A Comparison Of Auditory And Visual Graphs For Use In Physics And Mathematics

STEVEN C. SAHYUN
University of Wisconsin - Whitewater
rev.7 1/27/16

This paper describes the investigation of auditory graph presentation methods. Two pilot studies and one treatment study evaluated the success of subjects when answering questions with graphically presented data and used a combination of interviews and Web based tests composed of math and physics questions. An additional follow-up pilot study relating to auditory graphing preferences is also described.

The ability to interpret graphical information is a prime concern in physics as graphs are widely used to give quick summaries of data sets, for pattern recognition, and for analysis of information. While visual graphs have been developed so that their content can be readily and concisely discerned, there is great difficulty when someone is unable, because of their environment or due to physical handicaps, to view graphs. An alternative to the visual graph is the auditory graph. An auditory graph uses sound rather than pictures to transmit information. The auditory graphs used in this study were of single valued x, y data constructed by mapping the y axis to pitch and the x axis to time. The addition of auditory indicators to mark first and second derivative information were useful in some graphs as well as further audio enhancement to indicate negative data values.

The treatment study shows that students with very little training can use auditory graphs to answer analytical and identification type questions. Students using only auditory graphs achieved 70% of the performance level attained by subjects using visually presented graphs. The performance level of several blind subjects using auditory graphs exceeded that of sighted first-year physics students regardless of the presentation format.

Categories and Subject Descriptors: ... [...]: ...

General Terms: ...

Additional Key Words and Phrases: ...

1. INTRODUCTION

1.1 Purpose of the Research

The purpose of this study is to develop and test auditory graphing methods in order to provide quick, easy to learn, and useful access to information. The intent of this research is to provide access to information for people who are blind, visually impaired, or otherwise not able to view data in a visual format.

This research asks whether current auditory graphing techniques are adequate to permit users to identify and make interpretations from graphed data. Specifically: Can these graphs be used to answer math and physics questions? If not, what modifications to auditory graphs are most helpful to increase understanding of the

Author's address: Steven Sahyun, Physics Department, University of Wisconsin - Whitewater, Whitewater, WI 53190 email: sahyuns@uww.edu

displayed data? What disparity arises when students use auditory graphs to answer questions compared to students using visually presented graphs? The working hypothesis for this study is that in many cases, sound graphs can be as effective as visual graphs to display, and make inferences from, data.

In order to answer these questions and test the hypothesis, four test instruments were developed: the TRIANGLE Pilot test, a Web Pilot test, a Web-based Auditory Graph test, and a Web-based Auditory Preference Pilot test. All of the auditory graphs presented in this study relate the y axis to pitch and the x axis to time. The effectiveness of the auditory graphs was determined from the results of these test instruments. The studies described herein are tests of the most basic auditory display techniques and are an effort to discover possible improvements for these displays.

1.2 Relationship of This Study to Previous Research

Many techniques of information design have evolved during the 500 years since the Italian Renaissance [Tufte 1990]. One of the more common and important methods of describing a single data set is the two-dimensional line graph. The large number of graphs displayed in a typical general physics textbook [Halliday et al. 1993] is an indication of the importance that these graphs play in physics instruction. Unfortunately, students have difficulty interpreting these graphs and there have been several studies devoted to finding out why students have problems analyzing information from graphs[McDermott et al. 1987][Peters 1982][Price et al. 1974].

Researchers are developing auditory display techniques, e.g. methods of displaying data with sound instead of pictures, for use in medicine, geological surveying, fluid dynamics, and many other subjects [Kramer 1994]. These techniques often rely upon sophisticated tonal patterns or modification of original wave patterns. Since most areas of physics require data display, auditory display techniques can be extended to the field of physics. Auditory graphs provide an alternate method of data display. These graphs represent data with sound and have the potential for quick production by computer programs. In addition, auditory graphs are easy to play and understand when properly designed. They are not limited to line of sight or to physical touch for access. Since auditory graphs represent equations, they can be stored either in that format or as sound files for later retrieval and playback.

The techniques employed in the experimental design and data analysis are common in behavioral sciences [E. 1968]. This study is largely about how people answer questions given some differentiating factor. In the case of the current study, the differentiation is the type of data presentation. The questions used in this study relate to difficulties students may have with graphical information. Questions were developed from material that first-year physics students would generally encounter.

While physics is the primary concern of this research, graph comprehension is of broader interest than only physics. In the area of mathematics, there have been enough studies to warrant at least one very comprehensive review paper [Leinhardt et al. 1990]. The field of economics has also had research directed toward graphical comprehension [Cohn and Cohn 1994][Price et al. 1974]. Thus, results of this physics-oriented study should be applicable to other disciplines.

Understanding of physics comes from the ability to reconcile models of what a person believes to occur in an event and the data produced by that event. The

Force Concept Inventory (FCI) study[Hestenes et al. 1992] and other studies have shown that student misconceptions significantly inhibit understanding of physics. Since graphs are heavily used in physics, difficulties in understanding physics may be a result of problems in understanding, and an inability to interpret, the graphical information that is presented. Alternate presentation methods of data may reduce the difficulty students have with graphs. Studies on alternate presentation methods have generally focused on the use of microcomputer based laboratories (MBLs) [Linn et al. 1987][Mokros and Tinker 1987][Brasell 1987][Thornton and Sokoloff 1990]. These studies focus on how immediately generated graphs of physical phenomena play a role in student understanding of those phenomena.

1.3 Goal of This Research

The primary focus of this study is to demonstrate that simple auditory display methods can be used to impart enough information so that people can make informed and consistent decisions based on what was heard. The goal is to demonstrate that auditory graphs can act as a practical replacement of, or supplement to, visual graphs.

In the process of searching for an intuitive auditory graph, the graphing techniques were adjusted by the addition of various auditory indicators. The additions provided solutions to specific difficulties with the auditory graphs that subjects had encountered. These additions can be incorporated when developing graphing software, as has been done with the Audio Graphing Calculator [ViewPlus Technologies 2007], to provide better display and interface options for users. With accessible graphing software, all people can take advantage of the power of visual and auditory graphing techniques. By using the additional auditory indicators and a minimum amount of training, users are able to achieve greater comprehension of the graphed data.

Students who have difficulty interpreting graphical information presented as visual plots may find auditory displays of the data more understandable. This greater understanding of the data may reduce the number of misconceptions associated with graphed information.

2. THEORY

2.1 Auditory Display Techniques

Auditory display techniques are methods that use sound to convey information. They can be simple such as the sounding an alerting bell, or complex, e.g. a tone pattern to represent a relationship between variable quantities. Comprehensive descriptions of various methods and developments using auditory displays can be found in the proceedings of the International Conference on Auditory Display (ICAD) and associated texts [Kramer 1994]. Recent research has focused on the ability of auditory displays to represent graphical information. The current study is an attempt to extend these results and to provide auditory displays that are more effective.

Auditory display techniques can range from a symbolic format to an analogous representation of data. A symbolic representation is one where a sound represents some item of importance, the meaning of which is established through learned association. An alarm or spoken language is an example of symbolic auditory display. Symbolic representations of information may employ the use of metaphorical association, such as a dripping sound as an alert that there is a memory leak in a computer program. An analogous representation is a direct correspondence between the data and the sound produced. Examples of analogous displays are: relating the noise in a room to the number of people, the sound that a car engine makes to its r.p.m., or a data points value to pitch [Kramer 1994].

2.2 Types of Sonification

Sonification is a term that is generally used when relating the process of converting data into an acoustic format for presentation. Many methods of data sonification exist including modifying the tonal qualities of the sound, repetition rates, or spatial locations and the main classes are described in Kramer [Kramer 1994]. The method of data sonification used in this study is a multi-dimensional analogous representation where data points are represented by set of characteristics of sound. Examples of sound qualities used in this study include pitch, amplitude (volume), timbre (clarity of the note, affected by the incorporation of higher harmonics), the spatial location. By combining the different qualities of sound, the multi-dimensional nature of the information is portrayed in a concise manner. When multiple sounds are played, e.g. two data sets represented by unique instrument sounds, the sonification is sometimes noted as having multiple voices. In the later studies described in this paper, two voices were used for the data (for positive and negative data point values), and a third for the derivative information. The data sound also was panned from the left to right channel speakers to produce a spatial separation as a further enhancement of the graphs, but as the auditory graphs were displayed on remote computers, some of which may have not had multiple, stereo speaker systems, this aspect of the auditory graphs was not tested.

2.3 Mapping Methods

A common method of creating auditory graphs from data, maps the dependent variable (y axis) to pitch and the independent variable (x axis) to time. Thus, auditory graphs are a series of tones played in time. Studies [Mansur et al. 1985][Flowers and Hauer 1995][Flowers et al. 1997] have demonstrated the ability of subjects to understand the general trends of these types of auditory graphs. The auditory graphs in the current study used the pitch-time mapping method as a starting point for graph development.

The exact relationship for the y axis pitch varies between experiments. However, there is always the association that high pitch (higher frequency values on the order of a couple of kilohertz) represents high data values, and low pitch (around 200 Hz) represents low values. This method provides a direct one-to-one mapping between pitch and data. In the Triangle Pilot, Web Pilot, and the Web Auditory Graph tests all of the graphs had zero or positive y axis data values. Thus, the lowest magnitude value had the lowest note, and the highest magnitude value had the highest note.

There are two common mapping methods: linear mapping, in which data have a linear correspondence to frequency; and logarithmic mapping, also called a chromatic scale mapping. Linearly mapping a value x to a frequency n, is accomplished by the relation

$$n = n_{min} + (n_{max} - n_{min}) \left(\frac{x - x_{min}}{x_{max} - x_{min}} \right). \quad (1)$$

where x_{max} and x_{min} are the maximum and minimum data values to be graphed, and n_{max} and n_{min} span the desired frequency range.

The chromatic mapping method has a logarithmic pitch relationship found in musical instruments. This method converts a datas value x to a frequency f (to distinguish from the linear mapping method) by the relationship

$$f = f_{min} \times \left(\frac{f_{max}}{f_{min}} \right)^{\left(\frac{x - x_{min}}{x_{max} - x_{min}} \right)}. \quad (2)$$

where the desired range of frequencies is from f_{max} to f_{min}.

The Auditory Preference Pilot test included graphs with negative y axis values so the lowest absolute data value was associated with the lowest pitch and a secondary audio cue, by changing the instrument timbre, was added to indicate to the listener when the data represented negative values. First derivative information was portrayed by auditory cues representing points where on a visual graph the data crossed grid-lines, and second derivative information was portrayed by changing the timbre of the auditory cue sound.

2.4 Hypothesis

The ability to present data with an effective auditory format is one of the prime goals of this research. The current research is directed towards demonstrating not only that people can understand auditory graphs, but that they can also be used as effective displays for understanding and analyzing information. The working hypothesis for this study is that: in many cases, sound graphs can be as effective as visual graphs for data representation and for making inferences about that data.

If graph types are highly equivalent, as suggested by the studies by Flowers [Flowers and Hauer 1992][Flowers and Hauer 1993][Flowers et al. 1997], then there should be little difference between a student's ability to identify and interpret information when given auditory or visual graphs. However, there is the possibility that there will be differences in performance due to unfamiliarity with the sound format. By asking questions based on graphical material, the effectiveness of auditory graphing methods can be measured.

To test the hypothesis, it is important to determine how well students are able to answer graph-based questions. One testing method is to have two equivalent groups of subjects answering questions. Each group receives either visual or auditory graphs with the questions. While identification of simple graphs is important, students ability to interpret what those graphs mean is also significant. Thus, this study includes two types of questions: those that involve interpretation to identify a function, and those that require analysis of the data for interpretation of the physics concepts that the graphs represent.

A comparison of the performance of subjects using auditory graphs to that of subjects using visual graphs may indicate a difference between the two display methods. In addition, subjects may have better understanding of questions when both auditory and visual graphs are used. Subjects may find that the combination of formats is a helpful method to enhance the graph. Thus, three testing groups are reasonable to provide comparative data: visual graphs, auditory graphs, and both auditory and visual graphs. When the number of subjects is sufficiently large, a random assignment to one of the three groups should produce equivalent testing groups.

Comparing the performance of subjects ability to answer graph-based questions with respect to which graph type they receive may yield several outcomes. The first, is that if student performance is equivalent among the auditory, visual, and the combination displays, then the display modalities are equivalent. They can answer and analyze questions equally well.

Studies by Flowers and Hauer demonstrated that there are several areas of perceptual equivalence between auditory and visual graphs. Thus, the possibility for equivalent performance when answering questions is a reasonable supposition. Turnage et al. [Turnage et al. 1996] also reported rough equivalence between auditory and visual graphs when subjects were asked to identify properties of simple periodic wave patterns.

A second, albeit unlikely, outcome also exists: auditory graphs could outperform their visual counterpart. This outcome could be the result of an increased salience from auditory cues. Flowers and Hauer noticed this effect in some parts of their graph discrimination studies [Flowers and Hauer 1995][Flowers et al. 1997].

A more likely situation, however, is that there would be a performance difference due to greater familiarity of the visual graphs. This is understandable as students are trained to recognize and use visual graphs for many years by the time they take university level courses. Auditory graphs, on the other hand, are a completely new experience, and the amount of training they receive may strongly influence their performance. A study investigating an upper limit of the use of auditory graphs to convey information would require subjects with extensive auditory graph training. Comparable, but not equivalent, performances for discriminating differences between data sets when using auditory or visual graphs have been shown in the aforementioned studies [Flowers et al. 1997][Turnage et al. 1996].

If subjects completely fail to understand data presented with auditory graphs, currently reported research would be called into question. A finite limit on the practicality of auditory displays may exist. Also, such a result may demonstrate that the understanding of auditory graphs is not intuitive. Even simple data comparisons and analysis would require that subjects have intensive training and alternate auditory methods would need to be investigated.

The current research addresses this issue by investigating how well students are able to answer physics and math questions based on graphed data. Many systems studied in physics use graphs to display data for analysis. Ideally, the data can be represented by mathematical equations. Physics is an ideal topic for the study of auditory graphs since there can be a separation between the identification of mathematical functions representing the graph, and the inferred properties of a system that the graph represents.

3. TEST METHODS

3.1 Relationship of Tests

This study comprises the results from four investigations. The first investigation was a pilot test to examine the possibility of using the Science Access Project's TRIANGLE program to provide auditory graphs to subjects. The second study was a pilot test to investigate the use of a Web based format to provide auditory graphs to subjects. Questions and techniques from these first two pilot tests were modified and used to create a Web based auditory graph study given to over 200 students. This third study will be called the Web Auditory Graph test. The fourth study was a Preference Pilot test involving a detailed investigation of subject reactions to aspects of auditory graphs and of graphs with additional complexity.

The general process of data collection in the first three studies consisted of providing each subject a statement of informed consent to read and agree to, a survey questionnaire (Survey) for demographic purposes, a pre-test to assess equivalence among three sub-groups (Pre-test), and a number of questions consisting of one or more randomly assigned graph types (Main test). Recruitment of subjects involved soliciting various Introductory Physics instructors to provide an extra-credit opportunity for students who volunteer from their classes. The Auditory Preference Pilot test differed as subjects taking the test were not assigned into graph type groups, there was no Pre-test, and subject recruitment was based on convenience.

The informed consent page consisted of a statement of the test procedure that was involved, the names of the principal researchers and contact numbers, and an agreement clause. This page was required by the Institutional Review Board as human subjects were involved. The Survey questionnaire was used to gather data such as gender, age, and the number of physics, math, or other courses relating to graphical information that subjects had taken. This page also queried whether the subject had musical training or any vision or hearing difficulties.

The Pre-test consisted of a total of five questions about two graphs, four questions for the first graph and one question for the second graph. The first four questions asked for the number of local maxima, the location of maximum slope, etc. and were used to determine whether the subject could properly read a graph. The last question was similar to the questions used in the Main test and concerned the interpretation of the physics described by a graph.

The Main test was presented in different manners depending on the study. For the Triangle Pilot test, the subjects were presented with a series of 14 multiple-choice questions, one at a time, on a computer screen, and either listened to and/or looked at a graph that the question related to. Subjects answers were recorded in a written format. Assignment of the graph presentation method was random, with the subject receiving a single method (visual, auditory, or both visual and auditory) for all questions. For the Web Pilot test, the subjects accessed a series of Web pages that presented the graph and multiple-choice question, with one question per Web page. Answers were transmitted by selection of multiple-choice "radio buttons" and the answer was recorded by a scripting program. For the Web Auditory Graph test, the same presentation and recording method was utilized as for the Web Pilot, although the number and type of questions were modified and extended due to reliability and validity issues.

The Auditory Preference test was also a Web-based testing process. However, due to technical limitations, this test was only presented to local subjects. The test consisted of nine question pages: four consisted of pair-wise auditory graph comparisons, four involved matching an auditory graph to a visual graph (two questions were matching a visual graph to a choice of auditory graphs, and two were matching auditory graphs to a choice of visual graphs), and one page with five-point Likert ratings of 6 graph types. Each question page had a text field for subjects to provide comments and reasoning for their choices.

The level of difficulty of the graph questions was gauged to the target population for which the graphing method was used. As the TRIANGLE program was designed for college level use, appropriate questions centered on introductory college level physics. The population for the study was drawn from subjects who had taken, or were in the process of taking college level physics courses.

The subjects recruited in this study were all volunteers, and no incentive for their performance level could be applied and there is no guarantee that the subjects performed at their best level when answering the questions. However, since subjects were randomly assigned to different treatment groups from the same population, on average, any performance issues should be the same for each group of subjects. In addition, the results of the test were adjusted for random guessing which should reduce the effect of any student apathy towards the tests.

3.2 TRIANGLE Pilot test

The first experiment was a pilot test to investigate the advantages, difficulties, and question layout of a study involving auditory graphs. This experiment used the TRIANGLE program to display the questions and the visual and auditory graphs to a majority of the subjects. The results from this experiment not only helped elucidate several inadequacies in the production and testing of the auditory graphs and provided the basic testing material that was used in the subsequent studies.

The TRIANGLE program's primary purpose was to provide a workspace for students and scientists to read, write, and manipulate mathematics. This program was developed by the Science Access Project at Oregon State University and included a text region where the questions could be displayed, as well as a display for visual and auditory graphs generated from a table of data points. The calculator evaluated y versus x functions and displayed the results in a plot window and sonification of the data was represented using a linear relationship of pitch for the y axis data values. In addition, data points were located in space by stereo speakers so that the sound panned from left to right. The resulting auditory graphs could be played either continuously, or by stepping through data points with keys on the computer's keyboard. TRIANGLE displayed visual and auditory graph formats, as well as providing functionality for DOS screen readers and haptic access through the use of a Braille display. With the use of a Braille display, there is a moving icon on the screen to provide information about the graph to users who are both blind and deaf [ICAD 96 International Conference on Auditory Display 1996].

The auditory graphs produced by the TRIANGLE program created the question of: How useful is this type of display to the intended user? To answer this question, it was necessary to develop a testing method between auditory graphs and visual graphs in the context for which they would be used in the program. The context is the investigation of properties of mathematical functions and the display of scientific data.

A Pre-test consisted of five questions to check subject understanding of basic graph concepts. The Pre-test was given in a printed form, and consisted of labeled graphs that subjects could easily identify. The Main test consisted of 14 multiple-choice questions. Additionally, there were fill-in-the-blank supplements for two of the multiple-choice questions. The questions were designed to be equally valid for either visual or sonified displays.

Subject matter for the questions centered on previously published research involving graphs and physics. Most notably, questions from the Force Concept Inventory (FCI) [Hestenes et al. 1992], Mechanics Baseline [Hestenes and Wells 1992] test, and questions in Beichner's [Beichner 1994] study were used after some modifications. Other questions were developed after an analysis of subject matter presented in several introductory physics text books. The final questions were reviewed for content validity by several physics and science education faculty known for their interest and excellence in teaching at Oregon State University. To check reliability, the test questions were constructed to be applicable for split-half analysis.

The TRIANGLE program imposed a limitation on the auditory graphs because at the time of the study, there was no method for describing negative values in an auditory format. Therefore, graphs used for the questions could reference only positive y axis values. In addition, when testing subjects, it was necessary to use a guided interview process due the difficulty of subjects in learning and using the TRIANGLE program to display questions and graphs.

Subjects were randomly assigned to one of four graph category groups (three subjects per group): visual graphs printed on paper (P, Print group), visual graphs displayed on the computer (V, Visual group), auditory graphs produced by the computer (S, auditory graph group), or both visual and auditory computer graphs (B, Both group) presented on the computer. The presentation method with both sound and picture graphs was to check for any increase in students' ability to answer questions due to multi-modal presentation. Two graduate students participated using auditory graphs for purposes of testing validity, and for estimation of time allotment for scheduling purposes. Subjects in the test groups that used sound graphs were given a brief description about auditory graphs but no specific training was performed.

3.2.1 *Conclusions From the TRIANGLE Pilot Test.* The guided interview process for the test was necessary so that the interviewer could answer questions about the test, set up questions and graphs on the computer display, and observe if there were any particular difficulties with the 30-minute testing process. One subject noted that the proximity of the interviewer was uncomfortable in a performance-type setting. The proximity issue was immediately resolved, but question set-up became more time consuming. The subsequent Web Pilot test removed this issue by providing a self-running testing environment that could be accessed from remote locations and did not require the intervention of an experienced user for setting up test questions.

Other results from the pilot test suggested a large disparity between subjects who had auditory only graphs compared to the other subjects. While it was observed that subjects were able to distinguish absolute values by sound (i.e. pitch being higher or lower), and that subjects were able to identify the first derivative of the function (increasing or decreasing) they did not seem to be able to be able to correctly identify the second derivative (the rate at which a function is increasing or decreasing). These conclusions were indicated by poor results on graphs portraying curved functions. More experienced subjects (graduate students) were able to interpret the shape of the graph from the limited information because they immediately converted the sound into a picture. Even with this experience, they still had difficulty interpreting graphs that had a positive curvature.

A solution to the problem of identifying curved graphs is to enhance the derivative information. One method is to add sonic indicators for the first and second derivatives such as with a series of clicking noises, where the rate (tempo of the clicks) represents the first derivative. The tempo is set by how often the curve crosses some y axis interval. The pitch of the clicks can indicate the second derivative. This method was used in the auditory graphs for the Web Pilot, Web Auditory Graph, and the Auditory Preference Pilot tests.

It was evident even in the limited trial that the original hypothesis of the equivalence between simple graphs produced with TRIANGLE's basic sonification technique and visual graphs was not realized for a significant part of the pilot test material. One factor for these results seemed to be the lack of familiarity with auditory graphs for subjects in the auditory graph group. While some training is evidently necessary, the goal of this auditory display is to have a method that is reasonably intuitive. This pilot test demonstrated the need for first and second derivative information to be incorporated into auditory displays in order to increase the distinction between curved and linear graphs. Modifications to the questions and testing method, additional initial information, and altering the display formats formed the basis for the second pilot test called the Web Pilot.

3.3 WEB PILOT

The second experiment conducted was a pilot test to investigate the advantages, difficulties, and question layout of a study involving auditory graphs using the World Wide Web (Web) as a testing environment. The Triangle Pilot test suggested that there would be logistical difficulties when having a large number of subjects take a test with the computer generated auditory graphs used in the pilot. Also, a more flexible testing environment was necessary than that provided in the Triangle Pilot. It was suggested to the author that a Web-based test could overcome these difficulties and provide many advantages. Such a test would allow access by many student subjects as well as provide a flexible testing environment. Pictures, sounds, and text could be easily configured and changed, and multiple graphs could be displayed with little effort. In addition, it would be possible to record subjects responses with Web-based forms [Cebula 1997].

This experiment, named the Web Pilot, used a standard Web browser program, such as Netscape or Microsoft's Internet Explorer, to display the introductory materials, questions, and visual and auditory graphs in a series of Web pages. The results from this experiment helped show where revisions were needed when creating a test which at the time of this study was a new medium. Currently, there are many Web based survey possibilities, but at the time of this study, such services did not exist and all of the Web based features for displaying questions and recording survey responses were developed by the author. Unlike the previous pilot, subject volunteers accessed the Web Pilot test site from remote computers at various locations, at times of their choosing. Several PERL scripting programs recorded data from subject responses to questions presented on the Web pages.

As the testing process was designed for introductory physics students, instructors of these courses were solicited during the Fall 1997 quarter for the possibility of letting their students participate in this study. It was arranged with one instructor of an introductory algebra-based physics course at Site A, a large Pacific coast state university to provide extra-credit homework points to students taking Web Pilot test. The instructor announced the location of the Web Pilot tests introductory Web page in class and posted a link on the courses information Web page. Student volunteers were given one week after the initial announcement to complete the test. From this single course, 221 out of about 400 enrolled students completed the Web Pilot test. At most, six students who logged into the test, due to technical difficulties or lack of interest, did not complete all of the questions. Only subjects completing all questions had their data recorded. Of the 221 recorded subjects, 74 subjects received the Main test with auditory graphs, 75 received visual graphs, and 72 received both auditory and visual graphs. These numbers allow for statistically significant results at the $p = 0.05$ level since $n > 62$ for each group.

Given a normal population distribution, the approximate size of a group, n required for a 95% chance of the average measurement, $\overline{X}$, is within the limits of $\mu \pm L$, where μ is the target population mean and L is the error limit (5%), and a standard deviation of 20% due to an average of 5 answers per question is given by:

$$n = \left(1.96\frac{\sigma}{L}\right)^2 = \left(1.96\frac{0.2}{0.05}\right)^2 \approx 62. \tag{3}$$

When subjects accessed the Web address announced in their class, they were presented with a welcoming page stating the purpose of the test. The welcoming page also contained a brief description of auditory graphs, and an optional link to a Web page containing further descriptions and examples of auditory graphs.

After the introductory Web page, subjects were presented with a Web page to record their name and a school class code into text entry form fields. The names, class code, and an identification (ID) code number were appended by a PERL script program called namepage to a secure file that contained previous subjects names. This program also randomly assigned the subject to one of three graph test groups: "V: visual graphs", "S: auditory graphs", and "B: both auditory and visual graphs". ID and graph codes were passed to subsequent Web pages.

As in the prior experiment, there were three instruments: an initial survey questionnaire (Survey), a Pre-test, and a Main test. The content and subject matter of the Survey, Pre-test, and Main test were similar to those of the Triangle Pilot, but with some revisions to the questions resulting from difficulties in wording found during the Triangle Pilot. There was no paper presentation method as was performed in the Triangle Pilot due to that tests similarity in scores between the visual and paper presentation methods.

The TRIANGLE Pilot test questions were converted to a Web-based format. At the time, methods to display the questions' text and visual graphs were not difficult as the standard Web browser has this ability built into the display. However, due to bandwidth issues, MIDI sound files in conjunction with Apple's QuickTime player were used to embed sound files into the Web pages to reduce page file size and load times. The MIDI protocol uses data streams to trigger stored sound-wave patterns on the host computer. Each sound wave represents an instrument, or voice, whose pitch, onset, duration, and decay are triggered by the data. For the Web Pilot, y axis data values were represented with a piano voice that varied in pitch. To provide alternate access to the sound files, links were included so that subjects could download the MIDI formatted file or a much larger .wav formatted file. The visually presented graphs were produced with the KaleidaGraph program from Synergy Software. These graphs were converted to a .gif file format for display on the Web pages.

One of the TRIANGLE Pilot subjects made the suggestion to add tick marks to represent the y axis values and this suggestion was incorporated into the new auditory graphs. When data values passed certain intervals, a tick mark was sounded. The tick mark sound was represented by a drum instrument voice. The resulting frequency, or tempo, of tick marks represented the magnitude of the slope, or first derivative, of the graph at a given point. A small magnitude slope resulted in a slow tempo of the sounding of the tick marks, while a large magnitude slope resulted in a fast tempo. The sign of the slope was easily determined by listening to whether the data value pitch was increasing or decreasing.

While the tempo of the drum beat to indicate slope provided much needed information, the second derivative was also easily incorporated by modifying the pitch of the drum voice. To reduce the auditory load, it was decided to only use three pitches to represent the second derivative, one for negative values, one for positive values, and a third for 0. The optimal choice of pitches is a matter some debate and is the subject of future research.

For this study, the investigator chose to represent negative values of the second derivative with a high drum pitch, positive second derivative values with a low pitch, and 0 was represented with a pitch in between the two. Thus, the graph of $y = x^2$, from 0 to 1, had an increasing piano voice tone and a low pitch drum that would increase in tempo, and the graph of $y = 1x^2$, from 0 to 1, had a decreasing piano tone and a high pitch drum that also increased in tempo. The graph of $y = x$ had an increasing piano tone with a constant tempo drum beat whose pitch was between the high and low drum pitches. The reasoning for this choice of the tick mark pitches was that, aside from areas with inflection points, negative curvature occurs at local maxima, while positive curvature occurs at local minima. Thus, the tick mark pitch would reinforce the data pitch in those areas.

The auditory graphs used in this study were created prior to the test and included as embedded QuickTime or linked .wav files. They were produced in a multi-step process. The $x - y$ data sets used to create the graphs in the Triangle Pilot were converted by a small program written by the author into text formatted audio instruction files. A sound program called MIDIGraphy [Tontata], was then used to import the text files and converted the text data to MIDI sound files. The converter program set the instrument, time duration of the notes, length of the play time of the data set, and calculated and set the drum tick mark derivative information.

Group	# subjects	% Correct Pre-Test	Main Test	Std. Dev.
S: Auditory graph only	74	85	41	16
V: Visual only	75	85	66	17
B: Both Auditory and Visual	72	80	66	19

Table I. Results of the Web Pilot test.

The MIDI file was converted into the .wav format with SoundMachine [Kennedy], and was converted into Apple Computers QuickTime format with the MoviePlayer program though more modern applications are now available.

3.3.1 *Web Pilot Results.* The Auditory Graph group performed at a lower level (40% correct) than did the Visual (66%) or Both auditory and visual graph (66%) groups.

ANOVA of the Pre-test results showed $F = 1.71 < F_{critical} = 3.04$ which indicates no significant difference between groups at the $\alpha = 0.05$ level. ANOVA of the Main test results showed $F = 32.37 > F_{critical} = 3.04$ which indicates there was a significant difference between groups. T-test pairwise comparison between Auditory and Visual groups yielded $|t| = 8.13 > t_{critical} = 1.99$ and for Auditory vs. Both yielding $|t| = 6.91 > t_{critical} = 1.99$ indicating a significant difference between the auditory graph group and the other two groups. In the Both vs. Visual comparisons $|t| = 0.21 < t_{critical} = 1.98$ indicating that there is no significant difference between these two groups. Thus, the difference in performance between the auditory graph only group (S) at 40% correct and the others (V, B) at 66% correct is a significant effect.

It is interesting to note that the Auditory graph group took 1.8 minutes longer to answer the 14 questions of the test than did the Both group. This averages out to about 8 seconds more per question. Since both of these groups had similar times for download and display of the graphs, the extra time may indicate the extra time required for understanding the graph when there is no visual cue. However, more likely explanations are that on average, the Both group did not play the graphs, or that the Auditory group replayed the graphs an extra time.

3.3.2 *Conclusion From the Web Pilot Test.* It was strikingly apparent from this pilot study that using the World Wide Web as a testing environment had enormous advantages. An automated display and recording system was able to provide results that would otherwise have required over 100 hours of guided interviews. The Web-based test also eliminated scheduling conflicts and provided reasonable participation. Even though only about a quarter of the class was in attendance the day the test was announced, over half of the enrolled students participated.

Several subjects e-mailed comments about how interesting and enjoyable the auditory test was. The Web-based testing method eliminated any effects of pressure due to the proximity of an investigator as well as allowed for an unlimited time to complete the test. While this method produced many good results with relatively few problems, the method was not perfect. Approximately 10% of the subjects attempting the test either were not able to complete it, or had to try multiple times due to technical difficulties.

The results showed a difference between the Sound and Visual groups average correct response rates of 26%. It is evident that the auditory graphs used in this test were not as effective as visually displayed graphs. One possibility is that difference was caused by the lack of a proper introduction to the new graphing technique. Since subjects were not forced to understand the auditory graphs before starting the test, they may have found the graphs confusing.

It should be noted that had the auditory graph group been simply guessing, the correct response rate would have been about 20% instead of 41%. Thus, subjects were able to use these graphs to a limited extent even without training.

It was also evident from this pilot test that there were too few questions to provide a useful comparison between the test groups performances on linear, curved, and more complex graph patterns. Also, it was not clear from these questions how well the subjects were able to understand the shape of the graph versus their ability to draw conclusions from the graphs. Therefore, the Web Auditory Graph test used an expanded set of questions, including separate sections devoted to math or physics based graphs.

3.4 WEB AUDITORY GRAPH STUDY

The Web Auditory Graph test used the techniques described in the pilot tests and the Survey and Pre-test were identical to those of the Web Pilot. However, from the Web Pilot test it was evident that a better introduction to auditory graphs were needed for the Main test section so an introduction to auditory graphs page was presented to all test subjects before they began the test.

The Main test section however had been considerably altered from those used in the pilot studies. To determine how well subjects were able to identify graphs versus how well they could use graphs for interpretation of physical phenomena, the Main test was divided into two sections, a Math and a Physics section of 17 questions each. The Math and Physics sections had virtually identical graphs, and order of graph presentations was the same for the two sections. The rationale for having two sections of similar graphs was so that split-half analysis of the sections could be performed in order to investigate consistency and performance issues relating to identification or analysis type questions. In general, there were eight pairings of similar graph types. Thus, each graph type appeared twice in each test section. Graphs were grouped in the following categories: linear, step function, simple positive curvature, simple negative curvature, linear and curved composite, simple curved peak, complicated functions, and multiple peaked. Questions were also modified to explicitly state the data value ranges since the auditory graphs had no labels for their axes and a link to a MIDI file that contained the pitch representing zero was included with the auditory graphs so subjects could compare the zero pitch to the pitches of the auditory graphs. Graphs and the associated questions were reviewed by Math, Physics, and Science Education faculty for content validity.

The entire test was checked for compatibility with the JAWS screen reader and with Microsofts Internet Explorer. As noted in the Web Pilot, the auditory graph sound files were displayed in three formats so that users could pick the format that was most compatible with their system. The test was also checked for keyboard access to all links and text entry fields. These last issues were vitally necessary so that the blind subjects could access, take, and understand the test. Blind subjects were sent information packets containing graphs portrayed in the pre-test and

Site & Physics course	Institution type	# subjects	Approximate Course Enrollment	Date
A, algebra	University	189	350	Spring 98
A, calculus	University	2	200	Spring 98
B, agebra	Community College	4	20	Spring 98
C, algebra	University	28	44	Fall 98
C, calculus	University	8	30	Fall 98
A, Grad. Student	University	6	N/A	98
Blind	N/A	5	N/A	98

Table II. Population of Subjects taking the Web Auditory Graph test.

auditory graph instruction page as high-resolution tactile graphic images.

3.4.1 *Sample Population.* The subject population for this experiment included: undergraduate students from several institutions, graduate students to check the reliability of test questions, and several blind volunteers. The goal of this auditory graph study was to compare the performance of auditory graphs for the blind so this test included a small group of blind subject volunteers to evaluate the effectiveness of these graphs for the intended user.

Instructors of introductory physics courses at several Pacific coast educational institutions were solicited during the Spring and Fall 1998 terms for the possibility of letting their students participate in this study. It was arranged with one instructor at Site A, a large university, and one instructor at Site C, a small university, of introductory, algebra-based, physics courses to provide extra credit homework points to students taking Web Pilot test. Although there was no statistical effect on the test results, in part due to a substantial rewriting of the test questions, of the 189 subjects in the site A Algebra-based Physics course, 85 had previously taken the Web Pilot test. One instructor of a calculus-based introductory course at Site C also had her students participate for credit. An instructor of an algebra-based physics course at Site B, a local Community College, and a professor of a calculus-based course at Site A mentioned the study and Web address in class but did not offer credit for participation. Graduate students at Site A were informally solicited throughout 1998 for their participation. Six graduate students, taking the test using only auditory graphs, were used to check the test's validity.

Student subject participation from each physics course was not uniform. There were two factors for this. The most important factor for participation was the willingness of the instructor to issue extra credit for participation. When extra-credit was given for the test, class participation was generally over 50% even though the credit that subjects received played virtually no part in their overall grade. When extra-credit was not given, participation was greatly reduced. The second most important factor for participation in the Web Auditory Graph test was course size. Table II shows the distribution of the subjects among courses, schools, and approximate course sizes from which they were drawn.

Subjects from physics classes were randomly assigned to one of three test groupings. Of the 231 subjects, 74 subjects received auditory graph, 76 received visual graph, and 81 received both auditory and visual graph presentation methods. These numbers allow for statistically significant results at the $p = 0.05$ level.

A blind physics professor was consulted during test development, and acted as

Group	Number of subjects	% Correct Pre-Test	Main	Std. Dev.	Subtest Math	Phys	Avg. Time (min.)
S: Auditory	74	83	34	13	36	31	34
V: Visual	76	73	49	17	55	43	24
B: Both	81	80	46	10	51	40	30
N: Blind	5	7†	73	0	76	68	79*
G: Grad	6	92	81	9	89	73	40

Table III. Results of the Web Auditory Graph test. Results have been corrected for possible guessing. † Only one blind subject completed the pre-test. * Several of the blind subjects began the test, and then completed it several days later. The results shown in this table only include those subjects completing the test in one session.

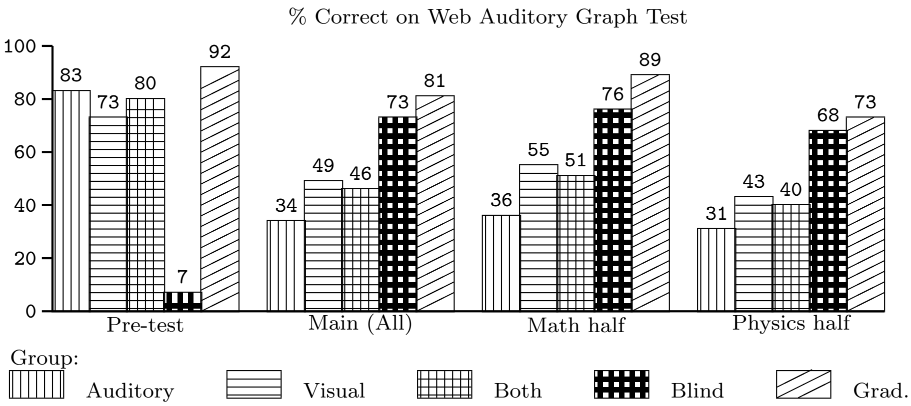

Fig. 1. Average Percent Correct per Group for Each Section Corrected for Guessing.

a critical evaluator of this study. Blind subject volunteers who had experience with college-level physics and who where willing to participate in a Web-based test were solicited by posts to e-mail lists, and through personal contact at conferences with the result that five blind subjects completed the Web Auditory Graph test during 1998. Interested subjects were sent a computer disk and Braille formatted information packets containing tactile graphs, introductory information, and the Web address location. Although this is a small number, the level of participation is a significant achievement as none of the test subjects participated locally. One of the subjects participated internationally from Europe while the other four were domestic.

3.4.2 *Results.* The responses of the auditory graph only group did not display any increasing trend (either absolute values, or relative to the Visual group) that would indicate better performance as subjects gain experience using auditory graphs. This could be an indication that the introductory material explaining the auditory graphs was sufficient for the purposes of this test.

The average values for the test sections, standard deviations of the averages, and average time for test completion are given in Table III. The average Pre-test score for the Blind group reflects the result that only one blind subject completed the Pre-test. Also, the Blind group had several subjects who started part of the test and returned a day or two later to complete the test as their schedules permitted. The table only lists the average time for two blind subjects who completed the test in one session.

The results are categorized by results from the different test groups for the Pre-test and Math and Physics sections of the Main test. Labeling for groups is as follows: S for the group with auditory graphs (Sound), V for visually presented graphs (Visual), B for both auditory and visual graphs (Both), G for graduate student subjects (Grad), and N for non-sighted subjects (Blind). For the V, B, and S groups, the limit on the error of the averages is

$$L = \left(1.96\frac{\sigma}{\sqrt{n}}\right) = 1.96\frac{0.19}{\sqrt{76}} = 0.04 = 4\%. \quad (4)$$

This result indicates a 95% certainty that a given test question is reliable to within 4%.

To account for the possibility of subjects simply guessing on answers, the % correct scores here reported have been modified by

$$C' = C - \left(\frac{W}{A-1}\right) \quad (5)$$

where C' is the adjusted score, C is the number of correct responses on the test, W is the number of incorrect responses, and A is the number of answer choices. [Nunnally 1978].

3.4.3 *Reliability Issues.* The correlation between the Pre-test and the Main test for the Auditory, Visual, and Both graph groups were low. There are several possibilities for the poor correlation scores: the Pre-test questions were not in the exact same format as the main test, the average question in the Pre-test was not as difficult as the average Main test question, the questions in the Pre-test could be skipped, and there were too few questions in the Pre-test.

Split-half correlation values between and within the Math and Physics sub-test sections were performed. Results between the Math and Physics sub-test sections of the test did not in general show good correlation. The poor correlations may reflect that the math questions were more of a descriptive choice, whereas the physics questions involved interpretation and understanding of physics principles. Within the Math subtest there all groups except the Grad group showed good correlation. The low value for the Grad group is possibly due to the small number of subjects and the low error rate to questions for this group. Within the Physics subtest there was acceptable correlation except for the Auditory Graph group which had a large negative correlation. Thus the Auditory Graph group did not demonstrate high internal consistency to this section of the test.

ANOVA calculations for the Main test, $F = 10.78$, and for the two test sections, Math $F = 12.84$ Physics $F = 5.23$ were greater than $F_{critical} = 3.04$ in all cases, these results indicate significant differences between the groups for each of the tests. Thus, detailed $t{-}test$ comparisons are valid. $T{-}test$ comparison between treatment pairs indicate significant differences between the Auditory and Visual, and between the Auditory and Both groups for the test as a whole as well as the sub-tests. The Visual and Both groups did not show a significant difference between their results.

Group	Pre-Main	Math-Physics	Subtest Math	Physics
S: Auditory	0.46	0.24	0.87	-20.15
V: Visual	0.33	0.10	0.95	0.68
B: Both	0.28	0.48	0.92	0.68
N: Blind		0.49	0.96	0.66
G: Grad		0.69	0.29	0.72

Table IV. Split-half analysis of the Web Auditory Graph test.

The test results for the Blind (N) group were very good. This group, while small, performed at substantially better, on the order of 20%, than any of the undergraduate student groups. Because of the inherent differences in the group composition and number of subjects between the Blind group and the the first-year physics student subjects, the groups can only be compared in a qualitative way and should be viewed as anecdotal. ANOVA, Sheffé or t -test comparisons between the Blind group and the student groups would not produce reliable results. A $t - test$ comparison between Grad (G) and Blind (N) groups showed that the 7% difference in the results between these groups was not significant ($t = 0.97 < t_{critical} = 1.99$). This result should be viewed with caution as the subject sample size ($n_{Grad} = 6; n_{Blind} = 5$) was very small for each of these groups. However, this comparison, perhaps more than any other test conducted in this study, demonstrates the power of these auditory graphs. The blind subjects were able to access graphical information presented in an auditory format from around the world. They were able to comprehend and answer graph-based questions at a level comparable to physics graduate students at the local test site.

A slight difference was noted in the scores on the Main test when comparing subjects in the sound group who had musical training (47 subjects) to those in the same group, but without any musical training (27 subjects). Musical training was determined from responses to a question on the Survey. Subjects with some music background had an average score of 12.9 of 33 (corrected for guessing), whereas those without music had a score of 10.0 (corrected for guessing). Since there are 27 subjects in the smaller group, equation 3 gives a measurement error of less than 8%. $F - test$ results give $F = 3.04$ and $F_{critical} = 3.97$ with $P = 0.0085$ at the $\alpha = 0.05$ level. Since $F < F_{critical}$, there is not a significant difference between the groups. A two tailed t-test also supports the null hypothesis of the two groups being equal is retained. Thus, music training seems to have a small effect, but the difference does not reach a statistically significant level.

3.4.4 *Conclusion of Web Auditory Graph test.* There was a significant difference between Sound and Visual graph groups. The main test consisted of 17 math and 16 physics questions; one poorly designed question was thrown out due to random answering. These questions had a correct response rate of 50% for the Visual graph group and 35% for the Auditory graph group. The Auditory graph group thus performed at 70% the level of that shown by the Visual group:

$$\begin{aligned}\text{Performance Ratio} &= \frac{(\text{Average Auditory Group score})}{(\text{Average Visual Group score})} \times 100\% \\ &= \frac{35}{50} \times 100\% = 70\%.\end{aligned} \tag{6}$$

The effect of a brief, self-guided, introduction and training with several examples seems to have had a substantial increase in the performance of the Sound group between the Web pilot and Web Auditory Graph tests. While these results were from first-year physics students from several institutions, the majority of subjects were from a single course at Site A. Expert physics students were able to effectively use the auditory graphs to answer questions at an average of 84% correct for the valid questions. Although a larger number of subjects would be needed to verify this finding, the performance ratio between graduate students using auditory graphs versus those using visual graphs may be as high as 87%.

Blind users demonstrated a 9% difference in average scores on the Main test when compared to physics experts. This result is not a significant difference. However, it should be noted that the 95 % confidence limit for a group of 5 subjects allows the average values to have a $\pm 18\%$ error range which would mask any significant difference between these groups. Nonetheless, it is impressive that blind subjects were able to perform about as well as graduate subjects on this test. Perhaps even more importantly, they were able to answer the questions at 73% correct. While this was not at the 81% level of the sighted graduate students, it was considerably more than 49% for the Visual graph student group.

The large number of subjects that participated in this test demonstrates the feasibility, practicality, and usefulness of using the World Wide Web as a testing medium. In addition, because the test was available via the Web, blind subjects could participate even from very distant locations. This was particularly important due to the very limited number of blind subjects who have had some training in physics. Furthermore, the results between the Sound and Visual groups demonstrate not only that are auditory graphs practical in tests, but also that they can be used to achieve performances that are within 70% of those obtained when using visual graphs. The performance results for this type of auditory graph are from a very short, self-guided training session. The new exposure to auditory graphs is an important consideration given the years of experience that subjects have had with visual graphs. While many parts of this testing process were successful, especially in terms of demonstrating that graph-based physics questions can be answered, to a certain extent, using auditory graphs, there are many areas left to explore. Such questions include: What are the best methods for portraying these graphs? What preferences do people have for sounds used in the auditory graphs? What is the limit of usefulness for these types of graphs?

3.4.5 *Subject Comments About the Auditory Graphs..* At the end of the graph test, subjects were invited to e-mail comments to the author. The following quotes are taken from those notes. They are telling as to what subjects found interesting, and which areas still needed improvement.

“Its easy to picture the graph being presented with audio tones.”

“In general your audible graphs are the greatest thing Ive heard about for a long time, and I hope you will continue to work on improving them."

"I think the whole idea is great and I think the drum beats to show curvature and slope are particularly functional and innovative. It is really important to develop the ability to hear negative values."

"I appreciate the value of getting blind users to try this and I am determined to get completely through it. By the way, did you try it blindfolded or you also blind? I want to make sure that you have gone through what I am going through (smile)!"

3.5 AUDITORY PREFERENCE PILOT TEST

The Web Auditory Graph test was an effective test using auditory graphs but many unanswered questions arose. There was an arbitrary decision of the sound types to construct the auditory graphs used in the Web Auditory Graph test. Data was represented with a piano tone, while a drum tone represented the derivative information. Any of a number of MIDI instruments could have been chosen for these representations. Also, the information for the second derivative used a high drum pitch for negative curvature, and a low pitch for positive curvature. This was a subjective choice by the author as a useful and convenient working model to begin with. There was no indication that these choices were necessarily the best ones to make.

In order to assess the effectiveness and desirability of various auditory graphing techniques, a test was developed that used a combination of pair-wise preference comparisons, graph identification questions, and Likert preference ratings. The preference questions were used to indicate which graphing styles subjects liked best, or thought were most useful. The graph identification questions were used to indicate which graphing style had the highest rate of being answered correctly.

This test was created not only to find better elements for the auditory graph displays, but also to test and evaluate an alternative method of auditory graph production. This alternate technique utilized Microsoft's ActiveX controls to create "live" graphs that have the potential for greater user control, customization, and flexibility than the prerecorded graphs could attain.

The results of this preference test can be used to guide the development of software that uses auditory graphs. The Web Auditory Graph test demonstrated that basic auditory graphs could be used for answering questions. Tests such as the Auditory Preference Pilot can be used to discover what issues should be addressed for the best optimization of auditory graphs.

3.5.1 *Data collection and Sample..* Twelve subjects participated in this study. While this was a Web-based test, due to the nature of the Microsoft ActiveX for Internet Explorer browser plug-ins required to generate and control the auditory graphs, it was decided to only use local subjects for this test. The subjects used the same computer, but at different times, for the test. Subjects included five advanced undergraduate physics students, three science and math education graduate students, three employees of the toxicology department, and an employee of the group developing the graphs. The subjects were also chosen because they had not been involved in previous auditory graph research.

Data collection was then similar to the method used in the Web Pilot and the Web Auditory Graph tests. Subjects were then presented with a series of pages containing one or more auditory or visual graphs, a multiple-choice selection field, and a text entry box for them to comment on their graph choice. There answers were recorded to a server-side data file.

3.5.2 *Instrument Development.*. The auditory graphs were produced by two methods. The first method played prerecorded MIDI sound files that used a piano instrument to represent the data values. This was the same method as was used in the Web Pilot and Web Auditory Graph tests. For this method, the data were mapped to a chromatic scale. The second method for generating the auditory graphs was with the AudioPlot ActiveX control from Oregon State University's Science Access Project. The AudioPlot (AP) control generated auditory graphs on the subjects computer from equations specified in the Web page. This method allowed various graphing parameters to be set within the Web page code. The auditory graphs produced by the AP control used a linear scale, at the programmer's insistence, to map the data to sound.

Both the MIDI and AudioPlot methods played the auditory graphs when the subject selected a play button on the page. The buttons were identical so the subject had no indication of a difference between the methods used to produce the graphs. The MIDI graphs consisted more of a staccato piano note with a courser resolution. The derivative information was represented with a drum like tone. The AudioPlot graphs produced a smooth, continuously varying tone with optional clicks for the derivative information.

3.5.3 *Data Results and Analysis.*. The purpose of this pilot test was to discover where any difficulties in the testing process may reside and to evaluate the question statements. Because of the small sample, the results should be viewed primarily as anecdotal evidence. However, tentative conclusions about the graphing methods can be made. Comparing the results to the subjects written comments about the reasons for their choices was very informative and greatly aids the interpretation of the results.

One question compared MIDI and AudioPlot (AP) representations of a Gaussian curve. The results for the question indicated a preference (58 to 33%) for the MIDI graph over the AP graph. Subjects choosing the MIDI graph mentioned that the unique piano notes "seemed cleaner," and that the discontinuous sounds produced a more dramatic effect, making it easier to distinguish the maximum point on a graph. In contrast, at least one subject preferred the AP graph because the data were represented with continuous sounds. Stevens in Mansur [Mansur et al. 1985] note that pitch has a logarithmic association with height. Thus, the linear mapping method used by the AP graphs had a perceptual effect of flattening the graphs higher pitches and may have made them seem less distinct.

Two questions investigated the pitch mapping preference for curvature, or second derivative information. The first was a comparison between a Gaussian curve graph that used a low drum tone to indicate positive curvature and a high drum tone for negative curvature with a second graph had the reverse mapping. The results were mixed with some preference for the first method (33%) vs. the second (17%). Subjects who did not show a preference said that they found the graphs confusing, or had a difficult time distinguishing between the graphs. A subsequent question compared the AP and MIDI methods for graphs incorporating derivative information. The AP graph had a score of 33% and represented positive curvature with a high pitch click, and negative curvature with a low pitch click. The MIDI graph had a score of 50% with positive curvature represented by a low pitch drum voice, and negative curvature by a high pitch drum. Thus, these two questions both support the idea of having a low tone to represent when the second derivative of a function is positive, and a high tone when it is negative. Of the subjects commenting on the MIDI graph representation, a typical response was: "the distinct sounds in [the MIDI graph] were much more clear than in [the Auditory Plot graph]." Subjects also chose the MIDI graph because of the "you can pick up the slope/curvature better."

One question specifically investigated the reactions to a change in the graphs' data sound when the y value was negative. This question was developed in response to comments received during the Web Auditory Graph test. For this representation, the data sound of the graph of $xSinx$ changed from a piano voice for positive values to a harpsichord voice for negative ones. The reasons for not preferring the change cited complaints that the tone change created "too many options for the ear to play with"and "broke up the graph a little too much." Those who preferred the tone change found it very helpful. One comment was: "I liked how the pitch changed when the graph went below 0. I think it is important to change the sound when some major distinction (like the zero line) is involved." A tone change that is more pleasing and less distracting may greatly improve its preference.

There were four graph identification questions. In one question, subjects were asked to match a visually presented graph of $e^x sinx$ to one of five AP auditory graphs. These graphs included the derivative indicators. In a second question, subjects were asked to match a visually presented graph of $e^{-x} cosx$, to one of five MIDI auditory graphs. These graphs included the derivative and negative indicators. Thus, subjects seemed to be able to match a pictured graph to its auditory representation equally well, 58% correct, with both methods, though some improvements in presentation or training could increase the correct response rate as subjects indicated that the graph choices all seemed to sound similar. The MIDI auditory graphs included drum voice derivative markers. One subject who answered incorrectly indicated that "the drums in the background created confusion as to what was going on." These subjects had not had a training tutorial as in the Web Auditory Graph test due to the varying nature of the presented graphs. Not having had training in the auditory graph presentation method seems to have had a negative impact on using the supplemental graph cues.

The two other graph identification questions, subjects were given an auditory graph and were asked to choose between several visual graphs or a "None of the Above" choice. One asked subjects to match an AP graph of $Cosx$ and had a correct response rate of 75%. The second asked subjects to match a MIDI graph of $Sinx$ and had a correct response rate of 92%. In this question, the one subject who answered incorrectly commented that the graph "seemed to mostly fit B [correct response], but I don't think the derivative was correct."

A second set of questions asked subjects to rate different auditory representations of the graph of $xSin(x)$ on a Likert type ranking scale. The results are given

Rank	Avg. Rating (1 - 5)	Graph Type
1	3.83	MIDI with negative indicator.
2, 3	3.75	MIDI plain, AP plain
4	3.41	AP with derivative indicator.
5	3.33	MIDI with negative and derivative indicators.
6	3.03	MIDI with derivative indicator.

Table V. Ranking of preferred graph types on a scale of 1 to 5, where 1 was bad, 2 was poor, 3 was neutral, 4 was good, and 5 was great. With a 4 point range (5-1) in the rating scale, a sample size of 12, and an average standard deviation for the results of $\sigma_{avg} = 1.12 = 28\%$.

in Table V There was a slight preference for the negative indicator though the averages were all within one standard deviation of each other. ANOVA analysis of the average Likert scores shows no significant difference between the methods at the $\alpha = 0.05$ level ($F = 0.82 < F_{critical} = 2.35, P = 0.53$). All the average scores were between 3 (neutral) and 4 (good) indicating that all of the methods were acceptable but could still be improved.

Finally, several subjects provided general comments on what they found helpful or annoying in the auditory graphs. These comments tended to focus on the drum beat (or clicks) indicating curvature, and the change in tone indicating negative values. A few selected comments demonstrate the greatest strengths and some potential problems with these auditory graphs:

"They all represented the graph well, it just depended on if one was interested in slope and curvature."

"I like hearing positive and negative. I like having pauses between notes instead of one constant sound. I like really hearing the slope. I don't like the soft drums because it's hard to differentiate them from the sound of the computer [fan]."

3.5.4 *Conclusion for Auditory Preference Pilot.* The Auditory Preference Pilot demonstrated some useful innovations in the development, production, and comparison of auditory graphing techniques. While the focus of this test was to provide an initial comparison of several of the assumptions used in the Web Auditory Graph test, it also provided a testing medium for a new control module that produced auditory graphs. The AudioPlot controls have the potential to provide auditory graphs with dynamic flexibility and customization for use on the Web. The results of this pilot test indicate that a variety of graphing techniques is acceptable from a user's standpoint. Also, the results indicate that some auditory graph characteristics tend to be favored by a majority, but by no means all, of the subjects and that subjects' preferences seemed to change over the course of the test.

Comments and preference choices about graphing techniques showed a favoritism toward graphs where the sounds were clear and distinct with a wide tone variation as typified by one subject's comment of: "I really like the negative value changing tone. It really helped to see the graph with my eyes closed." However, there were also indications that by the end of the test, some of the distinct display techniques became bothersome. In one of the final questions, a subject remarked: "I am starting to find the drum beats to be annoying." From comments such as this, it is evident that there is an inherent need in the design of commercial graphing displays for user configurations of the graphs. Items such as pitch range, the ability to turn on and off derivative sounds, sound transformations at the zero point, and continuous or "broken" sound playback are all important features that should be considered.

4. CONCLUSION

4.1 Summary of Test Results

4.1.1 *Triangle and Web Pilots.* The Triangle pilot test was useful for gaining experience in auditory graph methods, question development, and survey presentation techniques. The Web Pilot test was important for gauging student participation in a self-guided test. Participation in the Web based test was not a problem when the test was offered for token credit. The Web Pilot demonstrated that the PERL scripting method used to display the test and record subjects results worked well. This pilot test also demonstrated the inadequacies of the initial set of questions when used in a full comparative test. Thus, while the testing environment did not need to be modified, the questions used in the test were extended and reworked to provide a more complete comparison with a higher level of internal consistency. The difference between the performance of the Auditory graph group and the Visual graph group was 25% for the Web based pilot questions.

4.1.2 *Web Auditory Graph test.* The Web Auditory Graph test effectively compared the performances of several groups of subjects when answering graph based questions. Because subjects were randomly assigned to the different testing groups, and because of the large number of subjects in each group, the subject groups were assumed to be equivalent for analysis purposes. The Web Auditory Graph tests Main test was shown to have strong indications of validity and reliability from Split-half analysis, K-R 20 tests, and comparison of the scores of novices to experts.

ANOVA and Sheffé tests indicated that there were significant differences between the Visual and Sound groups, and between the Both and Sound groups for the Main test as well as each of the sections. After correcting the scores for the possibility of guessing, the difference between the average percent correct scores for the Visual and Sound groups was 15%. Subjects in the Sound group performed at 70% of the level of the Visual group subjects. The difference between the Sound and Both groups 13%. While these are significant differences, the results demonstrate that subjects are able to use auditory graphs to answer many math and physics questions at a fairly high level given very little self-guided training.

The Web Auditory Graph test also demonstrated that blind subjects around the world could not only access the test, but could effectively complete and answer the questions. In addition they were able to do so at a level that exceeded the student subjects, and was not significantly different from local graduate students taking the test with auditory graphs. Although small in size, the Blind group had an average percent correct score of 75%, which was considerably greater than that of the Visual group, and is not statistically different from the Grad groups score.

4.1.3 *Auditory Preference Pilot.* The Auditory Preference Pilot test was an initial attempt to determine how well subjects liked the auditory graphing techniques that had been developed, and which elements of the auditory graphs they thought were most useful. The test was as a Web-based tool displaying auditory graphs in several formats, the questions, and related visual graphs. Scripting programs served to generate the questions and record subjects answers.

The Auditory Preference Pilot test results indicate that subjects opinions of which items in an auditory graph are important change as they gain familiarity with this new graphing method. Thus, the results indicate the need for flexible graphing displays that have the ability to play the data with and without certain indicators, such as the derivative markers. It was also shown that many of the subjects found the technique of changing the tone quality to indicate when a data value is positive or negative (the zero indicator) was more helpful than the derivative indicators.

4.2 Further Studies Suggested By Test Results

The Auditory Preference Pilot test explored only a few of many areas of interest for future research on auditory graphs. One alternate avenue of research involves multivariate graphs. All graphs in these studies have used single-valued, single data sets. Construction of auditory representations for multiple data sets, for comparisons between data sets, as well as for the display of multi-valued functions needs development

Another important area for further study is an analysis of the effect of training times on performance. The relative differences between the Sound and Visual groups scores on the Web Pilot and the Main Auditory tests indicate that the amount of training plays a role in auditory graph comprehension. It is unknown how much training is required, or how training times affect the relative performance. Furthermore, there may be an effect that certain graphing techniques or indicators are only valid or useful under unique circumstances.

Since the initial pilot studies and the Web Auditory Graph test concentrated on simple graphs, and the Auditory Preference Pilot on only marginally more complex graphs, it is unknown at what point the graphing techniques used in these studies are no longer useful. As graphed data become more complex, there may be a preference for audification (directly representing the data values as a wave pattern, and then using that pattern to drive a sound source) or other data sonification methods. The auditory graphs in these studies allowed only limited control of the sound parameters. Playback rate, the ability to listen to the sound forward or backward, and point by point control of data sonification, could affect graph comprehension. Ultimately, the goal is to have the graphing method controlled by the end user when he or she is selecting the styles that are most evocative for the particular graph that is being listened to.

4.3 Practical Application of Results From This Study

Several of the auditory graphing features used in these experiments have had direct application in current software development. The Audio Graphing Calculator [ViewPlus Technologies 2007] is a scientific graphing calculator for Windows. This program is designed to display functions and data sets not only with visual graphs, but also with auditory graphs as well. The Calculator implements the use of the derivative tick-mark display as was found necessary in the initial pilot tests. It also incorporates a method for the user to enable or disable the tick-marks. In addition, the Audio Graphing Calculator incorporates a method for altering the sound quality when representing negative y-axis values. This characteristic was met with general approval from the subjects involved in the Auditory Preference Test. Thus, the effect of this research has led to significant changes in the display methodologies employed in real world applications.

4.4 Final Comments

The series of experiments described in this work has been an effort to demonstrate not only why there is a need for auditory graphs, especially in scientific areas such as physics, but how these graphs can be implemented and used. The use of auditory graphs benefits not only visually disabled people who have the right, and with these techniques, the ability for quick access of data displays, but also allows anyone to effectively use the displays with very little training. With the equivalent of a short description and a few examples, subjects demonstrated the ability to perform at a level that was at least 70% of what they would have achieved with visual graphs. With more training or experience with graphs, this can easily be increased to 85% or more. It was also demonstrated that auditory graphs are not limited to displays in research laboratories with fixed environments, but can be effectively utilized throughout the country and world.

More detailed information regarding the studies described in this paper can be found at http://sahyun.net/phd/index.html. To listen to a sample of some of the auditory graphs, please visit: http://sahyun.net/survey/sonitype.html.

5. ACKNOWLEDGMENTS

Dr. John Gardner for support instrumental in the development, execution, and application of this research. Dr. Randy Lundquist developed and wrote the ActiveX auditory graphing controls as well as the TRIANGLE program that were used in this research and Dr. Vladimir Bulatov assisted with the PERL scripting programs used to generate Web pages and record information. Financial support from the National Science Foundation under grant number HRD-9452881.

REFERENCES

Beichner, R. J. 1994. Testing student interpretation of kinematics graphs. *American Journal of Physics 62*, 750 – 762.

Brasell, H. 1987. The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. *Journal of Research in Science Teaching 24*, 385 – 395.

Cebula, D. 1997. Personal communication.

Cohn, E. and Cohn, S. 1994. Graphs and learning in principles of economics. *Research on Economics Education 84*, 197 – 200.

E., K. R. 1968. *Experimental Design: Procedures for the Behavioral Sciences*. Wadsworth Pub. Co.

Flowers, J. H., Buhman, D. C., and Turnage, K. D. 1997. Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors 39*, 341 – 351.

Flowers, J. H. and Hauer, T. A. 1992. The ear's versus the eye's potential to assess characteristics of numeric data: Are we too visuocentric? *Behavior Research Methods, Instruments, and Computers 24*, 258 – 264.

Flowers, J. H. and Hauer, T. A. 1993. "sound" alternatives to visual graphics for exploratory data analysis. *Behavior Research Methods, Instruments, and Computers 25*, 242 – 249.

Flowers, J. H. and Hauer, T. A. 1995. Musical versus visual graphs: cross-modal equivalence in perception of time series data. *Human Factors 37*, 553 – 569.

Halliday, D., Resnick, R., and Walker, J. 1993. *Fundamentals of Physics, Fourth Edition.* John Wiley, New York.

Hestenes, D. and Wells, M. 1992. A mechanics baseline test. *The Physics Teacher 30*, 159 – 166.

Hestenes, D., Wells, M., and Swackhamer, G. 1992. Force concept inventory. *The Physics Teacher 30*, 141 – 158.

ICAD 96 International Conference on Auditory Display 1996. *TRIANGLE: A practical application of non-speech audio for imparting information.* ICAD 96 International Conference on Auditory Display.

Kennedy, R. Soundmachine. http://infomotions.com/musings/tricks/manuscript/0600-0010.html

Kramer, G., Ed. 1994. *Auditory display, SFI studies in the sciences of complexity.* Vol. XVIII. ICAD, Addison-Wesley.

Leinhardt, G., Zaslavsky, O., and Stein, M. K. 1990. Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research 60*, 1 – 64.

Linn, M. C., Layman, J. W., and Nachmias, R. 1987. Cognitive consequences of microcomputer-based laboratories: Graphing skills development. *Contemporary Educational Psychology 12*, 244 – 253.

Mansur, D. L., Blattner, M. M., and Joy, K. I. 1985. Sound-graphs: a numerical data analysis method for the blind. In *Proceedings of the Eighteenth Annual Hawaii International Conference on System Sciences.* 198 – 203.

McDermott, L. C., Rosenquist, M. L., and van Zee, E. H. 1987. Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics 55*, 503 – 513.

Mokros, J. R. and Tinker, R. F. 1987. The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching 24*, 369 – 383.

Nunnally, J. C. 1978. *Psychometric Theory.* McGraw-Hill, New York.

Peters, P. C. 1982. Even honors students have conceptual difficulties with physics. *American Journal of Physics 50*, 501 – 508.

Price, J., Martuza, V. R., and Crouse, J. H. 1974. Construct validity of test items measuring acquisition of information from line graphs. *Journal of Educational Psycology 66*, 152 – 156.

Thornton, R. K. and Sokoloff, D. R. 1990. Learning motion concepts using real-time microcomputer-based laboratory tools. *American Journal of Physics 58*, 858 – 867.

Tontata. Midigraphy. http://www.vector.co.jp/soft/mac/art/se031047.html

Tufte, E. 1990. *Envisioning Information.* Graphics Press., Cheshire, Conn.

Turnage, K. D., Bonebright, T. L., Buhman, D. C., and Flowers, J. H. 1996. The effects of task demands on the equivalence of visual and auditory representations of periodic numerical data. *Behavior Research Methods, Instruments, and Computers 28*, 270 – 274.

ViewPlus Technologies, Inc. 2007. Audio Graphing Calculator.
http://www.viewplus.com/products/braille-math/AGC/

...