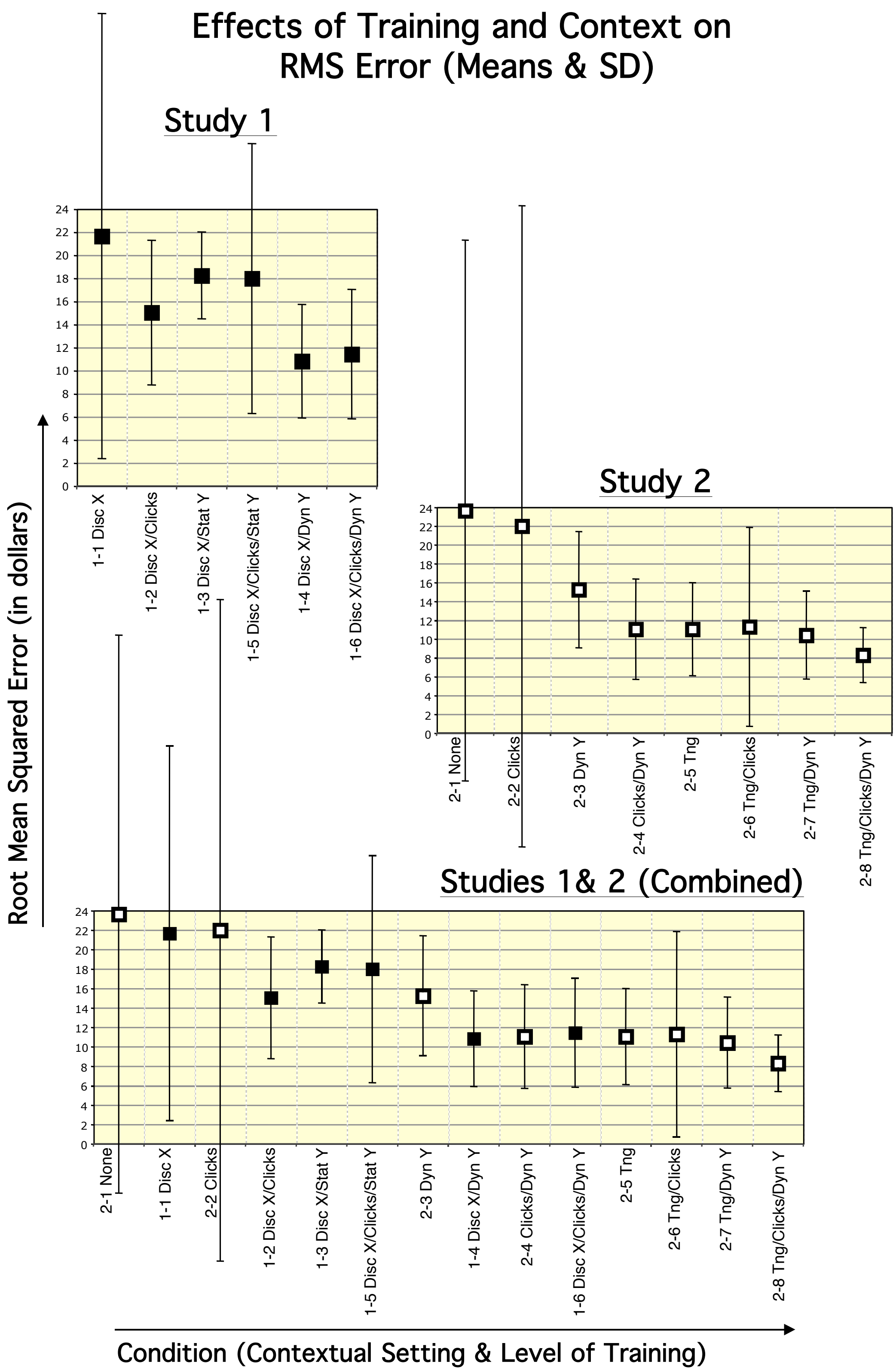


Figure9. Mean RMS error and SD for each study; also shown combined, in order of amount of additional useful information provided from least to greatest (left to right). Study and group numbers are depicted with study number first (e.g., study # - group #).

Next, compare the last two groups discussed to Group 3 of Study 2 (2-3 Dyn Y). In '2-3 Dyn Y', the only context is the dynamic (min/max-price) reference tone. But moving to the right, we again see that the additional x-axis context provided incidentally by the discrete data mapping of Study 1, and the intentional context from the added clicks in Study 2, have had visibly similar effects on the mean, as well as on the standard deviations.

Another interesting consistency deals with y-axis context. It is interesting that the groups having either type of x-axis context combined with the static (opening-price) reference tone ('1-3 Disc X/Stat Y' and '1-4 Disc X/Clicks/Stat Y'), have lower and visibly similar means in comparison to the groups with the x-axis context alone ('1-1 Disc X' and '2-2 Clicks'). Further, groups having either type of x-axis context with the dynamic (min/max-price) reference tone ('1-5 Disc X/Dyn Y' and '2-4 Clicks/Dyn Y'), also have lower and visibly similar means.

It is important to note that there are also interesting inconsistencies between studies. For example, in '1-2 Disc X/Clicks', it appears that when both types of x-axis context are combined, they result in a distinct drop in the mean in comparison to no context (2-1 None), as well as in comparison to discrete x-axis context or clicks ('1-1 Disc X' or '2-2 Clicks' respectively). But this effect is not seen when either type of y-axis context is present (see '1-4 Disc X/Clicks/Stat Y' and '1-6 Disc X/Clicks/Dyn Y' in comparison to their counterparts).

Another unanswered question pertains to a phenomenon hypothesized by Smith and Walker (2002) - that the continued addition of context could, at some point, begin to interfere with, clutter, or distract from more useful contextual information or from the data itself. The idea of clutter or "chart junk" is not new (Tufte, 1990). But it has yet to be conclusively demonstrated that the same principles will apply in the same manner to auditory graphs. Looking at the categories representing the addition of training ('2-5 Tng' to '2-8 Tng/Clicks/Dyn Y'), one might conclude that the information added and skills acquired in training were sufficient for the observed improvement in performance – and that additional context in those conditions caused no additional change in performance because it was redundant (merely cluttering the display). This may be true, but it is also possible that floor effects merely hid the effects of the context. It is possible 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 efficacy of any decrease in error can only be evaluated relative to the magnitudes and variations in the data being displayed. For example, Group 8 of Study 2 (2-8 Tng/Clicks/Dyn Y) showed a decrease in mean RMS error to $8.28. This is nearly a 50% decrease in error in comparison to its control. 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 both studies of this project, 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 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, and x and y-axis context, the error dropped to about 16% of the average price of the stock. But to truly evaluate the usability of such a graph would require replication and applied study with specific users, tasks, and systems. I believe this approach is essential to continued progress and improvement of auditory graphs, and it should be applied to each innovation and new idea, as well as to evaluate the usability of the resulting auditory graphs.

The purpose of this research was not to advocate any specific type of added context or training program. It was merely to take the next logical steps in the effort to verify factors governing performance with auditory graphs, describe their effects, and apply this knowledge in the innovation of potential solutions for improving performance; in addition, to acknowledge the user, task, display, and system specific demands that any sonification training program might incorporate.

This investigation tested theories of adding context and employment of training to alter human performance in a point estimation sonification task. The findings of this and future studies will further help us understand how to better design, implement, and employ auditory graphs and sonifications in ways that enhance system performance, give users more flexibility, and allow the improvement of system interfaces wherever sonification is appropriate.

Future research might seek to investigate the effects of this or other training programs in conjunction with other design techniques (such as other various ways to create context). Future researchers may also seek to investigate the effects of performance strategies other than a subtask breakdown approach. It is possible that the proposed strategy does not reflect the most efficient or accurate way to perceive an auditory graph. It may be the case that training users (or some users) to use a more holistic approach, or some other strategy could yield greater improvements in performance.

Finally future researchers might investigate individual factors, traits, or differences that govern performance among and between individual listeners - or affect how they perceive and use auditory graphs, added context, and/ training.

Appendix A: Testing Statistical Assumptions

It is important to test the assumptions of chosen statistical procedures and understand the impact of violations of any of the assumptions on the results. Each study in the project made use of a one-way analysis of variance (ANOVA) with group membership as the between subjects independent variable. Each study also made use of a two-way (in Study 1) or three-way (in Study 2) analysis of covariance (ANCOVA) procedure, with type of x and y axis context and (in Study 2) level of training as the between-subjects independent variables. Both ANCOVA procedures used pretest score as the covariate.

Presented here will be an analysis and discussion each of the assumptions of the ANOVA procedure (addressing each data set in order, Study 1 followed by Study 2). Next will be an analysis and discussion of each of the three additional assumptions of the ANCOVA procedure (again addressing each data set in order, Study 1 followed by Study 2).

Analysis of Variance

An ANOVA procedure assumes: (1) Independent Observations; (2) Normality of the distribution of each group on the dependent variable; and (3) Homogeneity (or equality) of variance within the distribution of each group on the dependent variable.

In both Study 1 and Study 2, independence of observations was achieved through random assignment of participants to conditions, and by assuring no interaction among participants during participation. In addition, for participant selection both studies relied on volunteer participation of undergraduates taking introductory psychology courses. This procedure is accepted as assuring a reasonably good cross-section, which is representative of the various groups and classifications possible in the available population.

To test the normality assumption we examined the histograms from each study. In Study 1, the group histograms reveal various violations of normality (see Figure A1). Each distribution shows some positive skew, and/or leptokurtosis or platykurtosis to one degree or another, but none could be strictly described as normal.

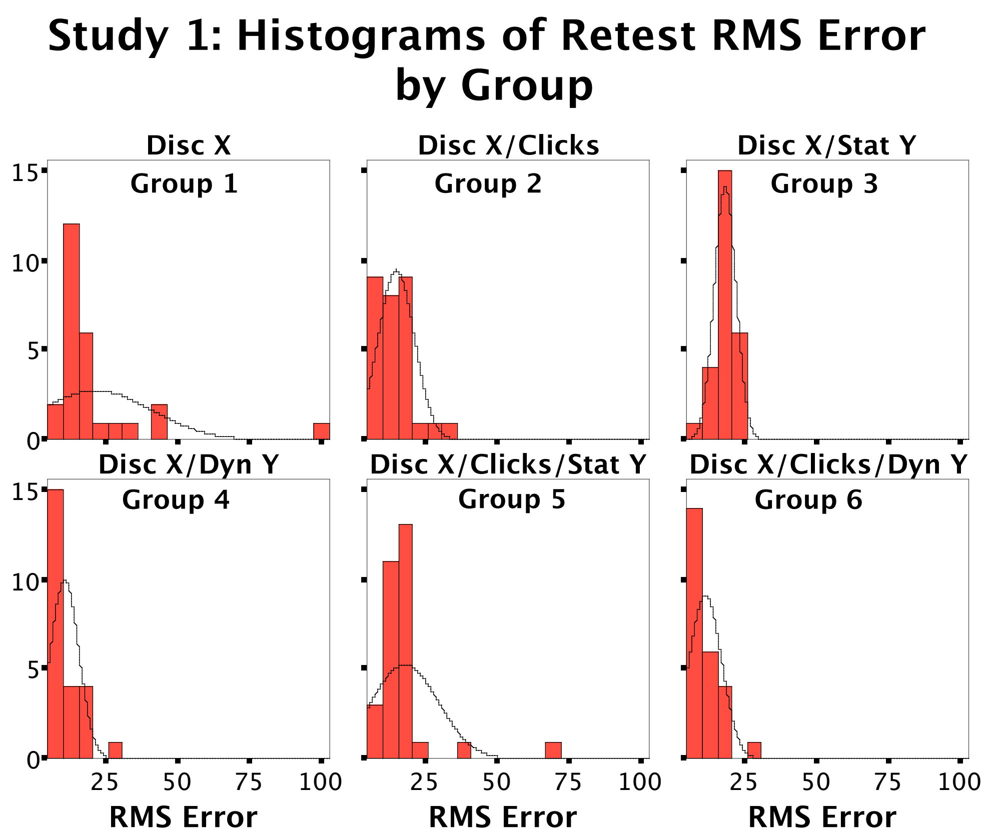


Figure A1. Histograms with normal curves for each group of Study 1 reveal various violations of the normality assumption

In Study 2, the group histograms reveal similar violations of normality. Again, each distribution shows some positive skew, and/or leptokurtosis or platykurtosis to one degree or another, but none could be strictly described as normal (see Figure A2 below).

In general, the chosen procedures are robust when the assumption of normality is violated. In specific, they are robust against type I error. The skewness has little effect, and

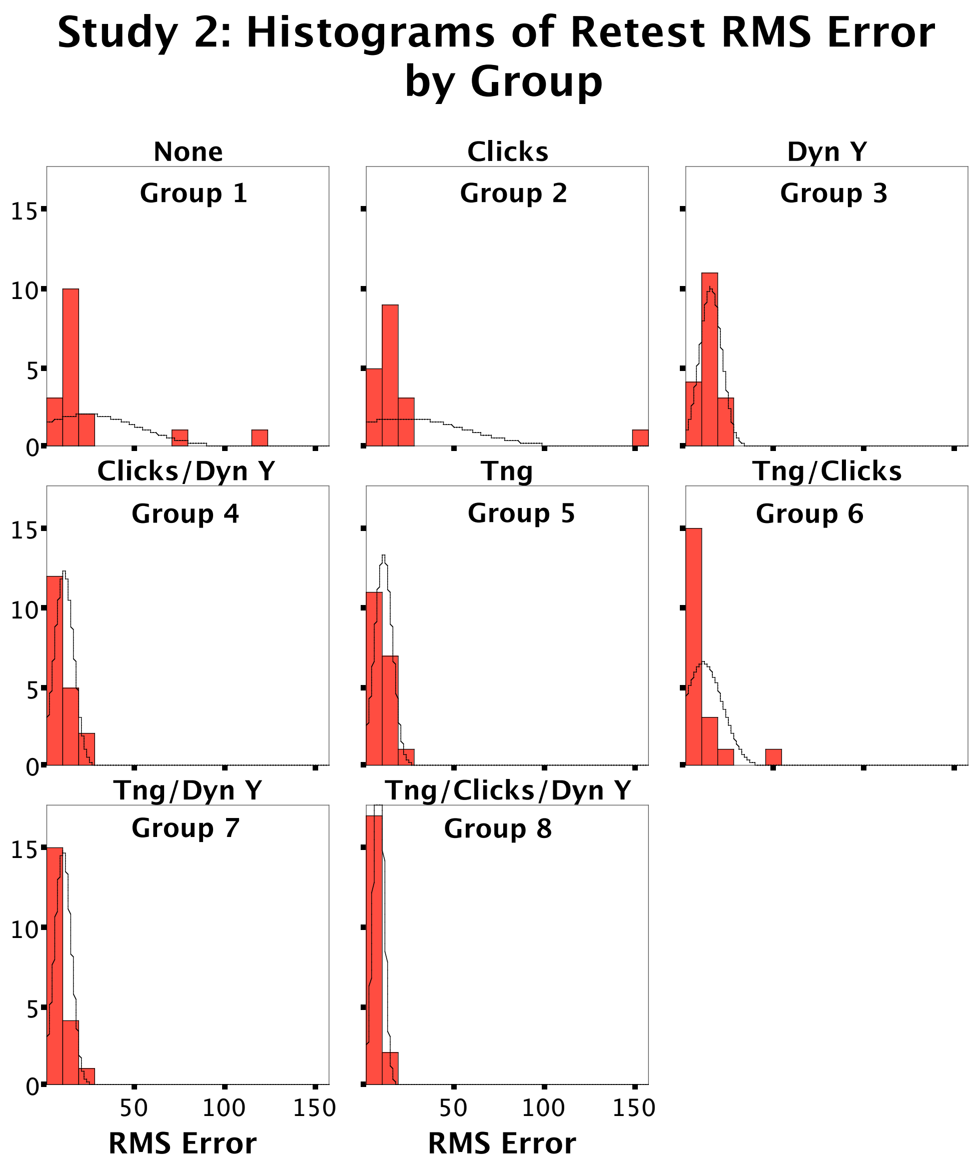


Figure A2. Histograms with normal curves for each group of Study 2 reveal various violations of the normality assumption

platykurtosis merely attenuates power making the procedures more conservative and robust against type I error.

The third and final assumption of the ANOVA procedure is homogeneity (or equality) of variance. To test for homogeneity of variance the data from the groups in each study was subjected to Levene's test of equality of error variances. In Study 1, Levene's test was not significant (p=.116), meaning that the group data from Study 1 did not violate the assumption. In Study 2, Levene's test was also not significant (p=.056), meaning that the group data from Study 2 also did not violate the assumption (see Figure 12). One might note that Levene's test is

Tests of Homogeneity of Variance

Study 1

Levene's Test of Equality of Error Variances

Dependent Variable: RMS Error 2

F	df1	df2	Sig.
1.799	5	154	.116

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Study 2

Levene's Test of Equality of Error Variances

Dependent Variable: RMS Error 2

F	df1	df2	Sig.
2.027	7	142	.056

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Figure A3. Levene's tests for equality of error variances were not significant, verifying that homogeneity of variance assumptions hold for both Study 1 (left) and Study 2 (right)

very nearly significant in Study 2. Taking note of this, it might be necessary to point out that our chosen procedures are robust (if there had been a violation), as long as group sizes are approximately equal. In this context "approximately equal" means that the number of participants in the largest group divided by the smallest group, is less than 1.5. This is condition is met by both data sets (Study 1 and Study 2).

Analysis of Covariance

In addition to the assumptions checked above, ANCOVA assumes: (1) That the covariate is measured without error; (2) A linear relationship between the dependent variable and the covariate; (3) Homogeneity of regression slopes between groups.

In both Study 1 and Study 2, the covariate (RMS error score on the pretest), like the dependent variable, were calculated from each individual participant's responses. The participants were under no time pressure and, in addition, data collection was automated. These techniques were judged as providing a strong degree of reliability that the individual responses analyzed were the actual responses each individual made. Thus, in both studies, the covariate has been measured with relatively high, and readily acceptable levels of accuracy.

In order to test the next two assumptions, pretest scores from each group were regressed on retest scores (see Figures A4 and A5). In Study 1, regression results reveal significant linear relationships between the dependent variable (retest score) and the covariate (pretest score). This verifies the linear relationship assumption held true for each group in Study 1 (see Figure A4).

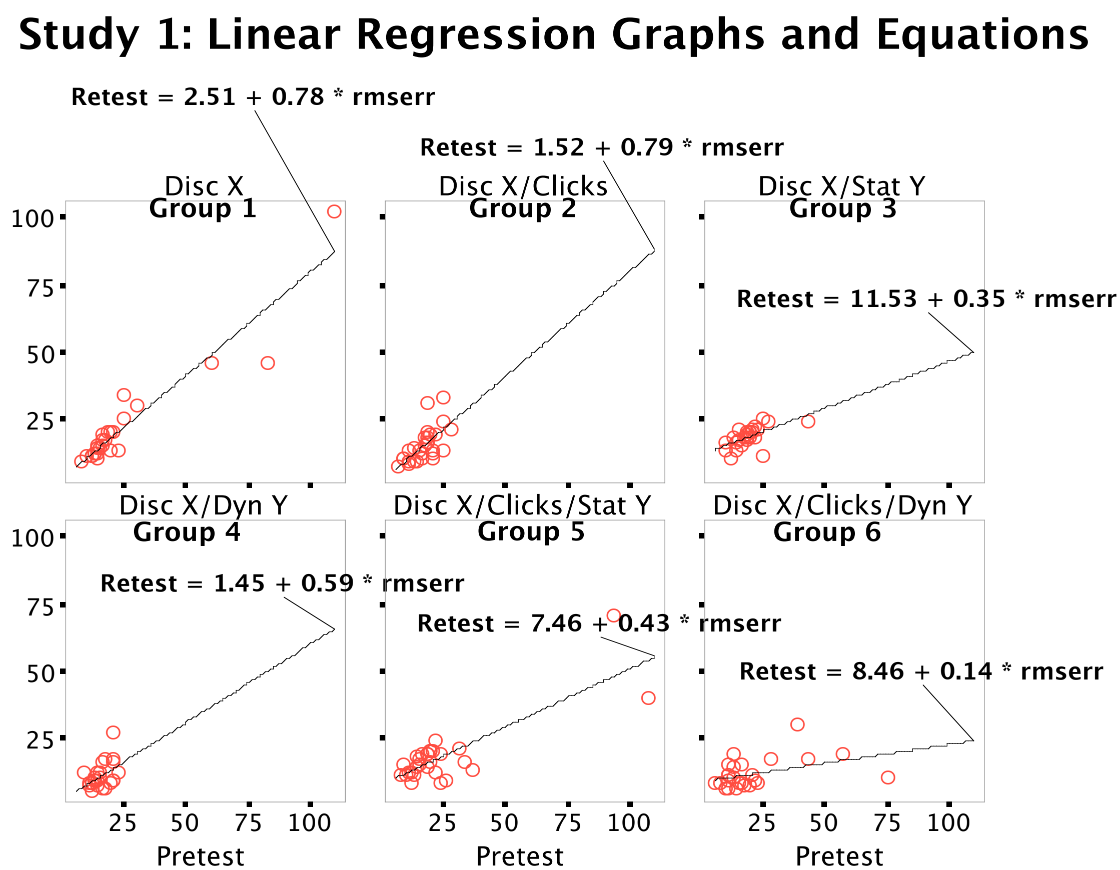


Figure A4. Linear regression results for each group of Study 1 reveal significant linear trends, but some groups violate of the assumption of homogeneity of regression slopes.

But contrasting the various graphs and equations reveals differences in group regression slopes in some cases – this is a violation of the assumption of homogeneity of regression slopes.

In Study 2, regression results also revealed significant linear relationships between the dependent variable (retest score) and the covariate (pretest score). This verifies that the linear relationship assumption also held true for each group in Study 2 (see Figure A5). But again,

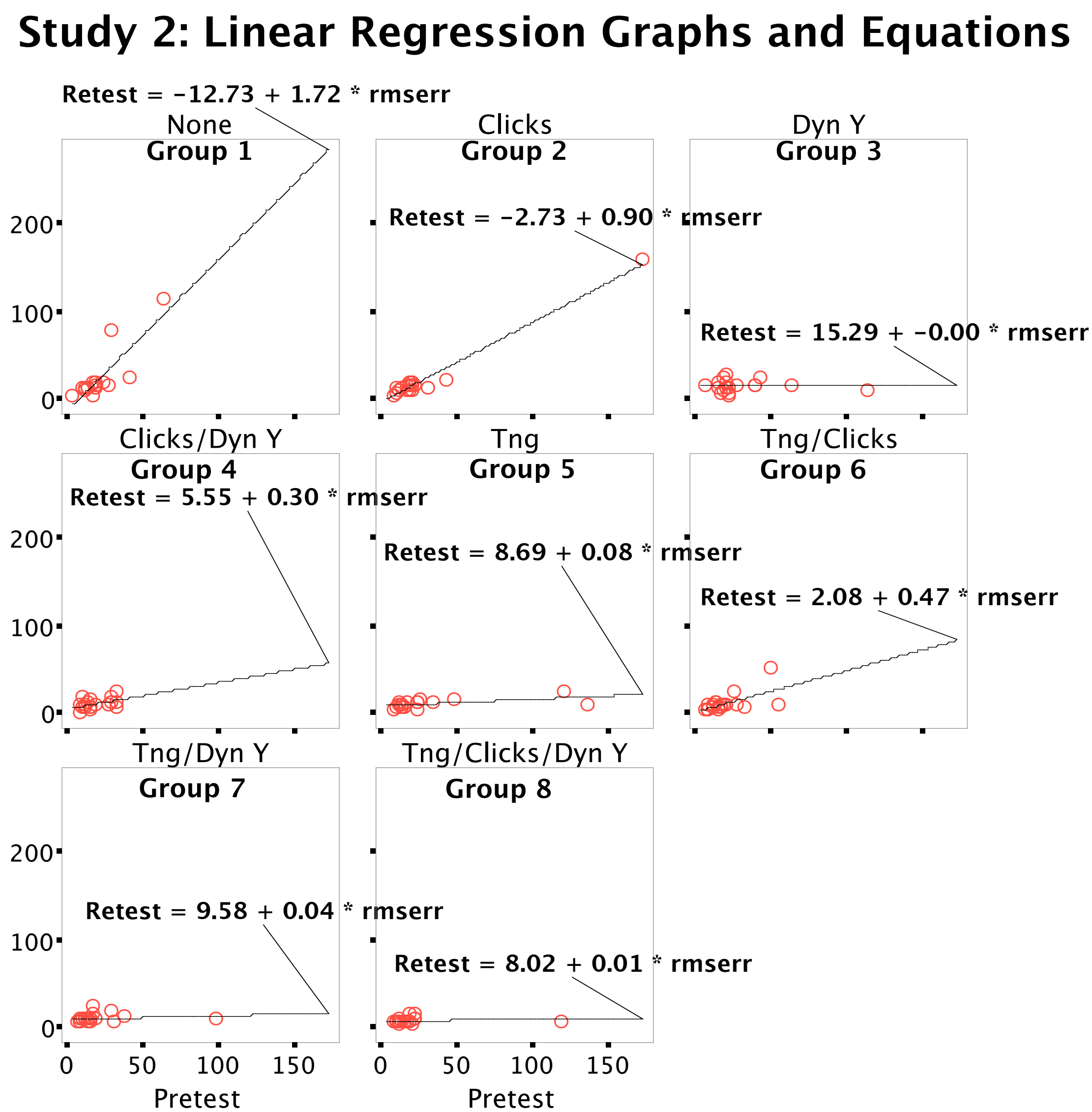


Figure A5. Linear regression results for each group of Study 2 reveal significant linear trends, but some groups violate of the assumption of homogeneity of regression slopes.

contrasting the various graphs and equations reveals differences in individual group regression slopes in some cases - another violation of the assumption of homogeneity of regression slopes.

Violations "of *all three* of the remaining ANCOVA assumptions" are serious, but violating an individual assumption of ANCOVA does not seriously affect results (Stevens, 2002). In both studies, two of the three assumptions are met. Further, in this case, the violation of homogeneity of regression slopes was in such a way that individual differences between groups may be overestimated in a limited number of cases, and underestimated in a larger number of cases (Stevens, 2002). For this reason, it is unclear in this situation whether the result would be to make the procedure more conservative, or if it would have no effect (Stevens, 2002).

It would appear that, in both studies, there is a treatment by covariate interaction such that individuals who answer with large errors in the pre-test are affected differently by the presence of training and context than are individuals who answer with small errors in the pretest. But this appearance is more likely created by floor effects, rather than being a true interaction (e.g., someone who answers with large error can improve a great deal in the retest, whereas someone who answers with small errors cannot answer very much better in the retest). These issues, in combination with the shortcomings of alternative procedures, prompted the decision to use ANCOVA - in combination with ANOVA conducted on difference or gain scores (Stevens, 2002).

Appendix B: Instructions and Materials

The following screen shots illustrate the instructions and materials participants saw as they executed the experimental task in the pretest and retest of each experiment.

Task Instructions and Materials

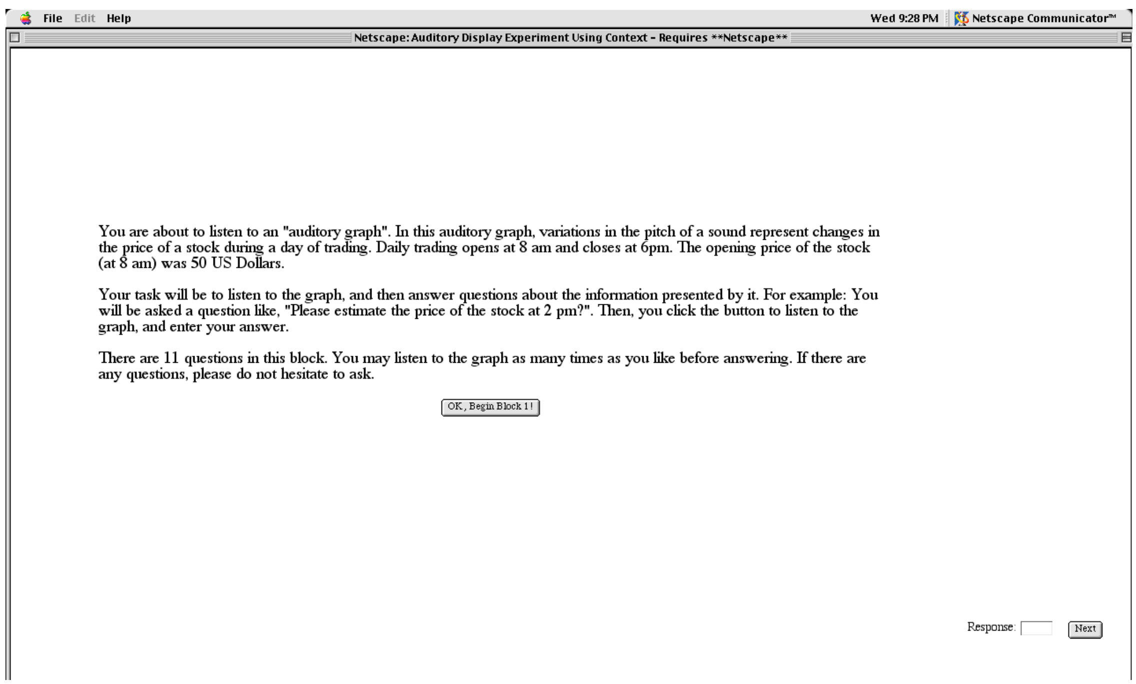


Figure B1. Instructions viewed by all participants prior to the pretest.

Figure B2. Task view for all participants during the pretest.

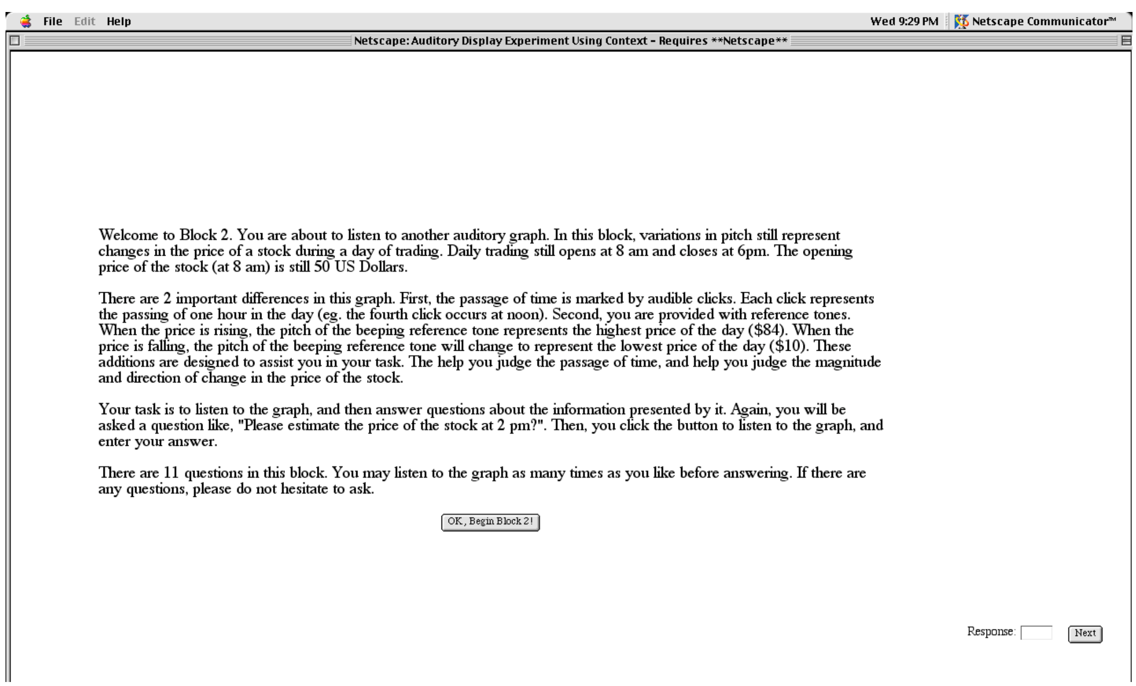


Figure B3. Typical instructions viewed by participants prior to the retest. Actual text viewed varied dependent upon the contextual setting assigned to each experimental group.

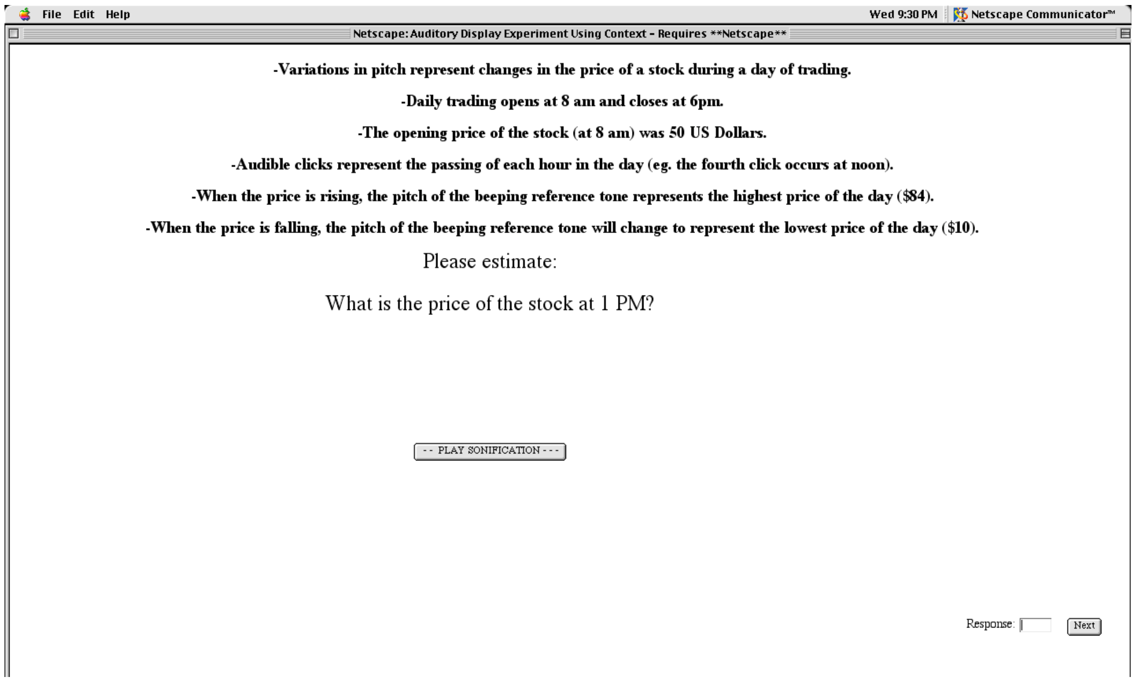


Figure B4. Typical task view of participants during the retest. Actual text varied dependent upon the contextual setting assigned to each experimental group.

Training Programs and Filler Task Materials

The following screen shots illustrate typical views of the material participants saw as they executed the training or filler task during Study 2.

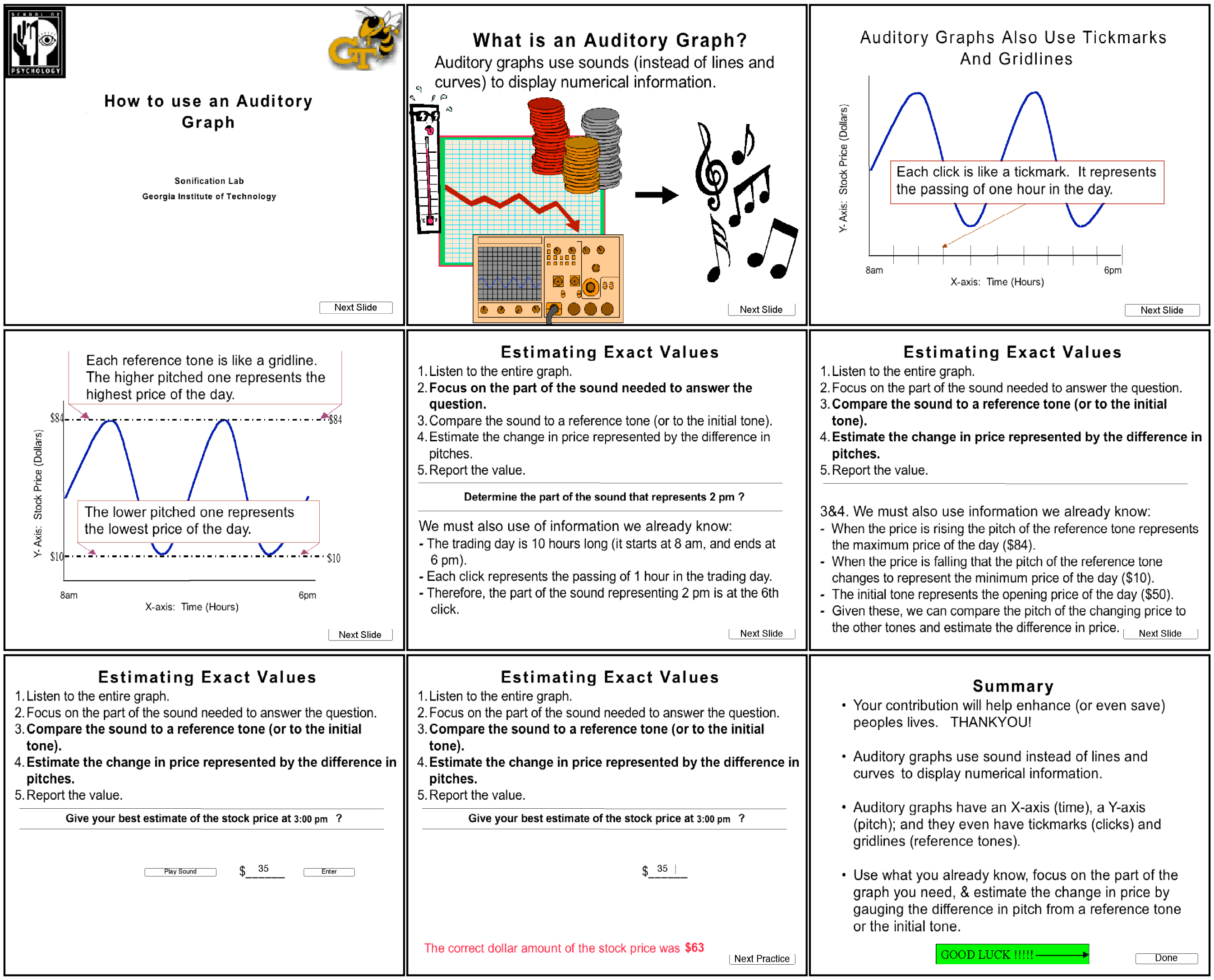


Figure B5. Typical views of a Study 2, Group 8 participant as he executed the training program. Actual text in each training program (Groups 5 - 8) varied, dependent upon the contextual setting assigned to each experimental group.

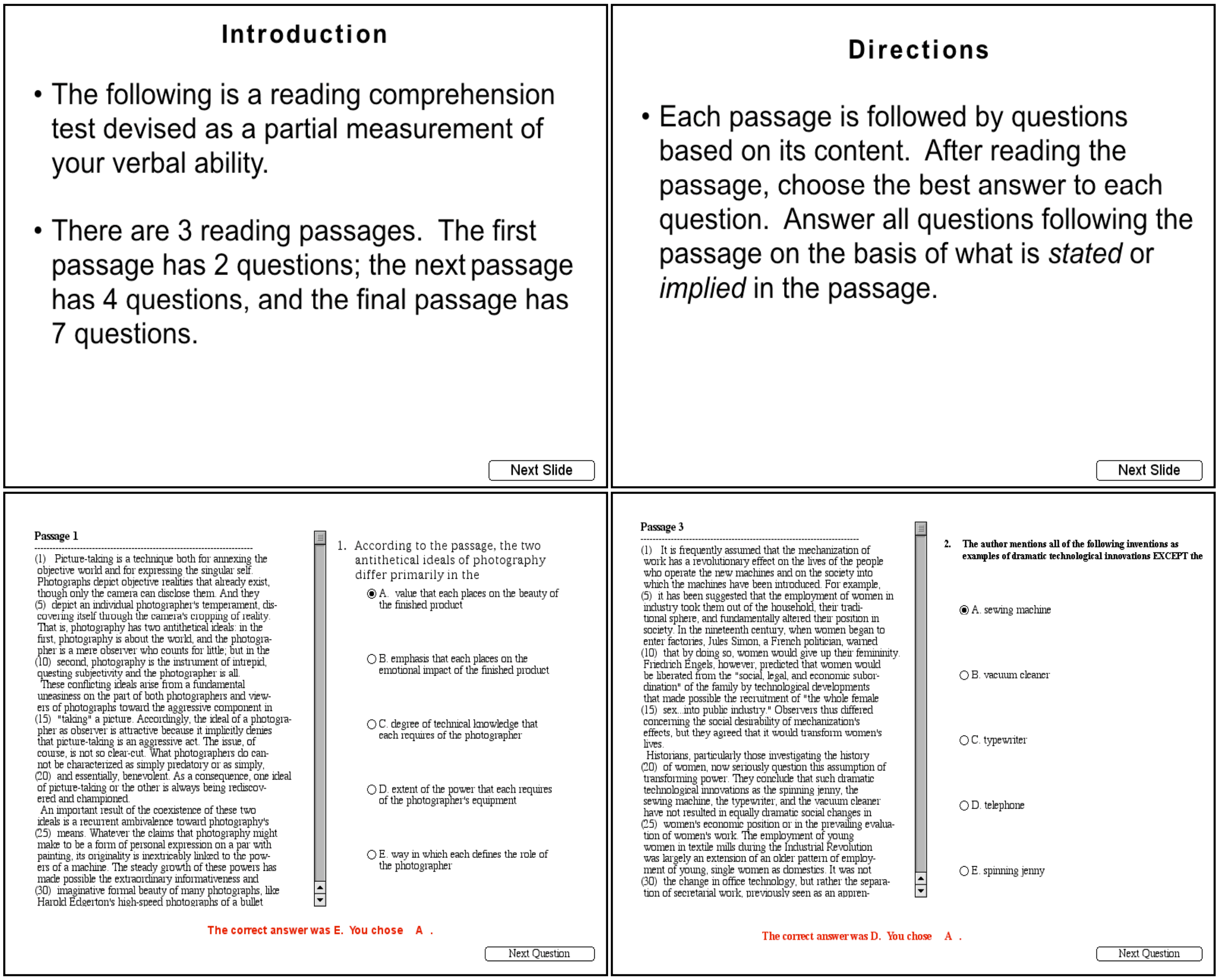


Figure B6. Typical views of a Study 2, Group 1 - 4, participant as he executed the filler task in the stead of one of the training programs.

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.

Annett, J. (1991). Skill acquisition. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 13-51). New York: Wiley.

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, July 29 - August 1). *Testing the effectiveness of sonified graphs for education: A programmatic research project.* Paper presented at the International Conference on Auditory Display, Espoo, Finland.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* (Vol. xiii). Cambridge, MA: MIT Press.

Carlile, S. (2002, July). *An overview of auditory dimensions that can be used to represent information.* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.

Cohen, J. (1994). Monitoring background activities. In G. Kramer (Ed.), *Auditory display : sonification, audification, and auditory interfaces* (pp. 449-531). Reading, Mass.: Addison-Wesley.

Cuddy, L. L. (1968). Practice effects in the absolute judgement of pitch. *Journal of the Acoustical Society of America, 43*(5), 1069-1076.

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.

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.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.

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)*(September), 341-351.

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.

Gardner, J. A., Lundquist, R., & Sahyun, S. (1996, November). *TRIANGLE: A practical application of non-speech audio for imparting information.* Paper presented at the International Conference on Auditory Display, San Francisco, CA.

Gaver, W. W., Smith, R. B., & O'Shea, T. (1991, April). *Effective sounds in complex systems: The ARKola simulation.* Paper presented at the CHI Conference on Human Factors in Computing Systems, New Orleans.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York,: Appleton-Century-Crofts.

Heller, M. A., & Auerbach, C. (1972). Practice effects in the absolute judgement of frequency. *Psychonomic Science, 26*(4), 222 - 224.

Johannsen, G. (2002, July). *Auditory display of directions and states for mobile systems.* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.

Kramer, G. (1993). *Sonification of financial data: An overview of spreadsheet and database sonification.* Paper presented at the Virtual Reality Systems, New York.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). *The sonification report: Status of the field & research agenda*: The International Community for Auditory Display.

Lunn, J. H. (1948). Chick sexing. *American Scientist, 36*, 280-287.

Marila, J. (2002, July). *Experimental comparison of complex and simple sounds in menu and hierarchy sonification.* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.

McAdams, S., & Bigand, E. (1993). *Thinking in sound: the cognitive psychology of human audition.* Oxford: Oxford University Press.

Moore, B. C. J. (1997). *An introduction to the psychology of hearing* (4th ed.). San Diego, Calif.: Academic Press.

Neuhoff, J. G., & Wayand, J. (2002, July). *Pitch change, sonification, and musical expertise: Which way is up?* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, Calif.: Sage Publications.

Quinones, M. A., & Ehrenstein, A. (1996). *Training for a rapidly changing workplace: Applications of psychological research*. Washington D.C.: American Psychological Association.

Salvendy, G. (1997). *Handbook of human factors and ergonomics* (2nd ed.). New York: Wiley.

Smith, D. R., & Walker, B. N. (2002, July). *Tick-marks, axes, and labels: The effects of adding context to auditory graphs.* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.

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, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Tufte, E. R. (1990). *Envisioning information*. Cheshire, Connecticut: Graphics Press.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, Conn.: Graphics Press.

Walker, B. N. (1994). *Auditory display and earcons* (Unpublished term paper). Houston, TX: Rice University.

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., & Kramer, G. (1996). *Mappings and metaphors in auditory displays: An experimental assessment.* Paper presented at the International Conference on Auditory Display, Palo Alto, CA.

Walker, B. N., & Lane, D. M. (2001, July). *Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners.* Paper presented at the International Conference on Auditory Display, Espoo, Finland.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An introduction to human factors engineering*. New York: Longman.

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.

Zotkin, D. N., Duraiswami, R., & Davis, L. S. (2002, July). *Customizable auditory displays.* Paper presented at the International Conference on Auditory Display, Kyoto, Japan.