AN ABSTRACT OF THE THESIS OF

Steven Carl Sahyun for the degree of Doctor of Philosophy in Physics presented on April 30, 1999. Title: A Comparison Of Auditory And Visual Graphs For Use In Physics And Mathematics

Abstract approved: ______________________________

John A. Gardner

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. This study shows that useful auditory graphs of single valued *x-y* data were constructed by mapping the *y* axis to pitch, the *x* axis to time, and by including drum beats to mark first and second derivative information. Further audio enhancement was used to indicate negative data values.

The study used a World Wide Web based test consisting of a series of math and physics questions. Each question was based on a graph and had multiple-choice answers. The test instrument was refined through a series of pilot tests. The main study compared the results of over 200 introductory physics students at Oregon State University, as well as other selected subjects. A computer program randomly assigned subjects to one of three groups. Each group was presented with the same test but had a different graph presentation method. The presentation methods were: only visual graphs, only auditory graphs, or both auditory and visual graphs.

This study shows that students with very little training can use auditory graphs to answer analytical and identification type questions. Student performance for the group

using only auditory graphs is 70% of the level attained by subjects using visually presented graphs. In addition, five blind subjects from remote locations participated in this test. Their performance level exceeded that of the first-year physics students.

This work also displays the results from a pilot study of various auditory preference choices. Elements of this test may be useful for future auditory graph research and development.

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

by

Steven Carl Sahyun

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 30, 1999
Commencement June 2000

Doctor of Philosophy thesis of Steven Carl Sahyun presented on April 30, 1999

APPROVED:

Major Professor, representing Physics

Chair of Department of Physics

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Steven Carl Sahyun, Author

ACKNOWLEDGMENT

I would like to thank the many people who have supported and helped me with this work. First I would like to express gratitude to my advisor Dr. John Gardner whose support and encouragement was instrumental in the development, execution, and application of this research. I would also like to thank my committee members: Dr. Lawrence Flick who provided much help with the test designs and their interpretation, Dr. Kenneth Krane and Dr. Henri Jansen for their review of the test materials and many helpful comments, and Dr. Michael Lerner. In addition, I would like to thank Dr. Belinda King for her assistance in the design of the Math sub-test questions.

Special thanks go to Dr. Randy Lundquist who developed and wrote the ActiveX auditory graphing controls as well as the TRIANGLE program that were used in this research. I would also like to thank Dr. Vladimir Bulatov for his assistance with the PERL scripting programs used to generate Web pages and record information, and Joe Alyea, who helped with the Physics Department's computers used to run the programs.

This study would not have been possible without the aid of the many students and other people who volunteered to take the many tests. I am grateful for their participation. I would like to thank Dr. David Griffiths and Dr. Teresa Burns not only for volunteering their students, but also for encouraging participation with the incentive of extra-credit. I would also like to thank Dr. Carl Kocher and Melanie Johnson for mentioning the tests to their classes.

I would especially like to thank my good friend David Cebula, who not only suggested the use and advantages of a Web based test, but also encouraged me with my graduate studies. He also volunteered his students and himself for the tests. His comments and help have been very important in my work.

Finally, I would like to thank my parents, Irene and Mel, my brother Michael, my grandmother Geraldine, and my wife Martha for their love, support, and encouragement. I would also like to thank my son Leif for his enthusiasm, joy, and happiness.

The financial support from the Physics Department and the National Science Foundation under grant number HRD-9452881 are gratefully acknowledged.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF APPENDICES

LIST OF APPENDICES (continued)

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

Table Page

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

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 not able to view data in a visual format due to environmental circumstances.

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 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, the Web Pilot test, the Main Auditory Graph test, and the Auditory Preference Pilot test. 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 Physics and Physics Education

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 [Kra94]. 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.

Physics education research is a special case of science education research. The areas contained within science education research that most directly relate to this study are physics misconceptions, which may or may not include graphs, and graphical interpretation, which may or may not include physics.

Interest in physics education research is sufficient to warrant several publications and journals solely dedicated to this sub-field. Most noteworthy are the American Journal of Physics and The Physics Teacher, both published by the American Association of Physics Teachers. Of course, articles and research are not solely restricted to these journals, and relevant information is often found in other education and human factors journals. Physics education research has not considered the effectiveness of auditory displays.

The techniques employed in the experimental design and data analysis are common in behavioral sciences [Kir68]. 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.

When studying graphical problems in physics, it is useful to examine general problem areas that students have with physics. Students' misconceptions of physics and students' misinterpretation of graphical information are two problem areas. Fortunately, many researchers have investigated these areas.

The Force Concept Inventory (FCI) [Hes92a] and the Mechanics Baseline Test [Hes92b] are often cited studies of physics misconceptions. These studies provide instruments that researchers have used to test large numbers of students. These studies also provide extensive baseline data to assess the effectiveness of physics instruction in introductory physics. Other studies [Tro81, Tru96] also demonstrate the results from the FCI and Baseline studies. These studies probe a wide range of problems that plague physics students and they employ graphical information as part of the tests.

There have been many studies directed specifically at the nature of how students read and interpret graphs. 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 [Lei90]. The field of economics has also had research directed toward graphical comprehension [Coh94; Pri74]. Thus, results of this physics-oriented study are applicable to other disciplines.

1.3. Description of Problems That Students Have With Graphs

Many techniques of information design have evolved during the 500 years since the Italian Renaissance [Tuf90]. 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 [Hal93] is an indication of the importance that these graphs play in physics instruction. Unfortunately, students have difficulty interpreting these graphs and there have been studies devoted to finding out why students have problems analyzing information from graphs [Mcd87; Pet92; Pri74].

The ability to interpret graphical information is a prime concern in physics as graphs are widely used to give quick summaries of data sets, allow for pattern recognition, and portray information in new formats. With the ability to interpret the information contained within graphs, students have a greater chance to confront any discrepancies between what they think is correct and the data that is displayed. 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. Graphs provide concise and effective methods of displaying data so relationships between data components are more recognizable.

College physics students have a great deal of confusion, and many misconceptions, about physics. The FCI [Hes92a] 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) [Lin87; Mok87; Bra87; Tho90]. These studies focus on how immediately generated graphs of physical phenomena play a role in student understanding of those phenomena. Unfortunately, the significant gains inherent in using MBL over non-computer techniques are not usually well demonstrated. However, these studies do demonstrate improved learning.

1.4. Situations Where Auditory Graphs Are Useful

Development of visual graphs has produced methods that readily and concisely display data. However, one failing with this display method is that there is great difficulty when someone is unable to view the graph in question. There may be many reasons that a person is unable to see a graph. Several situations are that the display item (paper, computer screen, etc.) is unavailable, not within the visual periphery, the focus of attention is directed elsewhere (medical procedures, driving, etc.), or the person reading the graph has some visual disability.

When visual graphs are inaccessible, an alternate form of display is beneficial. One solution is a haptic (tactile) display of the data or graph. An example of such a display is a raised line image that a person touches with his or her fingers. Tactile representation of images and graphs can prove useful, but much time and effort is generally required when exploring each image by touch. In addition, some method of tutoring or initial orientation is required as an explanation of the image. The process of orientation to a tactile graph can be accomplished by a personal assistant or by recently developed methods of computer annotation of images [Gar98]. Tactile images are difficult to produce and require time for the reader to comprehend the picture. Since there is a permanent record of each graph, storage and retrieval of multiple graphs becomes an issue.

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. All of the auditory graphs presented in this study relate the y axis to pitch and the x axis to time.

1.5. Use 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 Auditory Graphing Calculator [Sap99], 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. BACKGROUND INFORMATION: DEFINITION OF TERMS

There are two general categories comprising most of the material reviewed and used in this study. The first category concerns various statistical methods related to education research. The second category concerns definitions and techniques relating to auditory graphs.

2.1. Statistical Methods

Experimental instruments collect data. When researchers study the ability of subjects to learn, or when they evaluate subjects' responses to environmental conditions, the instruments they use are questionnaires, tests, observations of individuals, or a combination of these techniques. Questionnaires are measurement devices for items that are not directly observable, such as knowledge, attitudes, motivations, and feelings. They are documents that ask the same questions of all subjects. Tests are structured situations measuring performance and yield numerical scores to infer a relationship between the response and the construct of interest. The results are analyzed for differences in items measured by the test. Observations are recordings of various activities where the person acting as the observer may count action frequencies, make some inference from the subject's actions, or rate the quality of those actions [Gal96, pp. 332, 767 - 774].

As with any research, use, appropriateness, and consistency of inferences made from data are important. Also, there are always questions of how reproducible the data are. These issues are discussed in terms of a test's validity and reliability.

Validity is a check on systematic errors and is an interpretation of data arising from specified procedures. Validity is also a determination of the appropriateness of an instrument for the purpose which it is used. There are three main types of validity: criterion, content, and construct. Validity of a test is established by experts who are knowledgeable in the test's subject area or by demonstrating that measurements are consistent with theoretical expectations. Reliability of a test reflects how much measurement error is present in the scores and the ability of a test to reproduce results.

2.1.1. Criterion-Related Validity

Criterion-related validity correlates performance on a test with performance of a criterion variable. An example of this type of validity is a written driver's test to determine how well a person can operate a car. Criterion-related validity can be subdivided into concurrent and predictive validity.

Concurrent validity correlates a measure and the criterion at some point in time (e.g. a report of voting behavior compared to participation in an election). This type of validity is established by comparing the results of a new test to those of a currently accepted testing method whose characteristics are well established. Predictive validity correlates a future criterion to some measurement (e.g. tests for selection purposes). This type of validity is established by comparing performance on a test to the ability to succeed in some task and usually involves longitudinal (time) studies.

2.1.2. Content Validity

Content validity is the extent to which a measurement represents the content (or conceptual domain) that the test was designed to measure. Content validity is established through examination of related literature and through expert review. Face validity is a weaker form of content validity in that it is established by casual, subjective inspection of a test's questions. In addition, face validity is established when the subjects recognize and understand the purpose of the test.

2.1.3. Construct Validity

Construct validity is a measurement of how well a test evaluates what it claims to measure. There is no single technique for establishing this measurement, rather, it is developed through multiple types of evidence. Construct validity focuses on the extent to which a measure performs in accordance with theoretical expectations. It relates a particular measurement to other measures consistent with the hypothesis. A test with construct validity will show high responses for subjects who possess a trait (e.g. ability to perform calculus) or a low score for those subjects who do not have the trait. There are three steps to establishing this validity: a theoretical relationship is specified, the empirical relationship is studied, and the evidence is interpreted in terms of how it clarifies the measurement [Car79].

2.1.4. Reliability

Measurement error is the difference between a subject's true score (the average of an infinite number of scores) and the one actually obtained. It is a statement about how predictable and consistent a measurement is. Reliability is quantified as a coefficient r that ranges from 0 to 1. The square of the reliability is called the coefficient of determination. It gives the percentage of the variance that two variables share in common. When $r > 0.7$ more than half of the variance is shared [Ker73, p. 451].

There are several methods for checking the reliability of a test: alternate-form testing, test-retest checks, and investigations of internal consistency. Alternate-form testing requires two tests with questions differing in style and content. Measurement errors are estimated by computing a "coefficient of equivalence". If a test is given multiple times (test-retest), a correlation coefficient called the coefficient of stability is used to determine how a subject's score changes between different testing occasions.

2.1.5. Determination of the correlation coefficient

To evaluate the reliability of a testing method, the most common technique is to use the product-moment correlation coefficient, also called the Pearson r value, or simply r. This is a measurement of how closely two variables are related. The correlation coefficient is a computation of the slope of a line of best fit for two data sets. It is a quantitative expression of the similarity between two data groups and is given by the expression

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \tag{2.1}$$

where the sums are over the number of pairs and $x_i = X_i - \overline{X}$ is the deviation of the mean for a data set. The quantity $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ is known as the variance of a data set, and the quantity $\frac{1}{n}\sum_{i=1}^{n} x_i y_i$ is known as the covariance between two data sets. It should be noted that whereas the variance is strictly positive, the covariance may be negative. The reliability coefficient can vary between -1 and 1, with values above 0.7 being most desirable. Values of 1 are perfectly correlated, values of -1 are anti-correlated, and values of 0 are uncorrelated [Sne89, pp. 177-195].

2.1.6. Internal Consistency

One method of determining internal consistency is by calculating a split-half correlation coefficient called the coefficient of internal consistency. The correlation coefficient is determined by grouping test questions into two parts: half of the questions form the x values, while the other half form the y values. The correlation between the two halves is calculated. However, this method represents the reliability of only half the test, which is not a true reflection of the reliability of the test as a whole. The Spearman-Brown prophecy formula adjusts the correlation coefficient to account for the whole test. The formula is given by $r' = 2r / (1+r)$ where r is the split-half correlation coefficient.

Further checks on internal consistency are through methods of rotational equivalence. Cronbach's coefficient α calculates consistency for tests that have answers with variable scores. The Kuder-Richardson formula (KR20) is a specialized form of this test and is used when test items are scored dichotomously. This test is given by

$$\text{KR20} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum_{i=1}^{n} p(1-p)}{s^2}\right) \quad (2.2)$$

where p is the proportion answering correctly. The quantity s is the estimated standard deviation of the students scores on the test given by

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}}. \quad (2.3)$$

The value n -1 is referred to as the number of degrees of freedom [Car79].

2.1.7. Other Tests of Significance: *t* and *F*

A null hypothesis is one where there is no expected difference between two groups. One method for testing the null hypothesis is with the Student's t distribution. This test measures whether two sample means are distinct when a population standard deviation is not known. The t-test is given by

$$t = \sqrt{n}\left(\frac{\overline{X} - \mu}{s}\right) \quad (2.4)$$

where n = number of samples, μ = mean of a data group, and $\overline{X}$ is the sample mean (the mean of the data groups). The calculated t value is compared to tabulated values to find the probability level. If the tabulated probability is greater than the level of significance, the null hypothesis is not rejected. The tabulated t values depend on the desired probability level that the tabulated value will be exceeded and on the number of degrees of freedom used to calculate s. The t distribution was first published by W. S. Gossett in 1908 under the pen name of Student, and perfected by Fisher in 1926. [Sne89, pp. 54, 64-71,466]. The tabulated t values can be found in many math and statistics books, or from computer programs.

If more than one comparative test is required, analysis of variance (ANOVA) is required. This procedure compares the amount of variance between groups to the variance found within groups. Analysis of covariance (ANCOVA) is a technique that combines the features of analysis of variance and regression and is used for modeling interactions with multiple classifications. If the ratio of the variance between the groups to that within the groups, called the F, or variance ratio yields a non-significant value, then use of t-tests to compare pairs of means is not appropriate [Gal96, p. 392].

$$F = \frac{\text{mean square between classes}}{\text{mean square within classes}} \tag{2.5}$$

For significance at a given probability level, the data's F value should be greater than the calculated F value. This value is found from a table of values dependent on the number of degrees of freedom (number of samples and number of comparisons) in the numerator and denominator as well as the desired probability level. F becomes 1 as the number of degrees of freedom approaches infinity [Sne89, p. 223].

The probability level p is an indicator of the confidence limits for F and t-tests, and corresponds to the probability that a sample estimate has a true value exceeding the tabulated (integral table) value. A value of $p < 0.05$ is a common level for research, $p < 0.10$ is sometimes used for exploratory studies, and $p < 0.01$ is occasionally used as a very stringent value [Gal96, p. 183].

2.1.8. Estimation of Population Size

In social contexts, researchers refer to the population. In this case, population refers to the most restrictive group from which the sample was drawn, e.g. all students, all first-year freshmen, all first-year physics undergraduate students at OSU, etc. Another consideration is the sample size. Depending on the situation, the sample will either refer to individual subjects, or to class groupings. A determination of the number of samples n is roughly given by

$$n = \left(1.96\frac{\sigma}{L}\right)^2 \tag{2.6}$$

where σ is the standard deviation of all possible sample means and L is the error tolerance. Usually L = 5% of the average value to yield a 95% confidence limit. With n given by equation 2.6, there is a 95% probability that the average value will fall within the limits set by L [Sne89, p. 52].

Conversely, given n and σ, one can find the inherent error limits on the measurement

$$L = \left(1.96\frac{\sigma}{\sqrt{n}}\right). \tag{2.7}$$

Another error arising from the measurement process is the standard error of measurement. This error determines the probable range within which the individual's true score falls and is quantified as

$$s_m = s\sqrt{1-r} \tag{2.8}$$

where r is the reliability coefficient [Ker73, p. 453].

2.1.9. Corrections For Guessing

For multiple-choice questions, a certain number of subjects may arrive at the correct answer simply through random guessing. A model that attempts to estimate how the reliability of a test is changed by random responses reduces the number of correct responses that a subject receives by a value related to the number of incorrect responses. The adjusted score is given by

$$C' = C - \left(\frac{W}{A-1}\right) \tag{2.9}$$

where C' is the adjusted score, C is the number of correct responses, W is the number of wrong responses, and A is the number of answer choices [Nun78].

Once validity and reliability for a test are established, it is important to determine if any of the differences in scores are real or are only the effects of random fluctuations. This determination is accomplished by evaluating the statistical significance of a study, which usually involves testing and rejecting a null hypothesis when it is false. Thus, if a study is statistically significant at some preset probability level, it would be a false statement to say that two groups are equal. This gives no indication or validation of the magnitude of the difference other than to reject the null hypothesis [Sne89, p. 62].

2.2. 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 [Kra94]. 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 point's value to pitch [Kra94].

2.2.1. Types of Sonification

Sonification is a term that is generally used when relating the process of converting data into an acoustic format for presentation. One method of sonification is to use data values to construct a wave pattern. The pattern can then be used to directly drive a speaker system to reproduce the sound. For example, the air pressure on a membrane can be recorded via a voltage-measuring device, and the voltage level drives the speaker. In some cases, variations that occur over a very long or very short period of time such as ground movement during earthquakes, astronomical data, or vibrational modes in structures, are used as the data waveform pattern. The data can then be time compressed or expanded so that the resulting fluctuations will drive a speaker at an audible frequency. The process of translating a waveform into the audible range is sometimes referred to as audification.

Another method of sonifying data is to let individual data points represent some characteristic of sound. Examples of sound qualities include pitch, amplitude (volume), attack (onset of the sound), note duration, decay (how the sound fades away), timbre (clarity of the note, affected by the incorporation of higher harmonics), brightness (amplitude of factors influencing the timbre), the spatial location, vibrato (a slight oscillation in the pitch, usually at about 1 to 10 Hz), or replication of noises heard in the world (doors, footsteps, sirens, bells, etc.). By combining the different qualities of sound, many researchers hope to portray multi-dimensional information in a concise manner. Relating data to a sound quality is called mapping data to sound and can be accomplished in a variety of ways.

When multiple sounds are played, e.g. two data sets represented by unique instrument sounds, the sonification is sometimes noted as having multiple voices. If the sound is located at a particular location in space, an effect produced by stereo or quadraphonic speakers, the sound may be referred to as a beacon.

2.2.2. Mapping Methods

Often, numerical data are represented by tones at specific frequencies (pitches). 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

$$\frac{x - x_{\min}}{x_{\max} - x_{\min}} = \frac{n - n_{\min}}{n_{\max} - n_{\min}} \tag{2.10}$$

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. Thus, the resulting linearly mapped frequency is given by

$$n = n_{\min} + \left(n_{\max} - n_{\min}\right)\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right). \tag{2.11}$$

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

$$f = f_{\min}\left(\frac{f_{\max}}{f_{\min}}\right)^{\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right)}, \tag{2.12}$$

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

2.2.3. Description of Auditory Graphs

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 [Man85, Flo95, Flo97] 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 auditory graphs portrayed in the Triangle Pilot, Web Pilot, and Main Auditory Graph Tests, presented data with only positive y axis values. The Auditory Preference Pilot test included graphs with negative y axis values.

It should be noted that this is a new field of research, and other methods of constructing auditory graphs are being investigated. Alternate methods include modifying the tonal qualities of the sound, repetition rates, or spatial locations [Kra94].

2.2.4. Sound File Formats

In auditory research, computers are used to produce auditory displays. The computer-generated sounds are stored as sound files that are played to subjects in a study. Two techniques are generally employed for sound generation and storage: construction of a sound's wave pattern, or the use of triggered events to play pre-stored sound samples. When the entire sound pattern is constructed, two file formats are generally used: Microsoft's wave format or Sun computer's AIFF format. These file types are often identified with the .wav or .aiff extension to their file name. Many sound producing programs, for example CSound, will construct Wave pattern files given some initial parameters. Wave, AIFF, and other similarly constructed sound file types tend to have high fidelity (the ability to reconstruct the intended sound) as they are methods of digital recording. The disadvantages of these file types are that they can become quite large and do not readily lend themselves for random access of the sound data.

When sound is constructed through the use of triggered sound samples, data is converted into a set of instructions for playing the sound samples. The General Musical Instrument Digital Interface (GMIDI) format, which is an extension of the MIDI protocol, is the most commonly used method. This popular protocol specifies certain data streams to trigger sound events such as an instrument type, and timing, pitch, and duration of a note. MIDI is a serial interface used on many musical devices such as electronic pianos and has been incorporated into personal computer sound systems. Sound fidelity is not consistent with MIDI because various computer systems employ different methods to record, store, and generate the sound samples. Another problem is sound resolution, as there are only 128 sound frequency steps available. Advantages of MIDI are that it is a common sound platform, has very small file sizes, and allows for random access of the sound file [Kra89, pp.57].

2.2.5. Other Terms

A Web browser refers to a program designed to access information pages on the World Wide Web network. The most common of these programs are Netscape, Microsoft's Internet Explorer, Opera, and Mosaic. The browsers often employ a special sub-program called a plug-in, which increases the functionality of the browser program. One such plug-in is Apple's QuickTime, which allows sound and movies in the QuickTime format to be displayed as an item contained within the Web page. Microsoft's

ActiveX is a set of control modules that a plug-in can use to enhance the ability of the Internet Explorer browser program.

3. REVIEW OF STUDIES ABOUT GRAPH PERCEPTION

There are many diverse studies related to the current research. The most relevant can be broken down into three major categories: studies about graph perception, studies on physics graph concepts, and studies on auditory graphs. While the last category is most directly related to the current research, much of the subject material and many questions used in the current study were derived from studies in the first two groups. Understanding how people perceive graphical information, and knowing which attributes of graphs are important in the visual sense, can lead to development of attributes that are also important for auditory graphs.

3.1. Leinhardt, Zaslavsky, and Stein

There is a large amount of literature devoted to the teaching and learning of graphical information. Enough literature has been published to warrant a review paper by Leinhardt, Zaslavsky, and Stein [Lei90]. Their paper described the research and theories related to the teaching and learning of functions, graphs, and graphing in high school mathematics. The functional relationships that are regarded as important constructs in the development of abstract knowledge have led to a body of research upon which their paper is based.

In their review of the literature, they take the viewpoint that there is a fundamental link between graphs and the functions that they represent. Graph interpretation is colored by the viewpoint from which it is taught or learned. The result is that students often have difficulty translating their mathematical prowess to scientific graphs, even though graphs are constantly used in science and social studies courses.

Leinhardt *et al.* report that there is no proven method for teaching graphs and functions, although there are several preferred sequences. Also, they note that the use of technology dramatically affects the teaching and learning of functions and graphs. Misconceptions may arise from students' tendency for overgeneralization, poor inferences, or incomplete learning of the material.

The complexity of the domain of functions and of graphing styles reflects the complexity of structure and of demands graphing presents both to students and to teachers. A graph's complexity is determined by the context in which it is presented, the variables utilized, and the focus of the data. A graph's complexity can influence interpretation depending on the number, type, and location of features within the graph.

There are several factors that Leinhardt *et al.* mention as hindrances to student learning of graphical information. Students can be distracted when a graph has features that can be seen as a pictorial representation of some aspect of the situation. Also, there is a tendency for students to become overwhelmed by the information presented in a new subject, making abstraction of the data representation more difficult.

Some examples of common student problems mentioned by the authors are confusion between interval and point representation, mistaking slope for height, and iconic interpretation of the graphs. Also, misconceptions are often seen when graphs contain pronounced features, such as a sudden rise or fall or a discontinuity.

Some of the difficulties that students face in understanding graphical information are due to their not viewing a graph as an abstract representation of the system under study. Resulting problems are that students give more meaning to the graph's scale than is mathematically warranted, do not comprehending the significance of the slope for a situation, and view a graph as a scene of an event.

The way that a student correctly interprets a graph often involves some level of algebraic construction in order to provide a proper interpretation. However, direct comparison of displayed points is also necessary.

When students encounter problems learning graphical information, the problem can usually be classified into one of three broad areas: a desire for regularity, a point-wise focus, or difficulty with the abstractions of the graphical world. Students also place a disproportionate emphasis on single points, such as maximum and minimum values, which distract them from other salient features such as intervals and especially slope.

3.2. Vernon

Some of the earliest studies on graph perception were by Vernon. [Ver45, Ver52a, Ver52b] The initial study concentrated on the ability of adult subjects to understand and acquire information about problems in which the information was presented in graphical formats. The information was presented in the form of charts, diagrams, or pictures rather than through verbal statements to both college students and soldiers. This initial study lead the investigator to conclude that acquisition of information becomes muddled and uncertain when the learner does not have a definite basis of ideas or background knowledge about the problem.

When knowledge is limited to a random collection of ideas, slogans, clichés, or prejudices, additional factual information may not be correctly retained. The information either is unassimilated, changed to conform to the preconceived ideas, or else remains an isolated fact that neither influences, nor is modified by, the main basis of the world view. Vernon came to the conclusion that people will tend to ignore facts when their ideas are mainly directed by emotionally toned opinions. When knowledge of the issue is vague, people are readily guided by preconceived ideas and prejudices. Thus, proper analysis of data is dependent upon sufficient education and impartiality about the subject matter presented.

The primary impetus in Vernon's later papers is to view how education affects a person's ability to interpret information presented in a graphical format. The first paper relies on the assumptions that various graphs and charts can present data accurately and vividly, and that people who see these graphs and charts can understand and assimilate the information [Ver52a, p. 22].

Vernon also claimed that a verbal argument seems necessary to provide a meaningful setting in most cases. Graphs are valuable only in so far as they can be perceived to corroborate or extend the facts upon which the argument of the text is based. [Ver52b] Useful conclusions about information portrayed in graphs or charts without written explanation can be made only when subjects have had a fair amount of education relating to the subject matter presented.

3.3. Wavering

To find what information should be contained in a new data display format, it is helpful to study the logical reasoning in graph construction. Wavering conducted such a study [Wav89] concerning line graphs. The premise was that once reasoning processes are known, student understanding of graphs can be improved. For increased understanding, graph construction and misinterpretation issues need to be addressed. The primary purpose of the study was to infer mental manipulations that students use to construct line graphs and to propose connections to theoretical mental structures.

Wavering's research design consisted of having students from grades 6 through 12 construct graphs from given numbers. The subjects wrote down information on what they were doing as they constructed the graph, and why they were doing it. Students were then asked to identify any patterns in the graphs and to state any relationships. Three types of graphs were developed having positive slope, negative slope, or exponential curves.

Wavering classified the responses to the graphing task into nine categories, determined from patterns in the responses. These categories broke down roughly into ability to draw, label, and state the relationships between variables. The ability to produce and accurately describe relationships was evaluated with categories based on developmental stages.

There are several implications from this study. First, the response categories appeared to be valid for the three types of graphs that were used. Second, student response patterns for grades 6 through 12 were similar for all instruments. Third, students in higher grades demonstrated an increased ability to provide more complete responses. The response categories were composed in Piagetian terms, with the lower categories representing concrete operational reasoning, and the upper categories representing formal operational reasoning.

3.4. Berg and Phillips

A study conducted by Berg and Phillips [Ber94] also investigated the relationship between thinking structures and the ability of secondary school students to construct and interpret line graphs. Graphing abilities were assessed through construction and interpretation of graphs with varying content and difficulty. This study again showed that students who utilized logical thinking structures, were better able to interpret questions based on the graphs, such as choosing the part of the graph with the greatest rate of change.

Other implications of the Berg study are that without "cognitive development, students are dependent upon their perceptions and low-level thinking," [Ber94, p. 340] and their responses revert to cueing off words used in the questions. With the development of mental structures such as proportional reasoning, logic overrides perception and students will no longer just see graphs as pictures, but can use them to make inferences.

An important note about the study conducted by Berg and Phillips is that they question the validity of studies that use multiple-choice instruments for determining how students learn about graphs. They advocate a testing process where students can supply their own answers and reasons for their answers. They also suggest that researchers should use a number of questions that address elements of graphing that conflict with perceptual cues.

3.5. Berg and Smith

Continuing the concerns about the validity of using multiple-choice questions for examination of the ability of students to construct and interpret line graphs, Berg and Smith [Ber94] compared the results between students' answers for multiple-choice or free response tests.

The purpose of the Berg and Smith paper was to challenge the validity of using multiple-choice instruments to assess graphing abilities, and was a report of two studies that addressed this issue. The studies were conducted on students in seventh through eleventh grades. The first study utilized numerous graphs to examine the subjects' abilities to construct and interpret graphs. The second study continued to investigate the questions of the first study and also attempted to learn about the differences in assessment when subjects drew their own graphs as opposed to selection from a multiple-choice instrument. This study compared the results for three graphing questions that asked students to either choose between, or draw, graphs representing various situations.

Their first study utilized three graphing questions that had been examined for subjects' responses to multiple-choice answers in studies by Barclay [Bar86] and by Mokros and Tinker [Mok87]. The three basic questions were modified for an interview method of data collection. The questions involved a distance vs. time graph of a person walking from and to a wall (Walk-Wall), and speed vs. time graphs of a ball rolling down a varied surface (Ball-Hill) and a bike traveling over a hill (Bike-Hill).

The second study constructed graphing instruments consisting of the three graphing scenarios used in the first study. The subjects also completed either a free-response instrument that had them draw a graph depicting a stated situation, or a multiple-choice instrument where they chose a graph to best represent the given situation. Student responses were categorized into one of three groups: correct graph, graph as a picture, or other. The response time for answering each of the questions was also analyzed.

The results from the first study showed substantial differences in the percentages of answer types between the free-response answers and those reported in the literature. This difference is what prompted the second study. Also, the responses to the first study provided categories of possible answers for scoring answers in the second study. The result of the second study was that the type of instrument used directly affected the response rate of correct answers in two of the three graphing questions studied. The free-response students drew significantly more correct responses on the Walk-Wall and Ball-Hill graphs, while the multiple-choice students chose more correct responses for the Bike-Hill graph, although the data was not presented in a clear manner. The study indicated that the percentage of "Picture Response" graphs was significantly and greatly reduced in the free-response choices.

From the results of the first study, Berg and Smith [Ber94a] concluded that the multiple-choice format used in studies might not encourage students to think through graphing questions in more than a superficial sense. During the interview method, Berg and Smith noted that the students would often answer a question quickly, but then change their answer as they explained their reasoning. Thus, the authors claim that the multiple-choice instruments often do not assess much more than superficial, first-reaction thoughts.

Berg and Smith's concluded in their second study [Ber94b] that there was a clear disparity between the results of the multiple-choice and free-response graphing instruments in terms of both correct responses and "picture as event" distracters. The results were that there was a statistically significant, 19% disparity between the response rate of correct answers between the two instruments.

3.6. Roth and McGinn

Part of the difficulty in assessing the ability of subjects to understand and interpret graphical information is that such an activity is a practiced skill. Roth and McGinn [Rot97] made several assertions to that effect in their research survey paper. They mentioned that graphs act as linguistic objects, whose relationship to the phenomena they represent is established through considerable work. The relationship holds because of convention. Students may misinterpret graphical information, not because they have not developed sufficient cognitive processes, but because they have not fully learned the conventions. Often, questions developed to provide objective responses showed signs of being socially constructed, and thus could not be pure measures of a subject's cognitive abilities. The assessment of subjects' competencies is affected by social factors (linguistics, motivation, testing climate, etc.) and cannot be held as an isolated measurement.

Roth and McGinn go on to point out that graphing ability is also a matter of practice. Since graphing is "one of an array of signing practices such as talking, writing, gesturing, drawing, or acting used extensively in scientific communities," [Rot97, p. 96] the more exposure one has to graphs, the better one can interpret their meaning. To develop graphing competence, students need to actively participate in the development of graphing practice.

3.7. Gillan and Lewis

There have been several studies concerning the methods that people use to interact with graphs in order to draw information from the images. One model of how people encode information is the Mixed Arithmetic-Perceptual model proposed by Gillan and Lewis [Gil94]. This model states that common processes people use to analyze and respond to graphical information are: searching for indicators, encoding the values of indicators, performing mathematical operations on the values, and performing spatial comparison between indicators. They performed two experiments to investigate a proposed linear relationship between response time and the number of processing steps used to analyze a graph.

Gillan and Lewis' investigation began with a questionnaire given to scientists and students asking them to recall recently used graphs and the purpose of their use. From the responses, they found that the uses for the graphs were often for quantitative purposes. To develop their categories for the perceptual and arithmetic processes, they conducted a series of task analyses of people interacting with graphs. These studies consisted of detailed verbal reports as subjects performed a task, and of observations of people answering questions about information presented in graphs. Tasks included identifying values of graphical elements, comparing the amounts of two or more indicators, summing, negating, taking ratios, determining the average values of indicators, and determining a trend.

Based on the task analyses, Gillan and Lewis decided that there was a limited set of component processes when performing frequently used tasks. The processes are: searching for the location of a data point of interest, encoding the value from the axis or associated label, performing arithmetic operations, comparing spatial relations between several indicators, responding with the answer. Their model predicted that, for different types of graphs, there is a linear increase in time to complete a task dependent on the number of processing steps. The results from their testing generally supported this model.

Gillan and Lewis concluded their paper with suggestions for reducing the time for users to make calculations relating to graphs. Of particular note are the suggestions to "organize the task so that users do not have to keep many partial results in working memory," [Gil94, p. 439] and to design graphs such that the number of arithmetic operations is minimized.

3.8. Milroy and Poulton

A related study by Milroy and Poulton [Mil74] concerns the use of labeling graphs to improve reading speed. This study looked at three techniques for annotating graphed data and the resulting time and accuracy of reading those graphs. The annotation techniques were: placing the key in the graph field, direct labeling of the lines, and placing the key below the graph. Their study indicated that for line graphs, direct labeling tended to produce the quickest readings. The authors speculated that this could be an effect that direct labeling, as opposed to use of keyed labeling, tended to reduce the amount of information that subjects had to commit to short-term memory.

3.9. Price, Martuza, and Crouse

A study of the acquisition and retention of quantitative information from a line graph was conducted by Price, Martuza, and Crouse [Pri74]. Their study was particularly concerned with three aspects of learning from graphs. These aspects were the nature of informational units, the relationship between the number of informational units and performance, and the relationship between study time and acquisition of information from the graph.

Multiple-line graphs of fictitious stock data were constructed from semi-random data that showed increasing, constant, or decreasing trends. A criterion test consisting of six sub-tests each having eight questions was used. Three of the sub-tests were based on point information and the others on slope information. The test question items were constructed using several rules to ensure balanced wording of comparatives (increased/decreased, more/less, etc.) and truth value. Two groups, differing in the length of time given to study the graph, formed the basis of the comparative study.

The number of correct responses for each item was averaged for all subjects in the separate groups. The data were analyzed with analysis of variance tests. Study time was compared to information type, number of informational units, and wording of logical opposites. This analysis showed that all four main effects were significant but that none of the interactions were. The mean score of the eight-minute study-time group was higher than that of the two-minute study-time group. Study-time and logical opposite pair wording comparisons had statistically significant results.

Price, Martuza, and Crouse explained the overall pattern of the results as not supporting their initial hypothesis that the point and slope items included in the criterion test measured distinct constructs. The final statement was that the information from data points seemed to be a more important factor than the type of information in determining a subject's performance level. They also conjectured that slope question items are more difficult than point items, and that subjects reconstructed slope information from recalled points.

3.10. Cohn and Cohn

In an attempt to find out if college students could accurately reproduce graphs shown in classes, Cohn and Cohn [Coh94] conducted an experiment with an economics course at the University of South Carolina. A second purpose of this study was to tell whether the accuracy of the graphs in students' notes affected their success on tests in which graphs were used. Lastly, the extent to which instructor handouts containing unique graphs presented only in lecture facilitated learning was also presented as a purpose of the study.

The general design of the study was to have students complete a one page questionnaire, attend an experimental lecture, and complete a post-test. Copies of the students' notes were obtained for comparative analysis. The design was essentially a post-test only control-group design on a single class of students. The comparison was having graphs provided in lecture versus the students writing their own graphs to determine which was more beneficial to the students. The authors used students' scores on exams given in class, SAT, and GPA scores as a pre-test indicator. The post-test results were compared to these values. Prior to the lecture, the students completed a general background questionnaire consisting of questions about scholastic standing, socio-economic background, and a self-assessment of their ability of read and interpret graphs.

Each student was randomly assigned an envelope consisting of a handout for taking lecture notes, and was requested to take class notes on this handout. In all cases, the handout contained an outline of the lecture. In half of the cases the handouts also included reproduction of the two diagrams shown in class. Following the lecture, students reviewed their notes for 10 minutes, and then completed a 15-item multiple-choice test. This procedure allowed a good comparison to test for the effect of teacher-supplied graphs as compared to presented graphs on short-term student learning.

The post-test consisted of two definition questions, and 13 items to test student understanding relating to the lecture. Reliability was mentioned in that all of the test questions significantly and correctly discriminated between the upper and lower quartiles of the class. Graphs drawn on the lecture notes were assessed for accuracy.

From this study, Cohn and Cohn claimed that while many students had a tendency to draw inaccurate graphs, students who drew more accurate graphs performed at a significantly better rate compared to the rest of the class. When instructor-supplied graphs for their notes were provided, students with the tendency for drawing inaccurate graphs had increased test scores. However, students who could drew accurate graphs tended to perform best with their own notes.

3.11. Analysis and Discussion

The studies reviewed in this chapter were found by the current author to be of great use when developing the auditory graph tests conducted for this work. The following discussion is a critique of how the reviewed studies helped define issues related to this work as well as some of their strengths and weaknesses. The papers reviewed in this chapter represent a selection of issues related to the use of graphical information. As noted by Leinhard *et al.* students often have difficulty viewing graphs as abstract representations, and instead focus on graphs as pictures of a situation. The other reviews of this chapter also contain information and subject matter that was useful in the development and application of the conducted studies.

Perhaps the most important conclusion that Vernon's three studies demonstrated was that education played a role in how well a person could derive factual information from graphs and charts when they were presented without written text. Another important conclusion was that there seemed to be no advantage in using pictorial charts rather than graphs to portray the information. Even when specific factual data are understood, it is often difficult for people to incorporate this new information into their general body of knowledge.

In the current studies comparing auditory and visual graphs, efforts were made to choose subjects who were compatible with the subject matter that was presented in the graphs. This was accomplished by designing a test and graphs from material that first-year university physics students might encounter in the course of their studies and then choosing first-year students as the test subjects.

The use of graphical material as a testing medium is valuable. Wainer asserts that "graphs work well because humans are very good at seeing things, they are so basic to our understanding that we cannot easily imagine a world without them." [Wai92, p. 15] However, one must be careful in the presentation of information. Evaluating performance based on information that is presented in a flawed format can be misleading. When data are presented in a properly displayed graphical format, most common questions can be easily answered, and deeper analysis of data can then follow. The reasoning for creating alternate displays of information is that a better graph of the same data should make interpretation easier.

The important point from Wavering's study is that by grade 10, a majority of students were able to interpret graphs at a level that is consistent with formal operational and early correlational reasoning. However, Wavering's claim would have been strengthened if much larger sample sizes had been used to provide better statistical results. The consequence of Wavering and Berg's research studies is that it is not unreasonable to expect first-year college students to answer questions based on complex relationships of graphical data. The subjects have had full cognitive development since at least grade 10. Thus, the first-year students used in the conducted studies have the abilities needed for pattern recognition and recognition of relationships between variables.

While Berg and Smith raise concerns about the appropriateness of multiple-choice questions when investigating graphical information, the current study's goal was instead to determine how well students can answer questions based on graph differences. Thus, reliance on free response answers was greatly reduced. In addition, a limitation of the Berg and Smith studies were that only three graphing questions were investigated. In one of those cases, the reported data was obscured suggesting that results may not have completely agreed with the author's conclusions. Also, the free-response answers were grouped in categories, with "correct" answers being rather loosely defined, thus allowing percentages to be manipulated to the authors' advantage. Their conclusion that the disparity in testing methods may be important when considering factors for test construction. However, a free-response method is difficult to implement on a large scale and may be not be an important factor when identification between presentation methods of graphs is the main goal. Also, while there was a reported difference in success rate with response time and instrument utilized, it was not demonstrated that this was a significant effect.

Studies by Gillian and Lewis and by Milroy and Poulton noted the importance of labeling graphs to increase their ease of use. This is an issue for the current study as labeling auditory and visual graphs in an equivalent manner is important.

Price, Martuza, and Crouse noted that students tended to focus on specific points when interpreting graphs. However, most of their study dealt with giving students variable time to memorize data. Perhaps much of the effect that was seen was the ability to remember the smallest useful unit of the data. The more time that was allocated allowed the students to remember more of the data points, and thus gain a better interpretation of the graph.

While Cohn and Cohn's study demonstrated the interesting effect that better students performed at a higher rate on tests when using their own graphs and poorer students performed better using teacher-supplied graphs, the use of SAT scores to determine the student performance categories is problematic at best. Their study could have been strengthened if a more appropriate pre-test method had been included. Such a test could have been easily accomplished as test items could have been included in the course's normal exam structure.

4. REVIEW OF STUDIES CONCERNING PHYSICS, GRAPH CONCEPTS IN PHYSICS, AND COMPUTER AIDS IN GRAPHING

4.1. Hestenes, Wells, and Swackhamer

An important study was conducted by Hestenes, Wells, and Swackhamer [Hes92a] concerning a method to probe student beliefs on the concept of force and how they compare to the Newtonian concept. The Force Concept Inventory (FCI) is a multiple-choice test instrument that provided choices between correct and commonsense alternatives to Newtonian concept questions regarding aspects of force. This test has been given to many high school and college students and generated much literature. The primary uses for the FCI are as a diagnostic tool for student misconceptions and for evaluating instruction on Newtonian concepts.

The FCI is important because several of the questions in this well-researched test served as the basis for questions asked in the auditory graph tests. Not all of the questions could be utilized due to the nature of the display format, and the FCI covered only material relating to the concept of force whereas the conducted study had a more general basis of questioning. More will be mentioned of how this test was adapted in the section on Experimental Design. The Mechanics Baseline Test is similar instrument by Hestesnes and Wells [Hes92b]. Several questions from this test were adapted for use in the auditory graph study.

4.2. Trowbridge and McDermott

There are several studies that try to characterize how students conceptualize motion and the role that graphed information plays in their understanding. The first, by Trowbridge and McDermott [Tro80], looked only at how students understand velocity of simple observed motions. This paper described a guided interview process with over 300 subjects. The subjects were asked a series of questions relating to demonstrations about the motion of simple objects. In several of the questions, subjects were asked to compare the speed of two objects. When responding to one of the questions, some students would spontaneously draw graphs to aid as a communication device. However, it was observed that students were unable to correctly incorporate their graphing skills into a successful understanding of velocity. From student responses to interview questions, it was stated that students have a disparity between what their graphs illustrate, and what they think their graphs illustrate. It is this disparity that provided an expanded study with additional research.

4.3. McDermott, Rosenquist, and van Zee

The expanded study by McDermott, Rosenquist, and van Zee [Mcd87], looked not only at velocity, but also at kinematics as a whole, and how students had trouble connecting physical concepts and graphical information. Their descriptive study with several hundred students involved identifying areas in which students have difficulty in their interpretation of graphical information. Their data were derived primarily from responses to questions given to the students, presumably as part of an exam. The results were mainly a categorization of the more prevalent difficulties observed.

There are two main areas of difficulty that were identified: connecting graphs to physical concepts, and connecting graphs to real world phenomena. In the first category, the identified problems were: difficulty discriminating between the slope and height, interpretations of changes in slope and height, relating graphs between position, velocity and acceleration coordinates, matching narrative information with relevant features of a graph, and interpretation of the integral, or area under the graph. In the relation of graphs to the real world, students drew graphs relating to the motion of a ball on various tracks. From these graphs common problems were: an inability to represent continuous motion with continuous lines, separating the shape of the graph from the path of the motion, representing negative velocity, representing constant acceleration, and distinguishing among different types of motion graphs (x, v, and a vs. t).

4.4. Beichner

A comprehensive study about students' interpretation of kinematics graphs was performed by Beichner [Bei94]. The primary purpose of this article was to report on a study aimed at uncovering student problems with interpreting kinematics graphs. A secondary purpose was the proposition of a model for creating research-oriented multiple-choice tests that could be used as diagnostic tools or as formative and summative evaluations of instruction. Parts of the multiple-choice test that were developed in the Beichner study were used as question templates for the current research

The test evolved in several parts. Draft versions of the test were administered to 134 community college students who had been taught kinematics. The results were used to modify several of the questions, and the revisions were given to 15 high school, community college, four-year college, and university science educators. These individuals completed, commented on the appropriateness of the objectives, criticized items, and matched items to objectives in an effort to establish content validity. The final tests were then given to 165 juniors and seniors from three high schools and 57 four-year college physics students.

The test instrument consisted of 21 multiple-choice questions divided into seven testing objectives. The objectives were chosen upon examination of commonly used test banks, introductory physics books and informal interviews with science teachers. The test was designed to focus on interpretation skills. Three test items were written for each objective, most of these being written by the author although some items were adapted from previously used tests. The test questions and results of student performance were appended at the end of the paper.

All of the statistical procedures indicated that the test was valid and reliable. Results of data analysis also indicated several other results. First, calculus-based physics students did significantly better on the test (mean of 9.8 vs. 7.4) than algebra/trigonometry-based physics students (t = 4.87, $p < 0.01$). Second, college students were not significantly better than their high school counterparts (t = 1.50,

$p < 0.13$). Third, the mean for males of 9.5 was significantly better than the 7.2 mean for females ($t = 5.66$, $p < 0.01$).

The developed instrument appeared to be generalizable to a wide range of students studying kinematics, from high school to university courses, across the country. The results allowed for objective grading and the ability to provide statistical analysis from large numbers of subjects.

In a later study, Beichner [Bei96] investigated the impact of students analyzing video motion on their ability to interpret kinematics graphs. In this study it was found that the greatest impact on student's ability to interpret graphical information comes from hands-on involvement in data acquisition. The study demonstrated a strong correlation between the amount of exposure to video graphing labs and students' scores on a multiple-choice test on graphs, indicating a better understanding of kinematics graphs.

4.5. Mokros and Tinker

A set of studies by Mokros and Tinker [Mok87] demonstrated that middle-school students could learn to communicate using graphs in the context of appropriate microcomputer-based laboratory (MBL) investigations. The first preliminary study attempted to locate graph-related misconceptions, the second investigated children's graphing skills, and the longitudinal study examined MBL intervention.

In the first study, 25 seventh and eight grade students in a suburban school participated. The students were given a carefully constructed set of graphing problems in an interview setting. The problems were developed from the results of a pilot test to ensure appropriateness in terms of language, difficulty level, and coverage of various problem types. The interviews consisted of six graphing items and lasted 20 to 40 minutes. A protocol summary was completed for each student's performance. The findings of this study were that students exhibited two major types of errors, which have also been observed in college populations: graph as picture confusion and a weaker indication of confusion with relating slope and height.

The second study investigated the ways in which students learn graphing skills through MBL. Data were collected by observing individual lab groups. Students' interactions were recording as narrative summaries and by an event sampling process that was subjected to quantitative analysis. The study utilized an MBL course unit consisting of five days of activities on position and velocity plotting. The observations and scores from a nine-question quiz in the second preliminary study indicated that after five days, students had developed graph interpretation skills.

The longitudinal study was designed to provide more evidence about the impact of MBL on graphing skills. This study involved a pre-test, treatment, post-test design, with each test having two components: a multiple-choice test of graphing skills and an interview where the students talked through their thought process. In the longitudinal study, scores on the 16 graphing items showed a small (Δ = 15%), but significant, improvement. This research showed that students could learn graphing concepts over a long time frame when using MBL's.

4.6. Brasell

A study by Brasell [Bra87] not only extends Mokros and Tinker to high-school students but also assesses the effect of a very brief exposure to a kinematics unit on the ability to translate between a physical event and the corresponding graphical representation. The study also evaluated the effect of real-time graphing in comparison to delayed graphing of data on student learning.

The sample was drawn from entire physics classes (of seven to 17 students each) in seven rural schools in north Florida providing a total of 93 students. The students were mostly seniors and were familiar with the concepts included in the experimental activities. It is suspected that the choice of the students was a matter of convenience as the author is from the University of Florida.

The experiment was conducted over a three-day period, one day for the pre-test and orientation, one for the treatment, and one for post-testing and discussion. The treatment consisted of several groups: a Test only, a standard MBL display where data were displayed as it was acquired, a delayed MBL group where a 20-second delay was introduced between acquisition and display, and a pencil and paper graphing group that plotted their own graphs on paper. The MBL groups used curriculum units designed for the software. The paper and pencil group graphed complex motion described on a worksheet. Each class at each school had one group of students for each treatment to provide a balanced design. Students were randomly assigned to each group on a class-wise basis.

Pre- and post-tests were described as consisting of content-specified, multiple-choice items requiring students to translate between a verbal description of a physical event and the graphic representation of it. The pre-test had been developed and used by a previous researcher for use with humanities college students. The post-test was conceptually similar to the previous study, but altered in format. Due to the format change, performance changes were utilized only as a covariant. SAT scores were recorded and used as a covariant. It was stated that neither the pre- nor post-tests were checked for reliability. Validity of the tests was not mentioned. Analysis of covariance was used to reduce error variance of post-test scores.

Factorial analysis of covariance was utilized. The pre- and post-tests were divided into two sub-tests, one for distance and another for velocity. It was found from F tests using 3 treatment degrees of freedom, and 68 degrees of freedom for the data, that overall scores for standard MBL treatment were significantly higher than scores from the other treatments (F (3, 68) = 6.59, $p < 0.001$). While it was shown that scores for both sections were higher, only the distance sub-test scores were significant with F (3, 68) = 6.47, $p < 0.001$. The velocity sub-test was not considered a significant difference, F (3, 68) = 1.80, $p = 0.156$. A table of the results as well as a graph of the mean error rates for the different groups were presented.

Brasell stated that 90% of the difference in the mean scores was due to the real-time nature of graphing provided by MBL. At no time was the performance of the delayed MBL graphing significantly superior to that of students in the control groups. It was found that even a short delay in displaying graphs dramatically reduced the effectiveness of the MBL on graphing skills. It was suggested that one of the effects of the delayed graphing was that students appeared less motivated, less actively engaged, less eager to experiment, and more concerned with the procedure, rather than the concepts.

4.7. Linn, Layman, and Nachmias

A study on the cognitive consequences of microcomputers on graphing skill development was attempted by Linn, Layman, and Nachmias [Lin87]. In their study, they explored how students' graphing skills changed after exposure to MBL intervention. Their study centered on an "ideal" chain of cognitive accomplishments. These were: graph features, graph templates or sequences of activities that are used repetitively to comprehend the graph, graph design skills which augment and consolidate the templates for new problems, and graph problem-solving skills. They found that the MBL intervention increased student's ability to identify trends and locate extrema, but did not compare their results to non-MBL methods. Exposure to the MBL graphs acted as a basis on which students built their graphing models.

4.8. Thornton and Sokoloff

A study that did attempt to compare the effectiveness of MBL techniques was conducted by Thornton and Sokoloff [Tho90]. The purpose of their study was to compare the effectiveness of curricula that take advantage of MBLs presenting data in immediately understandable graphical forms to the effectiveness of non-computer based courses. The ability to learn basic kinematics concepts was evaluated with pre- and post-testing as well as by observations.

The sample was drawn from more than 1500 college and university physics students taking non-calculus and calculus based General Physics courses at the University of Oregon and Tufts University over a three-year period. The research design consisted of testing students enrolled in a laboratory course involving microcomputers to display the graphical information and comparing the results from their post-test scores to those of students who were not enrolled in the lab. Data were collected by 50-item multiple-choice pre- and post-tests. It was not mentioned if the same test was given at both universities. The reported data showed dramatic reduction (up to 40%) in the error rates when compared to the non-MBL group.

4.9. Analysis and Discussion

The studies reviewed in this chapter concerned the interrelationship of how students learn physics when using graphs and the use of computers to display graphical information can affect student learning. Perhaps the most important studies with regards to the development of the auditory graph tests used in this work were those by Hestenes *et al.* and by Beichner.

The FCI questions were concerned with determining where students were having difficulties in physics and were more focused in their subject matter than those used in the current study. Beichner's study was of even greater aid in question development as it reported on a multiple-choice test involving kinematics graphs. While not all of the questions from these papers were compatible with the current research, they were valuable templates upon which to build the physics multiple-choice auditory graph questions. McDermott's studies were useful in their focus on understanding where students have difficulty, especially in the areas of connecting graphs to physical concepts and distinguishing among different types of motion.

Since the current research utilized computer portrayal of graphical information, some discussion of the research investigating how computers have played a role in graphing was included. These studies were valuable as they also provided a basis from which to draw material for questions used in the current studies.

While the MBL studies indicated that learning had taken place with the use of computer generated graphs, a major shortcoming of all these studies was the lack of comparison to equivalent non-MBLs. For example, in the study by Thornton and Sokoloff students who did not participate in the microcomputer lab did not participate in any lab experience, hence were not as practiced as the MBL group. In addition, for some of the subjects, the lab was a separate, and an optional course, so the students who took the MBL may have been self selected for better performance. Another explanation is that those students not taking the lab may not be as comfortable, practiced, or competent with physics as the MBL group, which would also cause a difference in scores between groups. These studies are useful however, as they demonstrate the prevalence of computer use for graphing in current physics courses. In all these studies, the students were comfortable with computers as tools for displaying information.

5. REVIEW OF STUDIES ON AUDITORY GRAPHING TECHNIQUES

There is a large field devoted to the representation of data with sound. Generally, this field falls under the heading of Auditory Display and can encompass a wide range of sound representations such as the use of auditory cues ("earcons") as locators to more direct representations of data. The field is large enough for conferences such as the International Conference on Auditory Display (ICAD) with published proceedings [ICA94].

The quest to find a useful auditory data display has been approached from many fields such as mathematics, chemistry, computer science, as well as physics. From the diversity of auditory display techniques, it is readily apparent that no single display will suffice as a universal presentation method, just as no single visual graphing method works for all data. The following studies are those that directly relate to auditory techniques that would otherwise use two-dimensional plots.

5.1. Pollack and Ficks

One of the first studies concerning auditory display of information was performed by Pollack and Ficks [Pol54]. In their paper they investigated the relationship between auditory display stimuli in order to find a satisfactory procedure for increasing the information that can be transmitted from elementary auditory displays. The basic task of their subjects was to identify different qualities of sound stimuli. There were eight sound qualities tested using tones and noise: frequency ranges of noise and of tone, loudness of noise or of tone, rate of alternation between noise and tones, duration of tone display, the fraction of time tone was on, and direction of origination of the tone. Subjects were students and military personnel. The sounds were binary coded, in that the tones were either high or low, alternation rates were fast or slow, sound intensity levels were loud or soft, etc. In half of the tests subjects responded as they listened to the display, while in the other half, they responded after the sounds finished.

Pollack and Ficks reported that their subjects found the multidimensional displays easy to learn, especially the binary coded displays, and that subjects tended to associate the sounds with verbal symbols (e.g. "chirping birds"). They also reported that the multidimensional displays were able to effectively transmit more information than unidimensional displays. However, there was little improvement in information transmission when the dimensions were subdivided (degrees of loudness or alternation rates). The average error in correct identification of the auditory dimension was lowest for the binary comparison of frequency of the tone, at 0.08%. This rate was dramatically lower than for the other dimensions studied. The next lowest values were for sound duration (0.9%) and repetition rate (1.1%).

Their conclusion was that the use of multiple stimulus dimensions is a satisfactory method for increasing the transmission of information via auditory displays. Another conclusion was that it is more useful to have a greater number of binary coding dimensions rather than subdivision of only a few dimensions.

5.2. Mansur, Blattner, and Joy

Mansur, Blattner, and Joy [Man85] reported on a very significant study for representing data by sound. Their study, which generally provided the template for the current investigation, used sound patterns to represent two-dimensional line graphs. They were investigating a prototype system to provide the blind with a means of understanding line graphs similar to printed graphs for those with sight. This study used auditory graphs that had a three-second continuously-varying pitch to present the graphed data. The auditory graphs were also compared to engraved plastic tactile graphical representations of the same data. The authors cited research by Stevens, Volkmann, and Newman [Man85] on the pitch response of hearing that showed an exponential relationship between pitch and perceived height.

Mansur, Blattner, and Joy found in their study that there were difficulties in identifying secondary aspects of sound graphs such as the slope of the curves. They suggested that a full sound graph system should contain information for secondary aspects of the graph such as the first derivative. Their suggestion was to encode this information by adding more overtones to the sound to change the timbre. They also suggested utilizing special signal tones to indicate a graph's maxima or minima, inflection points, or discontinuities.

Their main study consisted of several comparison tests to indicate the effectiveness of sound versus tactile graphing methods. These consisted of comparing the slope of lines, straight vs. exponential lines, monotonicity, convergence, and symmetry. There were fourteen subjects, half of whom were blind. The sighted subjects were blindfolded for the tests. The subjects were tested with one presentation method, and then re-tested with the other method. The type of graph subjects received first was by random assignment.

The results were that the tactile graphs had a small, but statistically significant, advantage to the sound graphs in overall accuracy (88.3% vs. 83.4%). This disparity appears to come mainly from the comparison of straight lines vs. exponential curves where there was a 12% difference in the accuracy of identification (96% vs. 84%). Also, a test of whether a graph was converging to some limiting value had a 9% difference in the scores (89% vs. 80%).

5.3. Lunney and Morrison

Lunney and Morrison [Lun90] describe an auditory alternative to visual graphs in order to provide access to instrumental measurements. Their system was to convert infrared chemical spectra into musical patterns. The translation method first converted the continuous spectral pattern into a "stick spectrum" in which absorption peaks are replaced with lines representing location and intensity. The spectrum was then mapped to a chromatic scale with the infrared frequency converted to pitch. The sound map was played in the form of two patterns. The first pattern was to play from highest pitch to lowest, with intensity represented by note duration. The second pattern was to play the spectrum in order of decreasing peak intensity, with equal note duration. The first pattern was played twice, and the second three times. The six strongest peaks were also played together as a chord at the end. The authors mentioned that this was an effective technique for chemical analysis of spectra.

5.4. Frysinger

A review paper by Frysinger [Fry90] details various research approaches to data sonification. The bulk of his review describes data sonification, the areas of psychoacoustics (the psychology of hearing), and sound perception issues. Several of the articles that were reviewed are summarized above. Frysinger provided some very general indications for research direction and methodology. Some of these suggestions are utilization of synthetic data generation to control parameters, the use of two sessions to compare the effectiveness of two display types, and using forced choice type questions.

5.5. Minghim and Forrest

For more complex sound mappings, Minghim and Forrest [Min95] presented a review of several studies and an analysis of data sonification development. They mentioned the following areas where sound can be a useful tool in aiding data visualization: adding further dimensionality to data, alternate perceptual properties, additional interactive processes, inherent time dimension for data, use of sound as a validation process, and increasing the ability to remember data due to additional modal encoding. They also described a sonification program called SSound which implements a number of sound functions for aiding surface-based data analysis. Various surface properties were mapped to sound qualities such as pitch for density, rhythm for change in a function, and timbre for data correlation. Sound was spatially located using quadraphonic speakers to indicate information depth. Users of this sonification system required training for interpretation of the complex sounds. The authors did not report any formal results as to the effectiveness of the system.

5.6. Wilson

A similar program to represent data by sound is the Listen data sonification toolkit described by Wilson [Wil96]. The primary goal of this program was to provide a flexible sound toolkit for use in sonification research. The Listen program is an object-oriented modular system designed on Silicon Graphics (SGI) workstations incorporating MIDI sound libraries. Listen was designed to be a component for incorporation into other data visualization programs. The main modules of the Listen program are: Interface, Control, Data Manager, Sound Mapping, and the Sound Device modules. Only the Interface module interacts with the Control module, which then interacts with the other three. With this program, data fields can be mapped to four types of sound parameters: pitch, duration, volume, and location. Pitches used the semitone scale. Data could also be given timbres relating to various MIDI instruments for further diversification.

5.7. Flowers and Hauer

There are several important studies relating to the success of auditory graphs for display purposes. Flowers and Hauer produced a set of studies investigating the perceptual similarities between visual and auditory graphs.

The first paper [Flo92] described a single experiment to study how effective information about central tendency, variability, and the shape of data distributions could be portrayed with auditory graphs versus a visual graph. Data in this experiment were presented as auditory histograms, auditory quartile displays, and visual histograms. The auditory histograms presented the data distribution with the numeric value mapped to pitch, and the frequency of the data mapped to the number of times a note was repeated. The visual histogram was presented on a computer screen as text characters, with the numeric value mapped to the x axis, and the frequency of the data distribution was presented with vertical stacks of asterisk symbols. The auditory quartile displays were a musical analogue of the Tukey box and whiskers drawing that coded the minimum, first, second, and third quartile, and maximum data values as a set of five musical notes.

Twelve psychology graduate student subjects performed 132 comparison trials in each of three sessions, with only one presentation modality per session. The subjects gave a 1 to 10 similarity judgment rating for each of the graph comparisons. The judgments were based on differences in central tendency, variability, and the shape of the data distribution.

This study specifically investigated the perceptual structure of plots through dissimilarity judgments of a graph's slope or level when depicted by visual versus auditory displays. The study consisted of three tests, labeled Experiments 1, 2, and 3. Experiment 1 investigated student's ability to distinguish visual graphs, while Experiment 2 investigated auditory graphs. Experiment 3 was similar to 1 and 2 except that it provided a more sensitive evaluation between visual and auditory graphs. Results showed that the correlation between judgements and stimulus parameters for the auditory histogram ($r = 0.36$) and quartile display ($r = 0.40$) graphing techniques produced a far greater dissimilarity rating than did the visual histogram ($r = 0.06$) graphs. However, the opposite was true for skew ($r = 0.11$, 0.06, and 0.39) and kurtosis (presence of long or short distribution tails, $r = 0.07$, 0.02, and 0.21). Judgments on the range of data values were similar for all three graph types. The authors commented that the surprisingly low correlations between the dissimilarity judgements may have been related to little variation in the standard deviations for the distributions used as stimuli.

The second study by Flowers and Hauer [Flo93] extended the first with two experiments investigating whether combined auditory and visual presentations enhanced discrimination of stimulus parameters, and whether the auditory quartile (Tukey box and whisker) plots provided an adequate distribution of information. In the first experiment, 25 paid student subjects, with normal hearing and vision, participated in a study similar to that conducted in their previous paper on comparative judgment analysis of visual and auditory histogram graphs. There were three display methods: visual presentation, auditory presentation, and a combined auditory and visual presentation, with the auditory histograms using the same method as previously described. Their results showed that visual graphs again had a greater reliability in the dissimilarity judgments than auditory graphs, and that there was no evidence that combined presentation led to a greater consistency of judgments than visual presentation alone. The second experiment consisted of the use of auditory quartile displays, slightly modified from the previous study in that these displays had an additional leading note, representing the median as a prefix to the five-note system. A comparison of dissimilarity judgments between the original quartile display method, and the leading note prefix method showed a greater attention to the median ($r = 0.58$ vs. 0.20) but reduced attention to skew and range ($r = 0.20$, 0.31 vs. 0.38, 0.41). Thus, focusing on the central tendency came at the expense of other characteristics.

In their third paper, Flowers and Hauer [Flo95] compared the perception equivalence between auditory and visual graphs and the ability to convey information regarding the profile of changes of an independent variable. At least two of the samples consisted of introductory psychology students at the University of Nebraska. In Experiment 1, there were 18 students (7 male, 11 female) who received credit for a research exposure requirement for their introductory course. Experiment 2 consisted of 14 student volunteers who were each paid $15. Experiment 3 consisted of two groups of students who were in a similar situation as those in Experiment 1. It was not stated if students in one experiment were also in another, or what size of a class population that these students were drawn from; thus the number of students involved could be from 19 to 51.

There was some discrepancy between the methods for comparing the graphs in Experiments 1 and 2. In the first, students were instructed to sort 68 graphs into no fewer than three and no more than 10 categories. In the second, students used a pair-wise numeric (1-10) dissimilarity rating procedure of all possible pairings of 34 of the 68 graphs. Half of the subjects (seven) received one pairing set, and the other half compared a second set. The auditory graphs used the same data sets as the visual plots.

In Experiment 3, 16 graphs were used for comparison purposes. In this trial, both the visual and auditory graphs were compared in a pair-wise fashion. The visual graphs were displayed at the same time while the auditory graphs were displayed sequentially.

The authors' conclusion was that the experiments illustrated a close correspondence between the perception of auditory and visual graphs with regards to gross differences in function shape, as well as slope and level (height) perception. The main result of this study was to demonstrate how to use auditory graphs to convey information about distribution central tendency, variability, and shape to observers who had not been previously exposed to auditory representations of data.

5.8. Turnage, Bonebright, Buhman, and Flowers

Turnage, Bonebright, Buhman, and Flowers [Tur96] reported on a study, comprising of two experiments, comparing the equivalence of visual and auditory representations of periodic numerical data. The first experiment investigated whether equivalence of auditory vs. visual presentations of wave form stimuli would parallel that reported for other graph types. Twenty-six undergraduate psychology student subjects participated to fulfill a course research requirement. The subjects were divided into two groups of 13. Graphs consisting of 100 data points were constructed with three shape patterns (sine, square, or combination), three frequencies (high - 8 cycles/100 data points, medium - 6/100, and low - 4/100), and two amplitudes (high, and low) for a total of 18 graphs. The visual graphs were constructed with Microsoft Excel and presented via overhead transparencies. The auditory graphs were played as a series of 100 musical notes with a two-octave range. The y axis was represented by pitch and the x axis as time. Each auditory graph had a 6-second duration.

The subjects were presented with the task of providing similarity ratings for all independent pairs (153) of graphs. They were initially presented all 18 graphs in random order and had three practice tries for familiarization to the process of discriminating graphs. They then rated the graph pairs on a 9-point similarity scale for each of the three conditional dimensions (1: shape, 2: amplitude, and 3: frequency.) Coefficients of congruence (CC), interpreted like correlation coefficients, revealed that the visual and auditory graphs were very similar for all three condition dimensions (CC 1= 0.96, CC 2= 0.98, CC 3 = 0.94.) Thus, the two graphing methods have high similarity for difference discrimination. There was also some indication of slightly greater discrimination between sine and composite wave patterns with the auditory display than with the visual display.

The second experiment investigated the relative performance accuracy of visual and auditory graphs on a task involving discrimination between similar wave forms. Thirty-eight undergraduate psychology student subjects participated to fulfill a course research requirement. The subjects were divided into two groups of 19, for each graphing method. The graphs were constructed and presented as in the first experiment. Forty pairs of wave form graphs were selected for a comparison task. Subjects were sequentially presented with two graphs, A and B, and then presented with a third graph, X, from which they determined whether *X* was the same as A, B, or neither. The subjects were given three practice trials for familiarization. Results showed a significant difference in the performance scores of the two groups with the Auditory graph group average of 81% correct, and the Visual graph average of 96% correct.

5.9. Flowers, Buhman, and Turnage

Most recently, a study relating to auditory graphs for display purposes was conducted by Flowers, Buhman, and Turnage [Flo97]. This study investigated the equivalence of visual and auditory scatter plots to explore bivariate data. Their study consisted of two experiments, the first examining the relationship between visual and auditory judgments for the direction and magnitude of correlation for 24 bivariate data samples.

The first experiment used 45 unpaid advanced undergraduate psychology student volunteers. Nineteen of the subjects, in groups of three to eight, judged visual scatter plots of data samples, while the remaining 26 were assigned in groups of five to 16 to judge auditory scatter plots of the same data. The graphed data samples consisted of 50 random numbers about a Gaussian distribution with a mean of 50. Some of the data samples were given transformations to produce various correlations between the resulting 24 sample plots. The standard deviation within data samples ranged from about 6.2 to 11.6. Sound generation was constructed using Microsoft Excel to compute parameters for use in the CSound program. Each auditory graphs had a five-second duration, with individual data points represented by 0.1 second guitar pluck note. The *x* axis was represented by time and the y axis by a pitch scale ranging one octave below to two octaves above middle C. The data was mapped to a chromatic scale.

Subjects rated the magnitude and sign of the correlation between the variables in the graphs. The judgment data were recorded as a distance from the zero point on the scale. The visual graphs were presented for 10 seconds, while each auditory graph was played twice for a total of 10 seconds of listening time. Pearson's correlation for the comparison between the actual correlation and the judged correlation was $r = 0.92$ for the visual group and $r = 0.91$ for the auditory group. A t-test showed no significant difference between the auditory and visual groups.

The second experiment was a direct evaluation between visual and auditory perceptual sensitivity to data points lying outside the main data groupings. This was accomplished by examining changes in the perceived magnitudes and direction of the correlation for scatter plots that were altered with the addition of data points. In this experiment, 32 advanced undergraduate psychology student volunteers participated, 20 in a visual graph group, and 12 in an auditory graph group. Eight data sets from the first experiment were modified by moving one data point: in half of the sets the data point was moved to an outlying position in the center of the plot, and in the rest the data point was moved to an extreme end of the plot. The eight original plots, the eight modified plots, and eight additional plots were used so the number of test stimuli equaled that used in the first experiment.

Of the 24 plots, two of the modified plots showed significant differences in the judgment of correlation magnitudes. The two were plots where the outliers were for moderately correlated data samples rather than for weakly or strongly correlated data sets. Both auditory and visual conditions gave similar results. Thus, this study seems to indicate that judgments between correlation effects for both visual and auditory scatter plots are very similar. Both are effective in conveying sign and magnitude of correlation, and they are similarly influenced by error variances and by single outliers.

5.10. Analysis and Discussion

The studies reviewed in this chapter were found by the current author to be of great use when developing the auditory graph tests conducted for this work. The following discussion is a critique of how the reviewed studies helped define issues related to this work as well as some of their strengths and weaknesses.

The use of pitch to represent data was shown by Pollack and Ficks to have a lower error rate in comparison to other auditory dimensions such as sound duration, repetition rate, or loudness. The ability for pitch discrimination has been used by several researchers to create auditory graphs where the *y* axis data value is represented by pitch and the *x* axis is represented with time. However, Pollack and Ficks also noted that greater information can be transmitted to the listener by increasing the number of binary coding dimensions rather than subdivision of the codings. Hence, if increased information in an auditory graph is desired, additional binary type sounds may be useful considerations.

Mansur *et al.* demonstrated the viability of auditory graphs by comparing auditory graphs to tactile graphs but noted some difficulty subjects had distinguishing between straight lines vs. exponential curves. Studies by Flowers *et al.* extended the study of auditory graphs in a series of comparisons to visual graphs. Their work included several graph types to histograms, scatterplots, and Tukey box and whiskers drawings.

The basic auditory graph served as a starting point for the auditory graphs used in the current research. The previous studies provided auditory graphing methods that had been found to be reasonably effective replacements for visual graphs. The studies concerning more complex methods for mapping data to sound were useful for gaining ideas of what had been investigated and for sound generation techniques and controls.

This chapter, along with chapters 3 and 4, have been an attempt to demonstrate that there is a wide range of literature related to the current research. Perhaps the most relevant studies are those concerning auditory graphing techniques, especially those by Mansur *et al.* and those by Flowers *et al.* The subject material for the questions on which to base the graphs came predominantly from those studies presented in the chapter on physics graphs and concepts. Those studies relating the use of computers in the graphing process demonstrated that the student subjects are familiar with the computer as a graphing tool, and that it need not be presented as an unfamiliar object.

While the studies concerning graph perception may seem the least relevant, they serve as an underlying basis for the foundation of this work. It is important to keep in mind the common structures that people are familiar with when creating new representations for data display.

6. THEORY

6.1. Hypothesis Development

Wavering noted that "graphing is a tool used in science to display data and aid in the analysis of relationships between variables. Also, graphs are part of our daily existence with their use in all media." [Wav89, p. 373] In spite of the prevalence of graphs, several studies have uncovered areas where students have difficulty interpreting graphical information that is used not only in physics, but also in mathematics and economics [Bei94, Mcd87, Lei90, Coh94]. Because of these difficulties, researchers have devoted considerable energy to the teaching and study of graphs.

While progress is being made in teaching graphical information, some attention to how the data is displayed is warranted. Tufte discusses the effectiveness of graphical display methods [Tuf90] and literature concerning optimal display methods was reviewed in chapter 3. Unfortunately, almost all of these studies concern visual display methods. There are several problems with focusing on only the visual data display aspect, most importantly: What happens when one cannot see the graph in question? Thus, it is important to explore other avenues of displaying the information contained in graphs. Auditory graphs are one method for presenting information in a non-visual format.

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. Previous research has focused on how people perceive graphs and how they use graphs to learn about physics. Several studies have also investigated how well people can make judgments about graphs. However, none of the previous studies have demonstrated whether auditory graphs can be practically implemented, and what sort of results could be drawn when students use auditory graphs to answer questions.

The ability to present data with an effective auditory format is one of the prime goals of this research. 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 [Flo92, Flo93, Flo97], 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.* [Tur96] 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 [Flo95, Flo97].

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 [Flo97, Tur96].

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.

6.2. Further Justifications for the Research

At the most fundamental level, this research provides a method for portraying graphical information to people who are unable to interpret a visual graph. While haptic (pertaining to the sense of touch) methods for creating graphical information have been used in the past, there are several difficulties including interactiveness, resolution, production, portability, and storage issues. Haptic graphs require a significant amount of time for identifying contained elements and often a tutor is necessary for explanation of the information. The auditory format can remove many of these limitations.

The basic auditory display used throughout this study, centers on mapping the y axis data value to pitch and the x axis to time. 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 Main 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 is a strong similarity between this mapping method and music notation.

The association of pitch to the magnitude of data values is common practice and has been widely used in research and in other data sonification programs. There are different sound mapping methods, but pitch is the most common, has been applied in many cases, and appears to be intuitive for most people. Investigation of other auditory mapping techniques was outside the scope of this work.

Previous studies have focused on general similarities, or the ability of subjects to identify trends or differences in comparative data sets. The next logical stage after the identification of parts of graphs is the interpretation of the graph as a whole and the analysis of the graph's meaning. However, previous studies have not investigated the ability of subjects to interpret auditory graphs. Interpretation of a graph includes identifying the trends of a graph, making conclusions based on displayed data, or using a graph to infer properties about the system used to produce that graph.

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.

6.3. Implementation of the Research

6.3.1. Genesis of the Testing Process.

Personal computers have been used in many auditory display studies because of their ability to generate a wide variety of sounds. The TRIANGLE program developed by Oregon State University's Science Access Project takes advantage of this sound capability to generate an auditory graph as a complement to, and substitution for, the visual graph display.

TRIANGLE's primary purpose is to provide a workspace for students and scientists to read, write, and manipulate mathematics. TRIANGLE contains a calculator that permits evaluation of most standard math expressions. The calculator also evaluates y versus x functions and displays the results in a plot window. An auditory graph of the function or of data provides a blind or visually impaired user with a quick semi-quantitative overview of the graph. The auditory graph contains a number of display options. In addition, there is a moving icon on the screen to provide information about the graph to users who are both blind and deaf [Gar96].

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 an unbiased 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.

Because the TRIANGLE program was the genesis of the research, the initial investigation centered on using this program as a testing medium. TRIANGLE displayed both visual and auditory graph formats, as well as a text region that could be used to display questions about the graphs. Hence, in the initial stages, it was a good candidate for implementation of the testing process. Later, a testing process based on the World Wide Web proved to be a more flexible alternative with many advantages and is discussed in chapter 8.

6.3.2. General Test Design

The first stages of the testing process required several assumptions. The first was that the auditory display method implemented with the TRIANGLE program would be sufficient to the task. This was not an unreasonable assumption given that previous research employed similar auditory mapping methods. Also, TRIANGLE had been tested by several people for usability and stability.

The TRIANGLE program was used as the initial basis for the study. It was necessary to formulate an unbiased testing process that would demonstrate the ability of subjects to answer and evaluate questions based on graph types. A standard testing method is the causal-comparative design. This method consists of a pre-test, treatment, and post-test. Subjects are given a pre-test to measure their initial state, some form of teaching or learning treatment, and a post-test to measure their final state. Comparing the pre- and post-test scores provides a judgment on the effectiveness of treatment methods.

Although the causal-comparative method is convenient, it only demonstrates the ability of subjects to learn and to use auditory graphs. It provides information about the type of training that the subjects receive. Auditory graphs have been shown in previous research [Man86, Flo92] to be useful for identification of basic graph types, such as linear or curved, and for dissimilarity judgments. These types of investigations are not the focus of this study.

A comparison between two or more groups of subjects can be combined with the pre-test, treatment, post-test method. The pre-test verifies that the groups have equivalent abilities, or can be used to give a basis for correction if the groups are found to be non-equivalent. In the current study, the treatment was the visual or auditory display of a graph. The post-test was a series of questions and their associated graphs. The results of the post-test were compared to judge the effectiveness of the graphing treatment methods.

An assumption of this study was that knowledge of the subject matter used in the questions would be an important effect. A subject's understanding of the material could affect his or her overall performance. Subjects were randomly assigned to different groups so that student performance in each group would ideally be similar. By comparing the performance of two groups on identical questions, any difference was thus focused on the ability of subjects to utilize the graphs, and not necessarily on the knowledge of the material in the questions. The questions acted as a basis for different reasoning structures that are important to physics and math such as identification of functions, discontinuity, implications of the slope, maxima, or other prominent features.

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 math and physics. The population for the study was drawn from subjects who had taken, or were in the process of taking college level physics courses.

One difficulty of this study is gauging subject involvement when answering the questions. Since the subjects recruited in this study were all volunteers, and no incentive for their performance level could be applied, there is no guarantee that the subjects performed at their best level when answering the questions. However, since subjects were randomly chosen 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 can be adjusted for random guessing which should reduce the effect of any student apathy towards the test.

6.3.3. General Data Collection Procedure

The specific methods used to collect data varied between the Triangle Pilot, the Web Pilot, the Main Auditory Graph, and the Auditory Preference Pilot tests. The methods are fully developed in the chapters relating to each test. The first two pilot test studies investigated the test environment and development of test questions used in the Main Auditory Graph test. The general process of data collection in the first three studies consisted of giving 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 the three 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 instructors to volunteer 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 multiple-choice questions 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 Main 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.

6.3.4. Testing Considerations

There is the possibility that a difference in performance levels between visual and auditory graph groups does not necessarily demonstrate an inability of subjects to understand and interpret the presented material. Instead, the difference could be attributable to training and familiarity effects. It was assumed that since the subjects were drawn from standard physics courses, they had been exposed to many visual graphs in the course of their studies. It was also assumed that the subjects had been exposed to virtually no auditory graphs as this is a new representational method. Thus, it was assumed that subjects had much more experience with visual graphs than with the auditory graph representation. Some explanation and training for the auditory graph representation was necessary, but the amount of required training remains undetermined.

The issue of performance effects due to the familiarity of graphs and subject material can be addressed by comparing the subjects' results with those from more experienced graph users. By looking at how well a group of experts (physics graduate students) perform on the questions when given auditory graphs, it can be demonstrated whether the questions are answerable, and what would be the best expected outcome for the groups. Any issues of unfamiliarity with the testing material can be eliminated, thus focusing only on the difference in the graph styles.

There are several reasons why expert subjects were not solely used for these experiments. First was the issue of the audience that the auditory graph representation is trying to target. This graphing method is envisioned to be used as a common tool to help students understand basic data graphs. As such, it is important to discover whether beginning students can understand these graphs with little training and experience.

Another issue was the number of subjects required for meaningful statistical results. Given a normal population distribution, for the 95% probability level, the approximate size of a group required for a 95% chance that the average measurement, $\overline{X}$, is within the limits of $\mu \pm L$ is given by:

$$n = \left(1.96\frac{\sigma}{L}\right)^2 \tag{6.1}$$

where μ is the target population mean and n is the sample size [Sne89, p. 52]. Now, assuming a standard deviation of 20%, since there are 5 possible answers to each question, and an error limit of 5%,

$$n = \left(1.96\frac{0.2}{0.05}\right)^2 \approx 62. \tag{6.2}$$

Thus, each graph test group should have a minimum of 62 subjects.

The prime target audience of the auditory graph representation is blind users. While it would be desirable to use 62 blind first-year physics students to gauge their ability to answer the questions using the auditory graphing technique and compare their results to sighted users of equivalent background, this is not possible. There are extremely few blind students meeting the conditions of ever having had physics at the college level, even on a national scale. Blind people who have completed physics courses were solicited for their participation via requests on electronic mailing lists. However, only a very small number of people (five) participated. This will be discussed in more depth in chapter 9.

In any test, there is a question of whether the test is reliable and valid. Validity of these tests was determined by review of the questions with experts, and by comparing test results between first-year and graduate students. By designing the tests so that questions can be divided for split-half analysis, a statement about the test's reliability can be made. If there is a high degree of correlation between the scores in the two halves, then there is a greater probability that similar questions will have similar results. This helps to indicate how well subjects can reliably use the auditory graphs to answer questions.

One difficulty with the testing process used in this study that should be noted was the high reliance on technology. While this posed certain challenges, the technological problems affected all subjects equally in the Web-based tests. In the preliminary Triangle Pilot, an instrument method that was not technologically dependent was utilized for initial comparative purposes.

It is possible that a better scheme for testing and producing auditory graphs can be developed. The Main Auditory Graph test was an evolution of the processes used in the Triangle Pilot and Web Pilot tests. As was previously stated, the auditory methods used in this study were chosen to a large extent by results from previous research, similarity to musical representations, and prior device development.

Research into better graphing techniques is necessary and is the issue of further studies. Some indications of possible graph questions as well as alternate auditory display methods are demonstrated in the Auditory Preference Pilot test discussed in chapter 10.

7. TRIANGLE PILOT

7.1. Overview

The first experiment conducted was a pilot test to investigate the advantages, difficulties, and question layout of a study involving auditory graphs. This experiment, named the Triangle Pilot, 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 but also showed that there were no insurmountable difficulties with the auditory technique. The Triangle Pilot provided the basis material that was used in later studies.

This experiment consisted of three instruments: an initial Survey questionnaire, a Pre-test, and a Main test. The purpose of the questionnaire was to provide basic demographic and other relevant information to aid in analysis of the responses. The 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. Appendix A contains a copy of the Survey (A.3), Pre-test (A.4), and Main test (A.5).

Subject matter for the questions centered on previously published research involving graphs and physics. Most notably, questions from the Force Concept Inventory (FCI) [Hes92a], Mechanics Baseline [Hes92b] test, and the Beichner [Bei94] 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.

There were four treatment methods for this study: graphs visually presented on paper, graphs visually presented on the computer, auditory graphs presented on the computer, and both auditory and visual graphs presented on the computer. The paper presentation method was to check for any novelty effects that the TRIANGLE program might introduce. 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.

All materials used in this study were submitted to the OSU Institutional Review Board (IRB) for review and approval. After receiving endorsement by the IRB, subjects were solicited for participation in the study. Two graduate students participated using auditory graphs for purposes of testing validity, and for estimation of time allotment for scheduling purposes.

7.2. Sample

Ideally, a random sampling from a wide variety of first-year physics students would be desirable. However, this was not possible due the scope of the pilot study. Since the Triangle Pilot was intended to determine the feasibility of a study on auditory graphs, it was decided to limit participation to local students. The use of local students was a matter of convenience and limited the generalizablity of the study. The Main Auditory Graph test included subjects from several educational institutions to allow for greater generalization of the studies' results. As the testing process was designed for first-year physics students, instructors of these courses were solicited for the possibility of letting their students participate in this study.

The sample was drawn from an introductory, algebra-based physics course at OSU during the 1997 summer-session. It was arranged with the professor of this course that the investigator would ask for student volunteers from the course's laboratory component. Subjects participated in the study during the same time as their normally assigned laboratory section. For their participation, students received full credit for the missed class. Names of volunteers were taken from each of three laboratory sections that met on the same day, with 43 of approximately 60 students volunteering. While some of the volunteers may have chosen to participate due to a higher motivational level, the offered incentive attracted many of the volunteers.

Due to time constraints and resources, the study limited the focus to twelve students. The number of subjects was chosen due to the number of students who could be tested by one researcher during the three two-hour laboratory sections on a single day. The time for completion of the questions for the test was estimated to be half an hour as this was approximately the time taken by graduate student volunteers on a previous day. A list was formed of the volunteers from each laboratory section and a computer randomly selected four subjects from each laboratory section for a total of 12 subjects. The chosen student volunteers were taken from their next lab session and came to a designated room at assigned 1/2-hour intervals.

7.3. Data Collection

Data were collected through a guided interview process. The interviewer served as a guide to answer general questions, such as those arising from ambiguous wording or instructions, and set up the questions and graphs on the computer. This last step was necessary as the TRIANGLE program was not designed as a testing environment. Subject volunteers selected from a random computer-generated list met with the interviewer at an appointed time and place. Subjects were randomly assigned to one of the four graph category groups (three subjects per group): visual graphs printed on paper (Print group), visual graphs displayed on the computer (Visual group), auditory graphs produced by the computer (Sound group), or both visual and auditory computer graphs (Both group). Subjects were shown each of the questions on the computer, except for the print group which had questions on paper, and given as much time as they wanted to answer the question before proceeding to the next question.

The testing area consisted of a room with a large table upon which a computer was placed, and several chairs at the table. A video camera recorded the sessions, and all subjects had been questioned and gave their consent to being videotaped; the subjects where shown where the camera was located. Upon entering the room each subject recorded his or her name on a log page and was given a Document of Informed Consent (Appendix A.2) to read and agree to. They were next presented with the Survey and the Pre-Test. After completing the initial questions, the subjects were given an answer sheet (Appendix A.6) to record their responses. For each subject, the Survey and answer sheet were marked with a unique code for identification purposes and for recording the type of graphs on the test.

Each subject was given one of the four graph formats for the Main test. The order of the type of test given was changed between groups of students. The listing is given in Table 7.1 where the representation is P for the test given on paper (Print), V for the test displayed on the computer (Visual), S for the test presented on the computer with auditory graphs (Sound), and B for the test presented on the computer with both auditory and visual graphs (Both).

Table 7.1 Test Type per Interview Time.

Student :	1	2	3	4
Group 1: 10-12 am	S	V	B	P
Group 2: 1-3 pm	B	V	P	S
Group 3: 6-8 pm	P	S	V	B

In cases where the computer was used (S, V, and B groups) the investigator changed the displayed question and graph after the subject had finished with the previous question. In the studies with auditory graphs, subjects had control of the graph playback via the computer's keyboard. Subjects were allowed as much time as they wanted to study and listen to the graphs and to answer the questions. Subjects were also allowed to return to previous questions if they wanted to change their answers. The interview process was videotaped for later study, most notably to check for leading by the interviewer.

7.4. Instrument Development

The test questions and graphs were displayed with the TRIANGLE program. This is a DOS-based program developed by the Science Access Project at OSU. This program has a text region where the questions can be displayed, as well as a display for visual and auditory graphs that can be generated from a table of data points and then plotted on the screen. While viewing the graph, a user can also listen to a sound representation (sonification) of that graph. In the case of the TRIANGLE program, sonification of the data was represented with a linear relationship of pitch to the *y* axis data values. The *x* axis values were converted to time, so that the graph was played from left to right. 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. Subjects in the test groups that used sound graphs were given a brief description of the auditory graphs but no specific training was performed. Screen images of the TRIANGLE display can be found in Appendix A.7.

Questions for the study's Main test proved to be a challenge to create. From a review of previous research, it was decided that a multiple choice question format would provide useful information for determining the effectiveness of auditory graphs. The primary difficulty in question development was creating multiple-choice questions that would reference a single graph. In addition, the graphs needed to be comprehensible displays both as visual pictures as well as sonified data sets. The TRIANGLE program imposed a further 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.

To provide testing situations that were as nearly identical to each other as possible, there was no difference in the information contained in each of the graph types or in the question wording. For example, title and axis representations were mentioned explicitly in the question text, rather than on the graph's axes, as the auditory graphs had no labeling method. Each graph was displayed separately from the question text, although this was partially an aspect of the program used to display the graphs.

Only simple graphical information was portrayed in the questions because previous studies had not determined the effectiveness of interpretation of auditory graphs. The restrictions placed on the development of questions and investigation of introductory texts and previous studies provided material for 14 questions. These questions underwent review by graduate physics students and professors in physics and science education at OSU.

The wording of the individual questions was designed to provide correct and clear distinction between choices with an emphasis on drawing conclusions from the information displayed in the graphs and not primarily on their background knowledge of physics. Two short-answer questions were included to probe their understanding of more complex physical issues (11b and 12b) but these were not the primary focus of the test. Answers to the multiple-choice questions were evenly distributed among five choices (A, B, C, D, and E). The answer sheet was developed to provide a consistent method for the subjects to record their answers. Extra space on the answer sheet allowed subjects to write any additional comments or questions about the wording of the test questions. A copy of the answer sheet is located in Appendix A.6.

To check reliability, the test questions were constructed to be applicable for split-half analysis. Table 7.2 displays the correspondence between the graph type and the question number. Test splitting was for similar graph type rather than for questions concerning similar physical phenomenon.

Table 7.2: Distribution of Graph Types.

Graph Type	Question Numbers
Linear: Constant	1, 4
Linear: Increasing	2, 7
Linear: Decreasing	5, 8
Segmented: Linear	3, 6
Segmented: $1/x^2$	11, 12
Nonlinear: x^2	9, 13
Nonlinear: Root, $1/x$	10, 14

The graphs used in the Pre-test and the Main test were created in a multi-step process. First the physics principle investigated was modeled by an equation or segmented graph. A small program was developed to aid in creating a two column table of numbers that represented the desired graphs. Each graph had 100 data points, as the TRIANGLE program could create an auditory graph lasting approximately three seconds with that many data points. Each table was imported into Microsoft Excel for collation into a larger table so that each question was represented by one column of data. The resulting table was converted into a format for use by the TRIANGLE program. Each column of numbers was plotted at the time that the corresponding question was asked.

The necessity of using an interview process arose from the difficulty in learning and using the TRIANGLE program to display questions and graphs. While the auditory display was straightforward to use once the data had been loaded, the process of loading and manipulating the data could have interfered with the interpretation of the graph. Thus, the interviewer was responsible for displaying the data so that the subjects needed only to be concerned with interpretations derived from the display methods.

7.5. Data Results

Table 7.3 is a summary of the results contained in Appendix A.8. The table is divided by results from the different test groups.

Table 7.3. Percent of Subjects Answering Given Questions Correctly.

Question	% Correct Print	Visual	Both	Sound	Grad
Pre-test:					
P1	100%	100%	100%	100%	100%
P2	100%	100%	100%	100%	100%
P3	100%	100%	100%	67%	100%
P4	67%	100%	100%	67%	100%
P5	0%	33%	0%	100%	100%
Average	73%	87%	80%	87%	
Main Test					
Q 1	67%	67%	0%	0%	100%
Q 2	33%	33%	33%	0%	100%
Q 3	0%	33%	0%	0%	100%
Q 4	0%	67%	100%	67%	100%
Q 5	0%	33%	0%	33%	100%
Q 6	33%	67%	67%	67%	100%
Q 7	67%	100%	100%	33%	100%
Q 8	67%	100%	100%	100%	100%
Q 9	100%	100%	100%	33%	50%
Q 10	67%	67%	33%	0%	100%
Q 11	100%	67%	67%	67%	100%
Q 12	67%	100%	67%	67%	100%
Q 13	100%	33%	67%	0%	100%
Q 14	100%	67%	67%	0%	50%
Average:	57%	67%	57%	33%	
Standard Dev.	38%	27%	38%	34%	

While the summary table provides an accurate listing of the data, it is helpful to view the same data as a bar chart to recognize patterns in the data and to easily see where any difficulties may lie. The following chart displays the percent correct scores of each test group vs. the individual test questions.

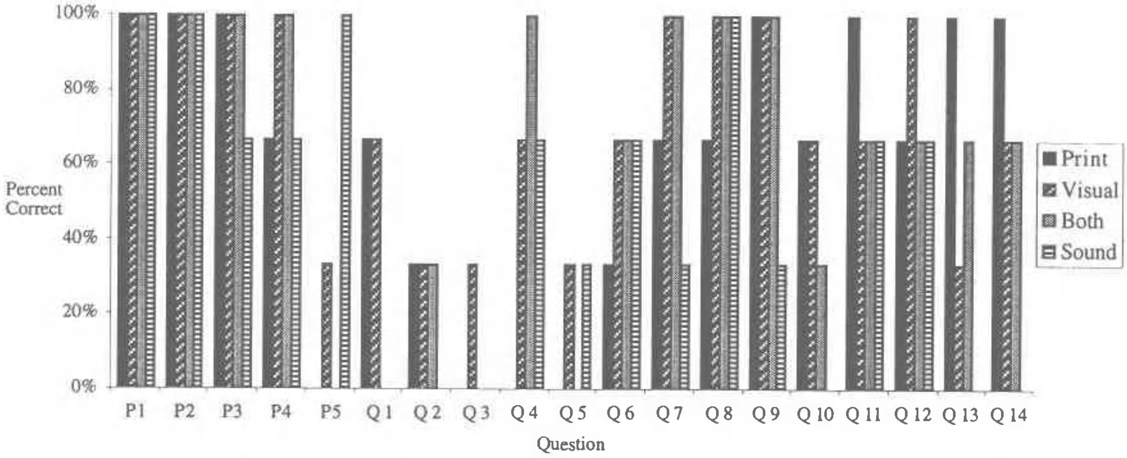


Figure 7.1 Comparison of Triangle Pilot Test Results for Different Presentation Methods.

Three subjects in each of the four groups is far below the required number to produce valid statistical analysis. However, keeping in mind that errors are greatly exaggerated, the pilot test leads to a number of insights. The first point to be noted is the striking difference between the Sound and Visual groups for Main test questions 1, 2, 3, 10, 13, and 14. These questions are reviewed in section 7.6.

The poor performance of the Sound group on a large number of questions gives an early indication that there may have been an oversight in the method of auditory graph production that was used. A dramatically lower score from the Sound group indicates that either the auditory graphs were not properly explained and understood, or that there was an fundamental display problem that prevented subjects from fully understanding the graphs. The two graduate students who participated using auditory graphs had perfect scores except for questions 9 and 14. These questions may not be valid, or the auditory representation is not adequate even for experts.

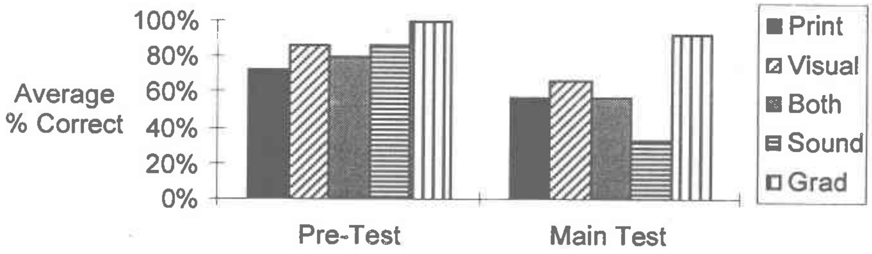


Figure 7.2. Average Scores of Groups on the Triangle Pilot Test.

The mean score on the Main test for all the undergraduate respondents was 53% answering correctly. While this may seem like a satisfactory value as it is in the center of the distribution, the average standard deviation for the student groups was a very large 34%. The large standard deviation is in part attributable to the small sample size, but may also be attributable to poor wording of the questions or lack of knowledge by the respondents. Analysis of individual questions is useful for determining any specific problems and is covered in section 7.6.

The average score for the Visual group was twice that of the Sound group. Although this difference may be attributable to random variation, it is more likely that it is due to subject difficulty with the auditory graphs. A revised auditory graphing method was used for later tests.

Table 7.4 Split-Half Analysis of Test Questions.

Question	% Correct	Split Question	% Correct
1	33	4	58
2	25	7	75
3	8	6	58
5	17	8	92
9	75	13	50
10	42	14	58
11	75	12	75

To comment further on the validity of the test questions in general, analysis of the question types is necessary. The percent of correct answers in question 11 is within one standard deviation of the score for its split question (12). The same is true for question pairs 1-4, 10-14, and 9-13. Thus, 4/7 of the pairs have differences within one standard deviation. When looking at two standard deviations, question pairs 2-7 and 3-6 are also included bringing the total to 6/7. These are reasonably close to the ideal limits of 0.66 for one standard deviation and 0.95 for two.

However, split-half analysis to check the reliability of problems of similar graph types generally shows a disappointing relationship. The correlation coefficient between the two groupings yields $r = -0.28$ which is very poor in comparison to the ideal of 1. The low r value indicates that either the question wording is ambiguous and needs to be rewritten, or that similar graph types do not lead to similar response rates. If the latter is the case, then the question's material plays a greater role than the graph in the subject's understanding of the graph.

7.6. Detailed Analysis of Selected Questions

The following is a commentary on the answers arising from the Survey portion of the pilot test. The Survey consisted of eight items that were a combination of multiple-choice and short answer questions. While the results of this Survey were not used for analysis purposes due to the small number of subjects, it was instructive to find where ambiguities lay. Refined versions of the Survey were used in later experiments such as the Web Pilot and the Main Auditory Graph tests.

Survey questions 1, 2, and 3 concerned general demographics of gender, age and high school physics experience. There was some confusion with question 4: *Number of years of college-level physics?* by the first-year students who took the test. They often asked if they should circle the 0 (because they hadn't completed a year yet) or 1 (as they were currently taking a first-year course.) This difficulty was corrected in later tests by restating the question as: *How many courses of college-level physics have you completed?*

Survey questions 5 and 6 asked subjects to state courses that they felt had been helpful in understanding graphical information. Responses often only stated course numbers which were difficult to decipher. These questions were rewritten for later studies.

Survey question 7: *Have you learned graphing techniques other than from academic settings?* was confusing to a large number of subjects. During the Triangle Pilot test, subjects were given examples as to what type of answer was being sought which tended to ease the confusion. This question was rewritten in later studies. There did not seem to be any difficulties with questions 8 or 9 which related to musical training or physical difficulties.

The Pre-test questions were interesting. It was expected that questions 1 to 4 would be answered correctly by everyone. They were designed to test the ability of the subjects to simply read a graph and locate points. Three of 14 subjects missed one of the questions and one subject missed two questions. This leads one to conclude that there may be a 9% error rate on the test due to students misreading, or simply not being careful with, the questions.

Question 5 on the Pre-test was designed to be similar to one of the moderately difficult graphs. It was a nonlinear graph referring to motion of an object. The number of correct responses was dramatically lower than the previous questions at 33% correct. This question has content validity as the graduate students had a 100% correct response. It should be noted that all but one of the subjects who had correctly responded to this question were randomly assigned to the auditory graph test group.

The text and answer choices for the questions in the Main test tended to be more complex. A full listing can be found in Appendix A.5, so only a brief description of the graph or changes to be made to the questions is described below. A listing of the percentage of correct responses per group is provided for each question. The categories are: the total results for the undergraduate test subjects (Total), the printed test group (Print), the group that only saw the test and graph displayed on the computer screen (Vision), the group that both saw and heard the graphs (Both), the group that only heard the sound graphs (Sound), and the graduate student subjects (Grad). The graduate students are used for validity comparison.

An important point to be stressed with these results is that the undergraduate subjects were divided into four groups, so that each group only had three test subjects, and that only two graduate student subjects participated in the test. Statistical fluctuations could account for many irregularities in the results as one incorrect response would manifest in a change of 33% for any given group's correct response rate for a question. A larger number of test subjects reduces the ambiguity and is demonstrated in the Web Pilot test described in chapter 8.

Main test question 1 was one of the more surprising results. The graph was designed to be the easiest to recognize, a straight flat line, and have common axes (distance, time) yet most students were not able to answer this question correctly. The averages for the Print and Visual group were equal whereas the subjects in the Sound and Both groups all missed this question. The difference, especially for the Both group could be attributable to a novelty effect of hearing the sound display and unfamiliarity with what the display was representing.

The auditory graph in this case was a constant tone. A training period for the sound graphs seems to be necessary for less experienced students, especially when considering the results from the next two questions. It was noted that answer *A) The object is moving with a constant non-zero acceleration* and *C) The object is moving with a uniformly increasing velocity* are the same. The second answer was changed to *C) The object is moving with a uniformly decreasing velocity* for use in the Web Pilot test.

Question 2 was again intended to be fairly straightforward. The Grad response shows that experienced graph readers using the auditory display can understand this question. The equality between Print, Vision, and Both groups shows that displaying the graphs on the computer may not be a significant effect. The low score for Sound, may be from random variations, or as a result of unfamiliarity with the auditory graph representation. The same change to answer C was later made on this question for the same reason as with question 1.

The results of question 3 were particularly interesting as this question was modeled on a similar question used in the study by Beichner [Bei94]. In that study, there was a 33% correct response rate whereas this pilot test had an 8% rate. Again, the Sound group completely missed this question. The results may not be as significant as the other groups also did very poorly. It is possible that the entire question should be rewritten. However, the Grad subjects found the question legitimate. Answer *B) The object doesn't move at first. Then it rolls down a hill and finally stops.* may have caused some confusion as it could possibly be construed as correct. A change *to B) The object doesn't move at first, then it moves away from the reference point, and finally stops.* may correct any misunderstanding.

Question 4 gives some indication that auditory graphs may be a valid form of data representation as the Sound group score was as high as the Vision group and the Both was higher still. However, the dramatic difference between Print and Vision scores may indicate that random fluctuations are greater than the one standard deviation previously mentioned. This question was designed to be a complement to question 1, and the Vision group did have the same score. Groups using sound had improved scores, perhaps indicating that a training period had taken place. Answer C should probably be modified as in Question 1.

The generally poor results on this question 5 can be attributed to the wording of the answers, which while correct, were meant to distinguish amoung subjects who had a good understanding of the concept of acceleration. Subjects chose only one of two answers: A) *The object is moving with a constant acceleration.* and B) *The object is moving with a decreasing acceleration.* The graph showed a linearly decreasing velocity, hence a constant, albeit negative, acceleration. Experienced subjects did not have difficulty with this question.

Subjects performed generally well on question 6 which had one of the most complex graphs. This question paralleled one of the Pre-test questions. As can be seen, the Sound group did at least as well as the other groups.

It is not understood why the Sound group did considerably worse on question 7 than the other groups, except that this question is the same graph type as question 2 and the results were also poor on that question. There could be a difficulty in recognizing the pitch, and hence the y value, as representing a linear increase.

Questions 7 and 8 were related by the physics principles that they questioned. Hence the similarity in the majority of the scores is not unexpected. The Sound group shows a dramatic change. This difference could be attributed to statistical fluctuations. There is also a possibility that auditory graphs that start with a high pitch and end with a low pitch may be more easily distinguished than in those that start low and end high.

Question 9 gives a clear indication that curved graphs are not well perceived in this form of data sonification. Not only did the Sound group do poorly, but the graduate students also noted that they could not tell the correct answer from the sound graph. The graduate students mentioned that they knew what the one correct answer was but answered according to what they heard.

The results of question 10 are interesting in that subjects only chose one of three answers: the correct response of 1/wavelength, the incorrect response of decreasing linearly with wavelength, or the answer of: not related to wavelength. The second answer may have been the result of confusion with the wording; this answer should be rephrased to *B) The frequency has a constant, linear decrease as the wavelength increases*. While the question appears to be generally valid, the low scores of both groups utilizing sound indicates that a better sonification technique is necessary for less experienced students.

Subjects did well on the similar questions of 11 and 12. These questions involved graphs that were more complex. Answers tended to be between those that were most similarly related to the graphs, perhaps indicating that better distracters need to be constructed. The Sound group performed as well as the other groups on these questions, indicating that more complex auditory graphs can be used.

There were second parts to these questions, *11b* and *12b. What does the peak on this graph represent?* The answers tended to be statements of the graph, rather than its physical interpretation, indicating poor question wording. This type of questioning was dropped in later experiments.

The large difference in scores between the Print and Vision groups in question 13 may be due to random fluctuations. Again, the Sound group shows that this sonification method was not adequate for curved graphs, and the Grad results also back this statement.

For question 14 the problem seems to be not in the question, but in the sonification of the graph. These results are perhaps the clearest indication that the method of sonification used in the Triangle Pilot was not useful for simple curved graphs.

7.7. 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 test. 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 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.

Subjects were able to answer all the test questions in the allotted time. The assumed time to complete all test parts was about 30 minutes, or about one question per minute. This was a reasonable guess as many of the problems were conceptual, multiple-choice questions that did not involve calculations. Subjects in the Print group tended to finish the Main test in less time than did the others due to not having to wait for the investigator set-up the computer display for the each problem.

While a cursory review of the data shows that the Sound group performed substantially below the level of the other groups, this problem may be able to be overcome. First, looking at the questions where the Sound group performed as well as the others and comparing the question types to those where they did not perform well provides important clues as to better sonification methods. It appears that subjects were able to distinguish absolute values by sound (i.e. pitch being higher or lower). Also, the subjects were able to identify the first derivative of the function (is it increasing or decreasing) but not the second derivative (the rate at which a function is increasing or decreasing). These conclusions are shown by the poor results on any of the graphs portraying curved functions. The more experienced subjects (graduate students) were able to interpret the shape of the graph from the more limited information. This came about 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, Main Auditory Graph, and Auditory Preference Pilot tests.

While the Print group was useful for comparison, and a slight difference in the scores was noticed, the difference did not seem to be significant. This result indicates that this fourth grouping is not necessary. Remote computer administration of the test, and greater subject participation, was accomplished more easily without this unnecessary group. With fewer testing sections, the group sizes were larger for a given number of subjects resulting in greater confidence limits.

Lastly, it was evident 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 this test. One factor for these results was the lack of training that subjects in the Sound group encountered before the test. While some training is evidently necessary, the goal of this auditory display is to have a method that is reasonably intuitive. Subject performance for the Sound group generally seemed to increase up to question 7, thus providing an initial estimate for the number of graphs for training. 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. The modifications to the questions, initial information, and display formats formed the basis for the second pilot test called the Web Pilot.

8. WEB PILOT

8.1. Overview

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 [Ceb97].

This experiment, named the Web Pilot, used a standard 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 with this new medium. The Triangle Pilot provided the basis material for this experiment; the Web Pilot provided the testing technique used in the Main Auditory Graph and Auditory Preference Pilot studies.

8.2. Sample

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 OSU to provide extra-credit homework points to students taking Web Pilot test. The instructor announced the location of the Web Pilot test's introductory Web page in class and posted a link on the course's 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 $n > 62$ (from equation 6.2) for each group.

8.3. Data Collection

Subject volunteers accessed the Web Pilot test site from remote computers at various locations at Oregon State University. Several PERL scripting programs recorded data from subject responses to questions presented on the Web pages. The data were written to secure files. The PERL scripting programs can be found in Appendix F.

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, a link to a Web page containing further descriptions and examples of auditory graphs, a copy of the informed consent document, and a link to the test. A copy of this page is located in Appendix B.2.

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 subject's names. This program also randomly assigned the subject to one of the graph test groupings, labeled by b, s, or v, and then passed the ID and graph codes to the next Web page.

The use of the ID code number provided anonymity of the final results (so that names would not appear with the test scores), yet allowed the investigator to identify students so that multiple attempts of the test could be eliminated. Also, the code provided a record of which students from a given class completed the test, and allowed for comparison of results between tests. Due to security, anonymity, and coding issues, subject's names were not written to the test's Web pages.

The Survey Web page contained text entry fields as well as radio-button type choice fields. A second PERL program called "surveyrecord" appended the subjects ID code number and any long text answers to a separate file when subjects chose the "Next page" button. The ID code, graph code, and several of the Survey answers were passed as a text string to the Pre-test page. The graph code was also passed as a variable to later pages.

A third PERL program called "prerecord" added the Pre-test page answers to the text string when subjects chose the "Next page" button. This program passed the answer string, along with variables for the graph code and the "start time" of the test, to the next Web page. The time that subjects took to answer the test was measured to provide some insight as to how long students took with the different presentation methods. The prerecord program generated the first question page for the Main test.

A fourth PERL program called "temprecord" generated subsequent Web pages for the Main test. The questions were read from individual files, and contained multiple-choice, radio-button style answer selections. Graph codes, previous answers, and starting time information were written to the generated Web pages. The graph presentation was determined from the graph code value and incorporated into the pages as well. When subjects chose the "Next" button, their answer for the question was added to the answer string and the next question was read from a predetermined file. When the subjects had completed the last question, the temprecord program calculated the total time, added this information to the answer string, and appended the string to a secure file of previous subjects' answers.

8.4. Instrument Development

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 the revisions as noted in the previous experiment. The presentation was through a linear series of Web pages. A copies of the Web Pilot Survey is located in Appendix B.3, the Pre-test is in Appendix B.4. A screen image of a typical question for the Main test is located in Appendix B.5. The questions for the Main test were the same as those in the Triangle Pilot test but with the revisions noted in chapter 7. There were three display methods for this study: visual graphs, auditory graphs, and both auditory and visual graphs. There was no paper presentation method as was performed in the Triangle Pilot due to that test's similarity in scores between the visual and paper presentation methods, and due to logistical difficulties.

The challenge of this experiment was to convert the testing process of the Triangle Pilot to a Web-based format. Methods to display the questions' text and visual graphs were not difficult as the standard Web browser has this ability built into the display. The method of producing auditory graphs that could be played from the browser window was more problematic. The difficulty lies in the ability of computers to produce sound from various audio file formats. One of the most common and useful formats, the Microsoft .wav format, leads to large file sizes (on the order of 100 Kb). Transferring large files on the Web resulted in long delays when displaying auditory graphs, especially if the subject was using lower speed modems to access the test. The use of MIDI reduced the auditory graph file sizes to about 2 Kb, which produced a page that would download and display more rapidly.

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. Thus, complex sounds can be reduced to small data files. For the Web Pilot, y axis data values were represented with a piano voice that varied in pitch.

One disadvantage of the MIDI format is that sounds are not completely consistent from one computer to another, as each computer may have different stored wave-pattern representations corresponding to a given MIDI voice code. Also, there is an inherent limitation on the resolution of sounds since MIDI uses a chromatic scale as a basis for the divisions between notes. Producing sounds between given notes greatly expands the file size.

As was shown in the Triangle Pilot, auditory test subjects had great difficulty distinguishing between linear and curved graphs. One of the Triangle Pilot subjects made the suggestion to add tick marks to represent the y axis values. 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 tone and a low pitch drum that would increase in tempo, and the graph of $y = 1 - x^2$, 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 produced in a multi-step process. The *x-y* data sets used to create the graphs in the Triangle Pilot were converted by the DataSonify program into an SLG formatted text file. SLG is a text file format and is an instruction set that the MIDIGraphy program [Ton99] can import to convert text data to MIDI sound files. DataSonify 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. The MIDI file was converted into the .wav format with SoundMachine [Sou99], and was converted into Apple Computer's QuickTime format with the MoviePlayer program.

The QuickTime sound file format was chosen because it allowed for a Web browser plug-in module that could display embedded play and pause controls in the Web page display. Also, since this module was available for several computer platforms, subjects would have little difficulty locating a computer from which to take the test. To provide alternate access to the sound files, links were included so that subjects could download the MIDI formatted file or the 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.

8.5. Data Results

Table 8.1 is a summary of the full results contained in Appendix B.6. The table is divided by results from the different test groups.

Table 8.1 Percent Correct per Question for Each Group.

Question	% correct per Group Visual	Both	Sound
Pre-test			
P1	77%	71%	78%
P2	97%	97%	96%
P3	99%	96%	95%
P4	83%	78%	81%
P5	71%	58%	73%
Avg.	85%	80%	85%
Main Test			
M1	68%	58%	30%
M2	67%	65%	36%
M3	59%	53%	22%
M4	71%	75%	57%
M5	11%	10%	16%
M6	68%	65%	23%
M7	81%	71%	64%
M8	84%	89%	73%
M9	69%	74%	54%
M10	69%	71%	41%
M11	71%	81%	41%
M12	71%	81%	41%
M13	67%	69%	42%
M14	67%	69%	31%
Average	66%	66%	41%

While the summary table provides an accurate listing of the data, it is helpful to view the same data as a bar chart to recognize patterns in the data and to easily see where any difficulties may lie. Figure 8.1 displays the percent correct scores of each test group vs. the individual test questions.

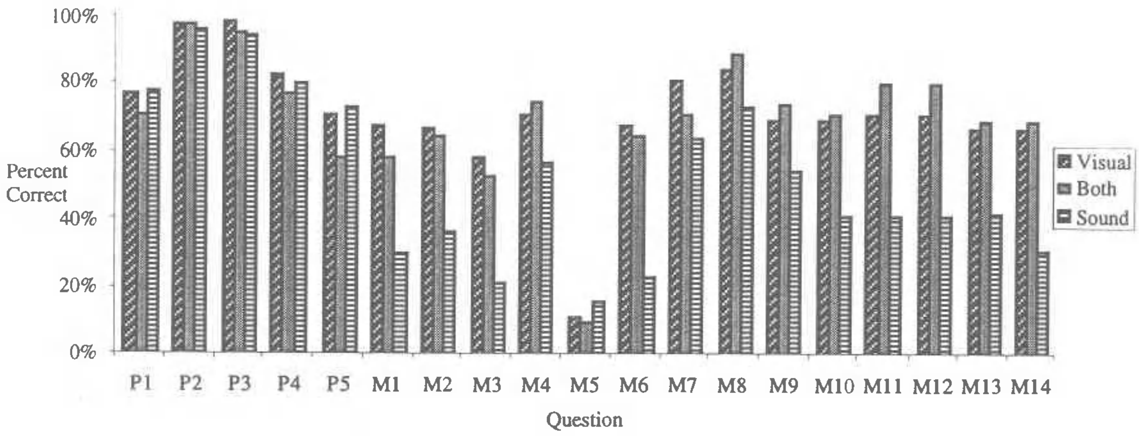


Figure 8.1 Histogram of the Results of Table 8.1: A Comparison of Correct Answers per Group.

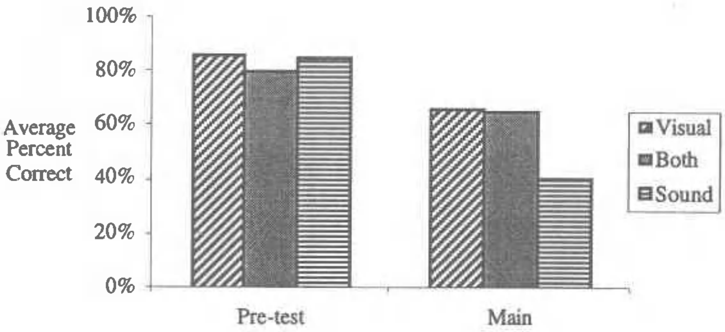


Figure 8.2 Average Scores for the Groups.

It is evident from the displayed results in Figure 8.2 that the Sound group performed at a lower level than did the Visual and Both groups. However, it should be noted that the Sound group for the Web Pilot had a greater percentage correct (40%) than did the Sound group of the Triangle Pilot test (33%).

8.6. Analysis

Analysis of the Pre-test was performed using Microsoft Excel on the data and is displayed in Table 8.2. The results are from a single valued analysis of variance (ANOVA) test at the $\alpha = 0.05$ probability level comparing the Sound, Visual and Both groups.

Table 8.2 Pre-test ANOVA Analysis.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Both	72	288	4.00	0.96
Sound	74	313	4.23	0.92
Vision	75	320	4.27	0.79

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	$F_{critical}$
Between Groups	3.04	2	1.52	1.71	0.18	3.04
Within Groups	193.76	218	0.89			
Total	196.81	220				

Since $F < F_{critical}$, H_0 is valid. No significant difference between groups

The numbers in the *SS* column represent the sum of the squares of deviations for all data. There are two degrees of freedom (*df*). The degrees of freedom value are one less than the number of items being compared. Since there are three groups to compare, there are two degrees of freedom between the groups. The number of degrees of freedom when looking for variations within all of the groups reduces the number of subjects in each group by one and then sums the resulting values. *MS* represents the mean squares and is simply the sum of squares divided by the degrees of freedom. The *F* value is calculated by dividing the between groups *MS* by the within groups *MS* [Huc96 p. 277]. The *P-value* indicates the probability of obtaining sample data that deviate as much or more from the hypothetical difference (in this case 0) than the observed data.

ANOVA of all three groups showed no significant differences. Since $F = 1.71 < F_{critical} = 3.04$, the hypothesis that the three groups are equivalent is accepted. Thus, the three groups can be considered identical.

Table 8.3 Web Pilot Main Test ANOVA Analysis.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Both	72	661	9.18	10.63
Sound	74	418	5.65	8.48
Vision	75	697	9.29	10.32

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	634.65	2	317.32	32.37	0.00	3.04
Within Groups	2137.06	218	9.80			
Total	2771.71	220				

Since $F > F_{critical}$, H_0 is rejected. Groups are not equivalent.

Analysis of the Main test produces a different result however. In this case, as shown in Table 8.3, ANOVA comparison at the $\alpha = 0.05$ level produces $F = 32.37 > F_{critical} = 3.04$, and the equality hypothesis is rejected. Since there is a significant difference between the groups, and since the experiment was designed to make such comparisons, analysis of the data can be made with a series of t -tests.

Table 8.4 Two Tailed *t* -test Between Groups for Web Pilot Main Test.

Group	Mean	Variance	*df*	*t*	$t_{critical}$	*P* (*T* <=*t*)
Sound	5.65	8.48	73	-8.13	1.99	0.00
Visual	9.28	10.45				
Both	9.18	10.63	145	-0.21	1.98	0.83
Visual	9.29	10.32				
Both	9.18	10.63	144	6.91	1.98	0.00
Sound	5.65	8.48				

In the Sound vs. Visual comparisons $|t| = 8.13 > t_{critical} = 1.99$. In the Both vs. Sound, comparisons $|t| = 6.91 > t_{critical} = 1.99$. These results indicate that there is a significant difference between the Sound group and the other two groups. In the Both vs. Visual comparisons $|t| = 0.21 < t_{critical} = 1.98$. This result indicates that there is no significant difference between these two groups. Thus, the difference in performance between the Sound group (at 40% correct) and the others (at 66% correct) is a significant effect.

It is interesting to note that the Sound 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 Sound group replayed the graphs an extra time.

There were eight questions where the Sound group fared particularly poorly in comparison to the other groups. The difference in average percentage of correct responses for questions 1, 2, and 3 was 38, 30, and 37% respectively. This might be attributable to unfamiliarity and lack of training with the display format. If subjects did not follow the optional links on the introductory page, the auditory format may have caused some confusion. The experiment had not been designed to record whether or not subjects had reviewed the supplementary material.

It is clear from Figure 8.1 that all subject groups had difficulty with question 5 of the Main test since the correct response rate was very low. Question 5 displayed a linearly decreasing graph plotted on axes of velocity and time. The question asked about an objects' motion. Analysis of the answers showed that a majority of the subjects (B – 81%, S – 68%, V – 72%) choose the incorrect answer of a linearly decreasing acceleration rather than the correct response of constant acceleration. This question demonstrates subjects' difficulty with concept of acceleration.

Question 6 had the largest difference in the percent correct between the groups (45%). The question displayed a complicated, segmented graph of an object's motion and asked subjects to find where the acceleration was greatest. The large difference in performance may have been due to a lack of training and a misunderstanding of the derivative indicators, or the results could indicate that this was a particularly poor question for auditory graphs.

Questions 11 and 12 involved composite graphs with linear and curved sections and each had a difference of 30%. The sound group may have had difficulty with these questions due to the difficulty representing the value $y = 0$ with sound. Questions 10 (29%) and 14 (36%) involved curved graphs where the Sound group may have again been hindered by lack of training and thus found these graphs confusing.

Several questions show that the auditory format has at least some promise, even when subjects have had virtually no training. Questions 4 and 7 involved linear graphs and had differences of only 14% and 18%. While not perfect, this may still indicate that the auditory format can be used even with very limited examples and training.

Questions 9 and 13 were both graphs of x^2, but had differences of 15% and 25%. The sound group tended to perform somewhat better with these curved graphs than with the others, but the 10% range is troublesome.

Table 8.5 Split-half Analysis for Web Pilot.

Question	% Correct	Split Question	% Correct
1	52	4	67
2	56	7	72
3	44	6	52
5	12	8	52
9	66	13	59
10	60	14	56
11	64	12	61

Split half analysis on the difference between the Sound and Visual groups gives a correlation $r = 0.47$. Using the Spearman-Brown fromula of $r' = 2r / (1+r)$ yields a correlation of $r' = 0.64$. This result shows some consistency between the questions but also shows the effects of the wildly varying performances.

8.7. 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.

This problem was addressed in the Main Auditory Graph test discussed next. 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 Main Auditory Graph test used an expanded set of questions, including separate sections devoted to math or physics based graphs.

9. MAIN AUDITORY GRAPH STUDY

9.1. Overview

The Main Auditory Graph test was the culmination of the techniques used in the pilot tests. Web-based testing techniques and instruments were developed in the Web Pilot test. However, from that pilot test it was evident that a better introduction to auditory graphs and more complex and complete test questions were needed for the Main test section. The Main test questions were rewritten to produce better data for analysis. The number of questions was expanded to include questions concerning mathematical functions, as well as to provide a wider range of questions that would probe subjects' understanding of physics concepts.

The original goal of the auditory graphing method was to provide visually disabled people with a method to quickly access information that is usually portrayed by picture graphs. The Main Auditory Graph test therefore included a small group of blind subject volunteers to evaluate the effectiveness of these graphs for the intended user. This group of subjects was not used in the pilot tests due to the extreme scarcity of subjects fitting the testing requirements. The subject population for this experiment included: undergraduate students from several institutions, graduate students to check the reliability of test questions, and blind volunteers.

9.2. Sample

As the testing process was designed for first-year physics students, instructors of these courses at several 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 Oregon State University (OSU) and one instructor at Pacific University (PU) of introductory, algebra-based, physics courses to provide extra credit homework points to students taking Web Pilot test. An instructor of a calculus-based introductory course at Pacific University also had her students participate for credit. An instructor of an algebra-based physics course at Linn-Benton Community College (LBCC), and a professor of a calculus-based course at OSU mentioned the study and Web address in class but did not offer credit for participation.

OSU physics graduate student subjects were informally solicited throughout 1998 for their participation. Six graduate students took the test using the auditory graph presentation. Graduate students were used as experts in order to provide data about the test's validity. Two additional graduate subjects used the wrong class code and received the test with visual graphs: their scores are not reported with the rest of the data. Two other graduate students attempted the test but due to technical difficulties their scores were not recorded.

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, participation was generally over 50%. 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 Main Auditory Graph test was course size. However, credit was by far the dominant factor.

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. Interested subjects were sent Braille formatted information packets containing tactile graphs, introductory information, and the Web address location. A computer diskette containing the same text as the Braille information was also included in the packet. Blind subjects participated throughout 1998.

A blind physics professor was consulted during test development, and had acted as a critical evaluator of this study. Five blind subjects participated as subjects. 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. From solicitations, 15 interested volunteers provided mailing addresses for the information packets. Of this number, six subjects decided to participate and were able to access the Web page test. One participant was unable to complete the test due to technical difficulties.

The Table 9.1 shows the distribution of the subjects among courses, schools, and approximate course sizes from which they were drawn.

Table 9.1 Distribution of Subjects per Course

Course	# subjects	Approx. Course Total	Date
OSU 203, algebra	189	350	Spring 98
OSU 213, calculus	2	200	Spring 98
LBCC 203	4	20	Spring 98
PU, algebra	28	44	Fall 98
PU, calculus	8	30	Fall 98
Graduate	6	N/A	98
Blind	5	N/A	98

Although there did not seem to be any effect on the test results, of the 189 subjects in the OSU 203 course, 85 had taken the Web Pilot test. 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 since for three test groups, the number of subjects in each group should be greater than 62 (from Equation 6.2).

9.3. Data Collection

Data were collected in a similar manner as in the Web Pilot test. After an initial welcoming and informed consent page, all subjects were given a short tutorial on the auditory graph presentation method with several examples for them to try. The tutorial consisted of a series of graph descriptions, images, and sound files of increasingly complex auditory graphs for them to experience. After the introductory page, there was a log-in page to record the subject's name and class code. PERL script programs recorded subjects' answers and presented them with subsequent Web question pages in an identical fashion to the Web Pilot test. Material pertaining to the Main Auditory Graph test can be found in Appendix C.

At the end of the test, subjects were presented with a page that thanked them for their participation. This page also contained links to a page of correct answers, an e-mail response form for any comments, informational pages on how the graphs were developed, and to the Science Access Project home page.

9.4. Instrument Development

The Survey and Pre-test were identical to those of the Web Pilot. 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, Math and Physics, 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. One question from the Math section contained a graph that was different from the corresponding graph in the Physics section. In the Math section, the graph displayed point discontinuities while the graph in the Physics section was that of a black body spectrum. 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.

The first question in each section consisted of a linear graph with a slope of zero. Aside from this graph, there were eight pairings of similar graph types. Thus, each graph type would appear 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. The rationale for having two graphs of each group was to allow for a split-half analysis of each subject test.

As this was a somewhat iterative process, questions were developed based on the graphs, and graphs were chosen based on the types of questions that could be asked of them. Several, but not all, of the questions from the pilot tests were used for this test. Questions were also chosen based on a diverse range of physical phenomena and their prevalence in the subject matter of introductory math and physics courses. The graphs and their questions were reviewed by Math, Physics, and Science Education faculty for content validity. The graphs and corresponding questions can be found in Appendix C.5.

There were a few modifications of the Web page display between the Web pilot test and the Main Auditory Graph test. As the auditory graphs had no label for their axes, the range of data values was explicitly stated in the questions. Also, several subjects in the Web Pilot test noted that it was difficult to locate the zero point on the auditory graphs. For this reason, a link to a MIDI file that contained the pitch representing zero was included with the auditory graph. The idea was that subjects could compare the "zero" pitch to the pitches of the auditory graphs. The zero sound for all graphs was identical. After taking this test, a couple subjects commented via e-mail that the technique of having the additional zero sound file was not particularly helpful.

Other changes to the test included annotating all images using the "alt" tag field. The graph images were produced with Microsoft Excel 5. Equations in the test were displayed using small graphic images of the equations. These images were created with Microsoft Word'97. All equation images were alt tagged with a linear notation for the mathematics.

The entire test was checked for compatibility with the JAWS screen reader and with Microsoft's 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.

The introductory material and Pre-test contained visually presented graphs. Blind subjects had been sent information packets containing these graphs. The graphs were represented as high-resolution tactile graphic images and were produce by the TIGER printer at OSU. Unfortunately, the informational packets were often ignored and the Pre-test questions went unanswered by a majority of the blind volunteers.

9.5. Results

Table 9.2 is a summary of more complete results for the Main Auditory Graph test contained in Appendix C.6. The table is divided 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, equation 2.2 yields a limit on the error of the averages:

$$L = \left(1.96\frac{\sigma}{\sqrt{n}}\right) = 1.96\frac{0.19}{\sqrt{74}} = 0.04 = 4\%. \qquad (9.1)$$

This result is the 95% confidence limit that the average values for each question are correct to within 4 percentage points. For example, there is a 95% certainty that the Main test question number 33 for the Sound group is between 60% and 68%.

While Table 9.2 provides an accurate listing of the data, it is helpful to view the same data as a bar chart in order to recognize patterns in the data and to easily see where any difficulties may lie. For example, question 34 has an unusually low result for all groups and requires careful analysis of its data. The original data for this question showed an even distribution of answer choices indicative of random guessing. Thus, the data show the effect of a poorly written question.

It should also be noted that the Sound group does 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.

Table 9.2 Table of Percentage of Correct Answers per Group for Each Problem.

Question	V - Visual	B - Both	S - Sound	G - Grad	N - Blind
Pre-test					
p1	72%	83%	78%	100%	20%
p2	95%	96%	97%	83%	20%
p3	93%	89%	93%	83%	20%
p4	67%	81%	89%	100%	20%
p5	55%	68%	66%	100%	20%
Main Test Math					
m1	84%	64%	57%	100%	100%
m2	86%	79%	46%	83%	100%
m3	82%	80%	77%	100%	100%
m4	78%	79%	76%	100%	100%
m5	80%	83%	69%	100%	100%
m6	61%	42%	24%	67%	80%
m7	83%	81%	70%	100%	100%
m8	76%	78%	68%	100%	80%
m9	66%	68%	45%	100%	60%
m10	62%	75%	55%	83%	80%
m11	58%	58%	36%	100%	80%
m12	38%	38%	14%	67%	40%
m13	68%	65%	49%	83%	80%
m14	28%	17%	22%	83%	40%
m15	66%	63%	59%	83%	80%
m16	49%	46%	46%	100%	80%
m17	30%	26%	28%	100%	80%
Physics Section					
m18	57%	51%	41%	83%	100%
m19	42%	35%	35%	33%	60%
m20	21%	28%	19%	83%	80%
m21	41%	41%	26%	83%	20%
m22	76%	81%	84%	100%	100%
m23	62%	44%	58%	50%	80%
m24	72%	79%	62%	100%	100%
m25	70%	62%	49%	100%	100%
m26	58%	62%	36%	100%	60%
m27	63%	64%	50%	100%	100%
m28	54%	58%	27%	100%	100%
m29	54%	43%	35%	83%	40%
m30	45%	48%	54%	17%	80%
m31	37%	36%	42%	83%	60%
m32	72%	70%	62%	100%	80%
m33	72%	70%	64%	100%	100%
m34	18%	16%	23%	17%	0%

The following charts display the percent correct scores for each testing group vs. the individual test questions. The charts are divided by test section. For Figure 9.1, it should be noted that only one blind subject completed the Pre-test, but that all his answers were correct.

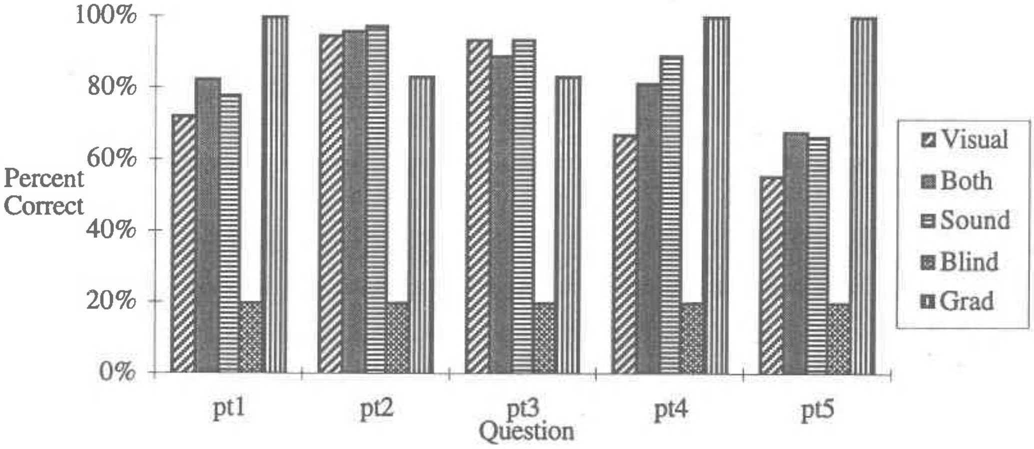


Figure 9.1 Pre-test: Average Percent Correct per Group.

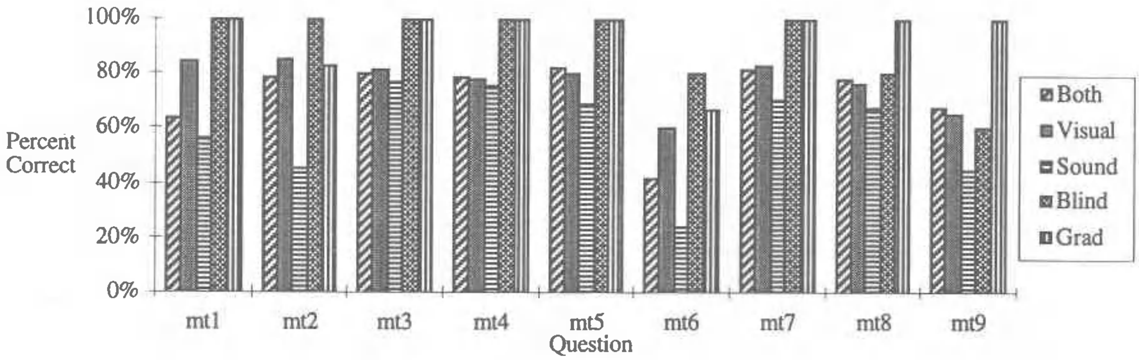


Figure 9.2 Math Section: Average Percent Correct per Group. Questions 1-9.

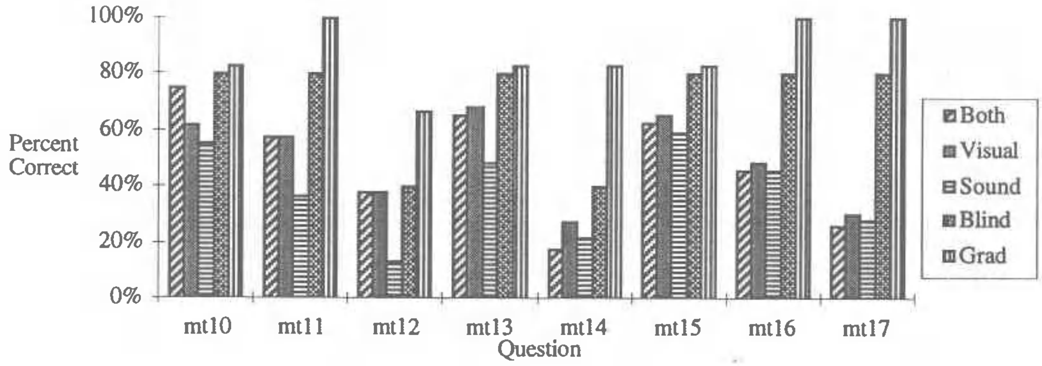


Figure 9.3 Math Section: Average Percent Correct per Group. Questions 10-17.

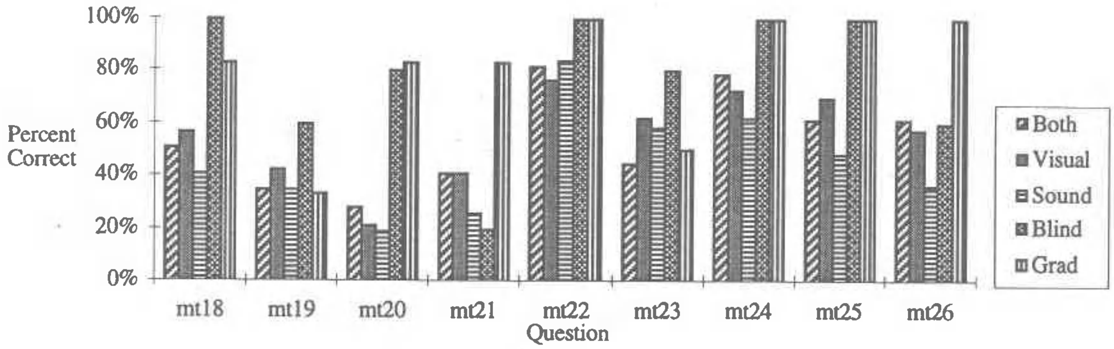


Figure 9.4 Physics Section: Average Percent Correct per Group. Questions 18–26.

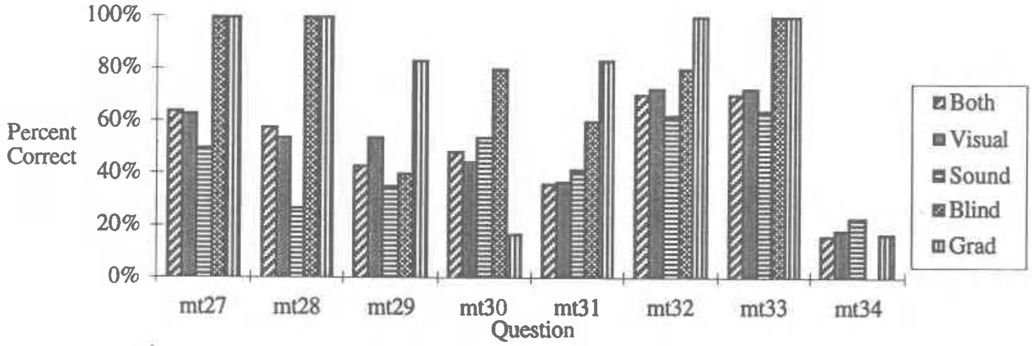


Figure 9.5 Physics Section: Average Percent Correct for Questions 27-34.

The average values for the test sections, standard deviations of the averages, and average time for test completion are given in Table 9.3. The average Pre-test score for the Blind group reflects the result that only one blind subject completed the Pre-test. The large average time for the Blind group was due to several of the subjects starting part of the test, and returning a day or two later to complete the test as their schedule permitted. The average time for the two blind subjects completing the test in one day was 79 minutes.

Table 9.3 Raw Average Percent Correct per Section per Group.

Group:	Both	Sound	Visual	Grad	Blind
Average, Pre-test	83%	85%	77%	93%	20%
Standard deviation (σ)	10%	13%	17%	9%	0%
Average, Main	57%	47%	59%	85%	78%
σ	20%	18%	19%	23%	25%
Average, Math Section	61%	49%	64%	91%	81%
σ	21%	20%	19%	12%	19%
Average, Physics Section	52%	45%	54%	78%	74%
σ	18%	17%	18%	30%	31%
Average time to Complete	30 min.	34 min.	24 min.	40 min.	41 hrs.

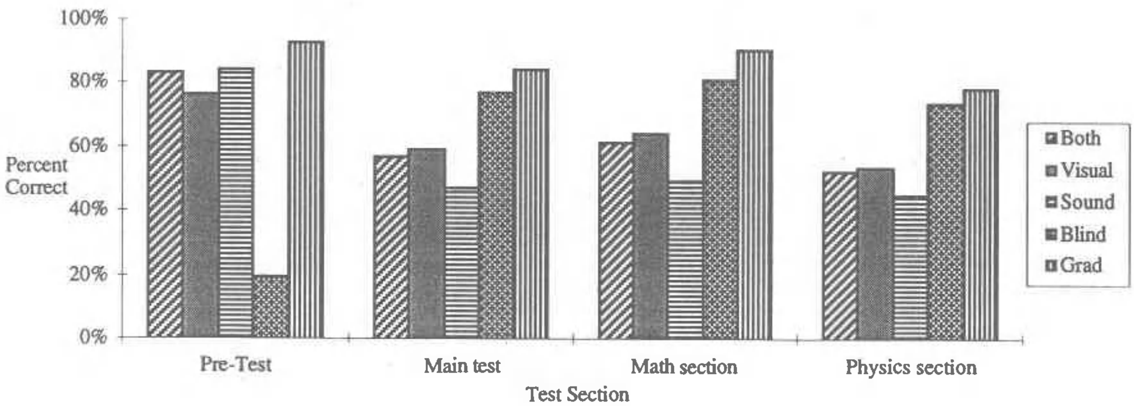


Figure 9.6 Raw Average Percent Correct per Group for Each Section.

The scores in Table 9.3 and Figure 9.6 represent raw averages. That is, they are not corrected for the possibility of subjects randomly guessing answers. To account for this possibility, the scores are modified as noted by Equation 2.9 in section 2.1.9. For the Pre-test, there were an average of seven answer choices, thus the adjusted Pre-test score becomes: $C' = C - W/6$, where C' is the corrected score, C is the percent correct, and W is the percent wrong. For the Main test, there were five answer choices so $C' = C - W/4$.

The adjusted scores are listed in Table 9.4 and shown in Figure 9.7.

Table 9.4 Average Percent Correct per Section per Group Corrected for Guessing.

Group:	Both	Sound	Visual	Grad	Blind
Average, Pre-test	81%	83%	73%	92%	9%
Average, Main	46%	34%	49%	72%	72%
Average, Math Section	52%	37%	55%	76%	76%
Average, Physics Section	40%	31%	42%	68%	68%

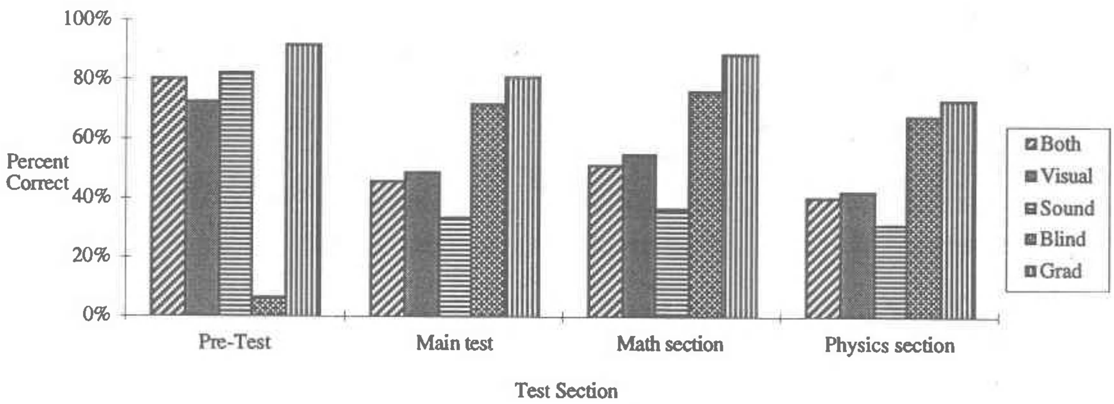


Figure 9.7 Average Percent Correct per Group for Each Section Corrected for Guessing.

9.6. General Analysis of Data

The data were analyzed in several ways. Section 9.6.1 is a description of relevant checks on the Main Auditory Graph test's validity. Section 9.6.2 is a description of issues relating to the reliability of the Main test. In addition, there are sections describing the effect of musical training on performance, the length of time for test completion, and relative performance between the test groups.

9.6.1. Validity Issues

9.6.1.1 Criterion-related Validity

Concurrent validity is a criterion-related validity. This form of validity is established by correlating a new test with a well established test. Unfortunately, there is no similar test for comparative purposes so this form of validity was not well established. Some of the questions in this experiment were modified from questions used in previous research, while the rest were developed to be of a similar format and style. The graph types for questions 1, 2, 3, 7, 8, 11, 12, 18, 19, 20, 24, 25, 28, and 29 were similar in the overall trend as those used in the Flowers study [Flo95], but questioning method was quite different.

The best comparisons of the questions that can be made are Pre-test questions 4 and 5, and Main test questions 18, 19, 20, and 21. For comparative purposes the results from the Visual group are cited as that group is closest to the testing method in Beichner's [Bei94] study. The percent correct values quoted in Beichner's study are approximate as they were displayed in a pie-chart format rather than tabulated values. The wording and graph for Pre-test question 4, which had 80% correct, was similar to Beichner's question 2 with approximately 65% correct. Pre-test question 5 and Main test questions 18, 19, 20, and 21 were variations on the theme of Beichner's questions 3 and 21. Pre-test question 5 had 55% correct while Beichner's questions 3 had 60% correct. The standard deviation for the Visual group's Pre-test was 17%.

Main test question 18 (with 57% correct) used linear graph with zero slope instead of the linearly increasing graph in Beichner's question 3 (with 60% correct). Main test question 19 (42%) was also similar to Beichner question 3 (60%) although it had velocity instead of distance for an axis. Main test question 20 (41%) was similar to Beichner's question 21 (20%) although the wording of the answers was altered. The average standard deviation for the Visual group's Main test was 19%.

The differences in the scores between this experiment and that of Beichner's study may reflect the changes in the wording of the questions, graph type, and axes. However, the results are almost all within one standard deviation and are an acceptable result. Thus, while the test as a whole cannot show concurrent validity, there are indications that at least some of the questions used demonstrate this trait.

9.6.1.2 Face and Content Validity

The objective of the test was stated when volunteers were requested. The objective was also explicitly stated in the welcoming Web page's initial paragraph and again in the Statement of Informed Consent. Thus, subjects were aware of the purpose of the test and face validity established.

Professors in the Math, Physics and Science Education departments at OSU, provided initial input during question development and reviewed the final test questions for appropriateness and content. Several physics graduate students also provided review and commentary of the test. Thus, there is evidence that content validity is established.

9.6.1.3 Construct Validity

Construct validity is a statement about how a measurement performs in accordance with theoretical expectations. Unfortunately, there is no well established theory relating auditory and visual graphs for use as a construct to generate test items. However, a hypothetical relationship was employed. This relationship stated that physics students could use auditory graphs at a performance level equal to that from visual graphs when answering questions that may encounter. Empirical evidence can be gathered by examining the performance difference between expert s and novices. In the case of the Main Auditory Graph test, graduate students served as the experts, and the first-year subjects served as the novices.

Graduate student subjects were solicited as subjects because of their experience with the physics material and graphs. They received the test with auditory graphs to determine what would be the best expected results for students using auditory graphs. If the graduate students consistently missed specific questions, then careful examination of that question's validity would be necessary.

Two graduate subjects inadvertently received the visual test. These subjects each missed one question (number 8 for one, and number 30 for the other) so these questions are of concern. More importantly are the questions where a majority of graduate students using the auditory graphs gave incorrect answers. This occurred for three questions 19, 30, and 34.

Question 8 involved the graph of *1/X*. It is not clear why one of the graduate students taking the visually presented graph test missed this question, but it could be attributed to misreading the question. No other graduate students missed this question so it remains valid.

Question 19 involved a linearly increasing graph with axes representing velocity vs. time. The most common answer (3 of 6) was "D: The object is moving with a constant velocity" whereas the correct response "The object is moving with a constant, non-zero acceleration" was only answered by two subjects. It is suspected that the subjects were not paying close attention to the statement describing the axes values and representation. Given that all other groups (including the S group) outperformed the graduate students on this question, and that the two graduate subjects did answer this question correctly, the question was retained as valid.

Question 30 involved the identification of an intensity pattern produced by a double-slit source. Five of six graduate students (and one of the visual test grads) identified the pattern as that of a single slit source. While there are similarities between the two patterns under certain circumstances, the other subjects groups correctly identified the pattern at a minimum level of 45% correct. The difference in results may have been due to the graduate students attempting to analyze the problem at a theoretical level that was more complex than was necessary for the test. Due to the response rate from the other groups, including the Blind group, this question was retained even though the construct validity may be in doubt.

Question 34 involved a determination the initial conditions for the motion of a mass suspended by springs on a cart. Several of the graduate students mentioned that the question was confusing, and responses from all groups followed a random distribution of the possible answers. Thus, question 34 was a poorly designed question, and was dropped from the analyses.

Recalculating the average correct scores for the Main and Physics sections without question 34 adjusts the Main and Physics sub-test averages. Table 9.5 displays the original average scores while Table 9.6 displays the averages after correcting for guessing.

Table 9.5 Recalculation of Raw % Correct Without #34.

	Both	Sound	Visual	Grad	Blind
Main	58%	48%	60%	87%	80%
Math section	61%	49%	64%	91%	81%
Physics section	55%	46%	56%	82%	79%

Table 9.6 Recalculation of % Correct Without #34 Corrected for Guessing.

	Both	Sound	Visual	Grad	Blind
Main	48%	35%	50%	84%	75%
Math section	52%	37%	55%	89%	76%
Physics section	43%	33%	45%	78%	73%

The graduate students result was over two standard deviations greater than that of the Sound group, and over one standard deviation greater than results of the Visual group. Thus, there is some empirical evidence that the test displays construct validity.

9.6.2. Reliability Issues

9.6.2.1 Split-half Analysis

One method of determining the internal consistency, and hence the reliability, of a test is with split-half analysis as noted in section 2.1.6. To reflect the reliability of the test as a whole, the correlation coefficient r determined from comparing the two test halves is used with the Spearman-Brown formula: $r' = 2r / (1+r)$. When a test is further subdivided, the Spearman-Brown formula becomes: $r' = 4r / (1+3r)$. A value of $r' = 1$ indicates perfect correlation; the two tests are identical. A value of $r' = -1$ indicates that a subject who answered correctly on one sub-test, answered the split question incorrectly and vice versa. A value of $r' = 0$ indicates no correlation between the two sub-tests. Ideally, the correlation value r' should be greater than 0.7 to be considered reliable.

Section 9.4 noted that the Main test had two sections, Math and Physics, containing similar graphs. Each of these sections was designed to be divided into two sub-tests containing graphs of similar nature such as derivative and complexity. For example a graph of $y = x$ was paired with a graph of $y = A - x$, and $y = x^2$ was paired with $y = 1/x$.

Table 9.7 lists the split-half correlation values between and within the Math and Physics sections for each of the groups (B - Both, V - Visual, and S – Sound). Correlation coefficients between the two section tests, and between the sub-test for each section, were calculated for the different groups. The correlation coefficients compared one set of questions, to a second set. The score of each question had been adjusted to account for the possibility of guessing prior to calculating the correlation coefficient. This was accomplished by applying the equation 2.9, $C' = C - W/(A-1)$, to the percent correct is the Sound group whose adjusted correlation value r' reflects the effect of a negative correlation value r. Obviously, this group does not show internal consistency between the Physics sub-tests.

There are very poor correlation values between the Math and Physics sections. Only the Grad group had a result that was close to the acceptable limit. 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. Poor understanding of the physics portrayed in a graph could have played a significant effect on ability to interpret the graph even if the subject could identify the graph.

The split-half analysis for the Math sub-tests demonstrated, for most groups, two equivalent tests. However, it should also be noted that the Grad group did not share the same level of correlation between the two Math sub-tests, in fact this group only had a poor correlation value.

The split-half analysis for the Physics sub-tests demonstrated a less successful attempt at developing two equivalent tests. The Sound group had a greatly reduced correlation result suggesting that the question and graph combination used in this section is least reliable for subjects who are new to auditory graphs and physics. Also, the reduced correlation coefficients may indicate that since the nature of the physics in the questions was different for each question, student understanding and performance ability varied, irrespective of the displayed graph. Subjects were answering questions about physical phenomena, and this was perhaps a more significant effect than the graph type.

9.6.2.2 K-R 20 Analysis

Another method for determining the internal consistency of a test is with the Kuder-Richardson #20 or K-R 20 as described in section 2.1.6. The K-R 20 result varies between 0 and 1 and is interpreted in a similar fashion as r. Table 9.8 lists the results of K-R 20 tests for the Main test well as for separate considerations of Math and Physics sections. The groups shown in the table include all of the undergraduate students as a group (All) as well as each group's separate results. All of the scores have been adjusted for guessing as described in section 9.6.2.1.

As can be seen from the table, all of the results show a high degree of internal consistency for the Main test, as well as each of the test sections. The scores are generally well above the 0.70 acceptance level. Thus, the Main test and the test sections can be regarded as internally consistent.

Table 9.8 K-R 20 Results for the Main Auditory Graph Test.

	All (v, b, s)	Both	Visual	Sound	Blind	Grad
Main	0.91	0.89	0.91	0.91	0.85	0.89
Math section	0.86	0.82	0.87	0.88	0.67	0.77
Physics section	0.84	0.84	0.86	0.82	0.80	0.80

9.6.2.3 Correlation of the Pre-test to the Main Test

The correlation values between the Pre-test and the Main test for each group give disappointing results. The correlation values shown in Table 9.9 are well below the 0.7 limit of acceptability. The low correlation scores indicate that the Pre-test is not a reliable indicator of a group's performance on the Main test. Hence, the Pre-test scores are not useful indicators for providing statements about group equivalencies. Unfortunately, these calculations were not performed for the Pilot tests, so the Pre-test was not modified in the Main Auditory Graph Test to provide more useful results.

Table 9.9 Correlation between Pre- and Main Tests.

All (B, S, V)	Both	Sound	Visual
0.29	0.28	0.46	0.33

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. All of these factors may have contributed to the poor results.

The format of the Pre-test questions was that the questions were all displayed on one web page, and that there were several questions relating to a single graph. The Main test only had one graph displayed at a time, and one question per graph. The average score on the Pre-test was 77% correct while that of the Main test was only 44% correct. This difference indicates a discrepancy in the difficulty of the questions and suggests that the tests were measuring different constructs.

The questions on the Pre-test were not mandatory as subjects could proceed to the next page before answering all questions. While this was not a common occurrence, seven subjects did not answer all of the Pre-test questions. Subjects were not allowed to proceed to the next question on the Main test until they completed their current question. The five Pre-test questions did not cover the breadth of material that the 33 Main test questions did. Only two of the Pre-test questions were similar to questions on the Main test. These questions did not display a high degree of correlation with the Main test.

Thus, the Pre-test acted as more of a familiarization with the testing interface, and not as a reliable indicator of group consistency. The only statement about the group's comparative ability makes the assumption that since the subjects were randomly assigned to the groups, the groups' average performances should be essentially equal if given the same test conditions. While this is not an ideal situation for the assumption, especially since the Pre-test did show a significant difference between groups, the groups were reasonably large so that any performance differences due to assignment to a group is small. The probability that one of the three test methods had the superior student for half or more of the three-student sets can be determined by looking at the possible combinations and their relative probabilities. For example, since each group had about 75 subjects, the total probability is found by summing the product of the number of combinations, the probability of r cases of the superior student being in the chosen group, and the probability of $75 - r$ cases of the student not being in the group [Sne89, p. 112]:

$$\sum_{r=37}^{75} \binom{75}{r} \left(\frac{1}{3}\right)^{r} \left(\frac{2}{3}\right)^{(75-r)} < 0.30\%. \tag{9.2}$$

The result of less than a third of a percent gives a good indication that it is unlikely that there was a significantly uneven distribution of students' abilities between the groups.

9.6.3. ANOVA Comparisons of the S, B, and V Groups

Microsoft Excel was used for ANOVA calculations for the Main test and for the two test sections. The ANOVA calculations were single factor with $\alpha = 0.05$ and compared the differences between the Sound, Visual, and Both groups. The results are given in Table 9.10. The numbers used for the calculation were corrected for guessing in the same manner as in section 9.6.2. Since $F > F_{\text{critical}}$ in all cases, these results indicate significant differences between the groups for each of the tests. Thus, it is worthwhile to make detailed comparisons between the groups using other analysis techniques.

Table 9.10.ANOVA Results

	F	F_{critical}	P-value
Main	10.78	3.04	0.00
Math section	12.84	3.04	0.00
Physics Section	5.23	3.04	0.01

9.6.4. Sheffé Tests for the S, B, and V Groups.

Comparison of the performance between groups can be performed by a series of t-tests. However, in order to perform any number of comparative tests, the Sheffé test is necessary in order to limit the probability of finding an erroneous significant result to at most 5%. The Sheffé test, compares the calculated t value to the critical value of:

$$\sqrt{(a-1)F_{0.05}} \tag{9.3}$$

where a is the number of comparison groups and $F_{0.05}$ is the 5% level of F dependant on the number of degrees of freedom. In this case, $a=3$, and there are 2 and 148 degrees of freedom. P is the probability of randomly finding a t value greater than the value calculated from the data. As can be seen in Table 9.11, all of the comparisons with the Sound group lead to significant differences. The comparison of the Visual group to the Both group does not have a significant difference. The data were compared using values adjusted to account for guessing.

Table 9.11 Sheffé Tests

group pair	test	$\lvert t \rvert$	t_{crit}	P	Sheffé	Sheffé P	significant
V-S	main	4.26	1.98	0.00	2.47	0.00	*
	math	4.60	1.98	0.00	2.47	0.00	*
	physics	3.02	1.98	0.00	2.47	0.01	*
B-S	main	3.71	1.98	0.00	2.47	0.00	*
	math	3.99	1.98	0.00	2.47	0.00	*
	physics	2.69	1.98	0.01	2.47	0.03	*
V-B	main	0.80	1.98	0.42	2.47	0.72	
	math	0.99	1.98	0.32	2.47	0.61	
	physics	0.45	1.98	0.66	2.47	0.90	

A t-test comparison between Grad (G) and Blind (N) groups showed that the 7% difference in the results was not significant ($t = 0.97 < t_{critical} = 1.99$). This result should be viewed with caution as the subject sample size was very small for each of these groups ($n = 6$). Equation 2.2 gives an estimate that the average for each question is correct to ±18%, so a significant result may be masked by this uncertainty.

9.6.5. Effect of Music Training

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 2.2 gives a measurement error of less than 8%. F -test results give $F = 3.04$ and $F_{critical} = 3.97$ with $P = 0085$ at the $\alpha = 0.05$ level. Since $F < F_{critical}$, there is not a significant difference between the groups. A two tailed t-test at the $\alpha = 0.05$ level also shows that $|t| = 1.74$ and $t_{critical} = 1.99$ with $P = 0085$. When $t < t_{critical}$ 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.

9.6.6. Test Completion Times

The Main Auditory Graph test showed a difference between the average times taken by the Sound and Both groups for completing the test. This difference was similar to that seen in the Web Pilot. The time difference between the groups for the whole test was 3.1 minutes, or about 6 seconds per question. This is about the length of time required to listen to the auditory graph once. Thus, the average subject in the Both group either did not listen to the sound graphs, or the average subject in the Sound group played the graphs an additional time. The time difference between the Grad and Sound groups was 6 minutes, or about 10 seconds per question. Thus, the graduate students may have listened to the sound graphs an additional time or given more consideration to the questions.

9.6.7. Blind Subject Performance

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 between the Blind group with the first-year students, there is no method for evaluating group equivalencies. Therefore, ANOVA, Sheffé or *t* -test comparisons between the Blind group and the student groups were not performed.

Although comparisons between the Blind and Grad groups are speculative, and should be viewed only as anecdotal evidence, ANOVA tests give no indication of significant differences. These results are displayed in Table 9.12.

Table 9.12 ANOVA Comparisons Between Blind and Grad Groups.

	F	$F_{critical}$	P	Significant
Main	1.12	5.12	0.32	No
Math	2.34	5.12	0.16	No
Physics	0.25	5.12	0.63	No

There are differences in the average scores between the Blind and Grad groups of 9% for the Main test, 13% for the Math section, and 5% for the Physics section. However, none of the differences between the Blind and Grad groups appear to be statistically significant. This result should be tempered with the reminder that there were only 5 subjects in the Blind group and 6 in the Grad group. 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.

9.7. Conclusion of Main Auditory Graph Test

There was a significant difference between Sound and Visual graph groups. The difference between the average percent correct on the entire test with the scores corrected for guessing was 50 – 35 = 15%. This difference was less than that of 25% observed in the Web Pilot, which was a shorter test and did not have the scores corrected for guessing. The entire test spanned the 17 math and 16 physics questions, with one poorly designed question thrown out due to random answering. These questions had a correct response rate of 50% for the Visual group and 35% for the Sound group. The Sound group thus performed at 70% the level of that shown by the Visual group:

$$\text{Performance Ratio} = 100\% \cdot \frac{(\text{Average Sound Score})}{(\text{Average Visual Score})} = 100\% \cdot \frac{35}{50} = 70\% \qquad (9.4)$$

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 Main 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 OSU.

Expert physics students were able to effectively use the auditory graphs to answer questions at an average level 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 ± 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 a level of 75% correct. While this was not at the

97% level of the two sighted graduate students, it was considerably more than the 50% level of the Visual 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? These questions are explored in the next chapter.

9.8. Subject Comments About the Auditory Graphs.

Finally, this chapter will end with several comments made by several test subjects. 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 need improvement.

"It's easy to picture the graph being presented with audio tones."

"In general your audible graphs are the greatest thing I've 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)!"

10. AUDITORY PREFERENCE PILOT TEST

10.1. Overview

While the Main Auditory Graph test was an effective test using auditory graphs, many unanswered questions arose. First, there were several assumptions inherent in the auditory representation. An example of an arbitrary decision in the Main Auditory Graph test was that 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 preference tests such as this can be used to guide the development of software that uses auditory graphs. The Main 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.

10.2. Sample

There were 13 subjects who participated in this study. As this was a Web-based test similar to the Web Pilot, one subject attempted the test from a remote location. Due to technical difficulties, the auditory graphs produced using the AudioPlot method were not active, thus the results from this subject were not included. There were twelve subjects that participated locally who used the same computer, but at different times, for the test. The subjects were solicited primarily due to their proximity to the research location at Oregon State University. The 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 Science Access Project. The subjects were also chosen because they had not been involved in previous auditory graph research. This choice was an attempt to reduce bias due to familiarity with previous auditory graphing techniques.

10.3. Data Collection

Subjects were invited to an office that contained a desk computer with a Web browser displaying the test's introductory page. Due to the nature of the ActiveX components for creating some of the auditory graphs, Microsoft's Internet Explorer was used as the Web browser. Subjects were told briefly what to expect from the test, that the experiment used a Web browser to display a test consisting of nine questions about auditory graphs. They were also shown the controls for adjusting the volume of sound produced by a pair of speakers next to the computer. The investigator indicated that he would be in a neighboring room in case any technical difficulties arose, and left the subject to take the test.

Data collection was then similar to the method used in the Web Pilot and the Main Auditory Graph tests. A Web browser displayed graphs and information and PERL script programs recorded the answers. The test consisted of an introductory page with the Informed Consent Document and a brief description of the test. Next, subjects were presented with a page to record their names. A scripting program appended the information to a file and assigned a code number. 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. Another scripting program appended their code and text answers to a second data file and passed the code and multiple-choice answer to the next page. After completing the last question, the scripting program appended the code number and the string of multiple-choice answers to a third data file.

10.4. Instrument Development

There were 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 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 Main 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 linear scale for mapping 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 to produce the graphs. The AudioPlot graphs produced a smooth, continuously varying tone with optional clicks for the derivative information. The MIDI graphs consisted more of a staccato piano note with a courser resolution. The derivative information was represented with a drum like tone.

It should be noted that a potential remote subject did not participate in this study citing security concerns with ActiveX control modules. The choice of using these controls to generate the auditory graphs on the Web was based primarily on the transport of Visual Basic code written for an updated version of the TRIANGLE graphing calculator. This code was able to be quickly modified to produce the AudioPlot control modules that were incorporated into the Web-based testing environment.

10.5. Data Results

Table 10.1 is a summary of the multiple-choice results for each question. The questions and answer choices are abbreviated for reference. The full text for the questions can be found in Appendix D.3.

Table 10.1 Summary of Answer Choice per Question.

Question	Answer Choice as Percentage of Total A	B	C	D	E	F
1. Gaussian curve: A = AP, B = MIDI, C = both, D = neither	33%	58%	8%	0%		
2. Gaussian curve with derivative: A = low +, high -; B = high +, low -, C = both good, D = neither good	33%	17%	33%	17%		
3. $x \sin x$: A = no change at 0, B = instrument change at 0, C = both good, D = neither good	50%	42%	0%	8%		
4. $x \sin x$: A = AP with deriv., B = MIDI with deriv. and pitch change a 0; C = both, D = neither	33%	50%	0%	17%		
5. Match visual graph of $e^{-x} \sin x$ to AP graph D: A = $\sin x$, B = $\cos x$, C = $x \sin x$, D = $e^{-x} \sin x$, E = $e^{-x} \cos x$, F = none	17%	0%	17%	58%	0%	8%
6. Match visual graph of $e^{-x} \cos x$ to MIDI graph E: A = $\sin x$, B = $\cos x$, C = $x \sin x$, D = $e^{-x} \sin x$, E = $e^{-x} \cos x$, F = none	8%	0%	0%	17%	58%	17%
7. Match AP graph of $\cos x$ to visual graph A: A = $\cos x$, B = $\sin x$, C = $x \sin x$, D = $e^{-x} \cos x$, E = $e^{-x} \sin x$, F = none	75%	0%	0%	17%	8%	0%
8. Match MIDI graph of $\sin x$ to visual graph B: A = $\cos x$, B = $\sin x$, C = $x \sin x$, D = $e^{-x} \cos x$, E = $e^{-x} \sin x$, F = none	0%	92%	0%	0%	0%	8%

9. Likert style 1- 5 preference of $x \sin x$ graph with different sound representations: 1 is bad, 2 is poor, 3 is neutral, 4 is good, and 5 great.	X avg.	std. dev.
A. MIDI	3.75	0.87
B. AP	3.75	1.14
C. MIDI, dx	3.08	0.79
D. AP, dx	3.42	1.08
E. MIDI, 0	3.83	1.34
F. MIDI, dx, 0	3.33	1.44

By equation 2.2, the error associated with the each of the Likert averages can be found. Using a 95% probability limit, the average of the standard deviations ($\sigma_{avg} = 1.12 =$ 28%), and the sample size of twelve subjects,

$$\text{Error} = \left(1.96\frac{1.11}{\sqrt{12}}\right) = 0.63, \tag{10.1}$$

or about 16% since there was a 4 point range (5-1) in the rating scale.

10.6. Analysis

The results would have provided more consistency if a larger number of subjects had been used. Because of the small sample, the results do not provide convincing evidence of the superiority of any of the graphing methods. The purpose of this pilot test was to discover where any difficulties in the testing process may reside and to evaluate the question statements. Thus, this test should be viewed primarily as anecdotal evidence. However, tentative conclusions about the graphing methods can be made. Comparing the results above to the subjects' written comments about the reasons for their choices was very informative and greatly aided the interpretation of the results.

The first question compared MIDI and AudioPlot (AP) representations of a Gaussian curve. These graphs used only the y axis to pitch mapping. The results for question 1 imply that there was a preference (58 to 33%) for the MIDI graph over the AP graph. This is a somewhat surprising result as great effort went to produce a pleasing smooth sound. The commentary is very interesting as unexpected factors played a role in the choice. Subjects choosing the MIDI graph mentioned that it "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.

Several subjects commented that their choice was at least partially based on the frequency ranges of the graphs. One subject who chose the MIDI preference noted that "the greater difference between the maximum and minimum tones made the graph easier to visualize." However, another subject chose the AP graph because "I seem to make the connection better for the higher pitches." Thus, future testing will need to be careful that the different graphing methods display the same range in frequencies.

These choices may also reflect the difference in data mapping methods used by the two auditory plots. As has been noted in previous research by Stevens in Mansur [Man85], 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.

Question 2 investigated the pitch mapping preference for curvature. A very brief description of what the drum tone represented was given at the top of the page. The first graph used a low drum tone to indicate positive curvature and a high drum tone for negative curvature. The second graph had the reverse mapping. The graphs were again of a Gaussian curve. The results were that 33% (four subjects) preferred the first graph (A), while 17% (two subjects) chose the second (B). However, 33% didn't have a preference (C), and 17% didn't like either (D).

The comments provide the additional feedback that those choosing C or D often did so because they found the graphs confusing, or had a difficult time distinguishing between the graphs. Also, two subjects preferred one graph type gave the opposite graph a higher preference rating in question 9. This indicates the necessity for providing better descriptions and for asking the same question about several different graphs to determine some consistency in the responses.

Question 3 investigated the preference of including a change in the graphs' data sound when the y value was negative. This question was developed in response to comments received during the Main Auditory Graph test. For this representation, the data sound of the graph of $x \sin x$ changed from a piano tone for positive values to a harpsichord tone for negative ones. There was a slight but non-significant preference for the tone change. The reasons for not preferring the change are very informative. Cited complaints were 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.

Question 4 compared the graphs of $x \sin x$ between 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, and negative curvature by a high pitch drum. The MIDI graph also incorporated a tone change for negative values. This feature was not included in the AP graph as it did not have a similar display option at the time.

Of the subjects choosing the AP graph (A) and providing comments for question 4, there is an indication that improvements were still desirable. Comments included: "A would be better if the drum pitch had those high harmonics for positive values instead of the negative ones," and "A sound is good to me. ... Sharp pitch is better to me, but this one also needs some different sound to express the ups and downs." Of the subjects commenting on the MIDI graph (B), they cited that their choice was because "the distinct sounds in B were much more clear than in A." Subjects also chose the MIDI graph because of the "negative change and [because] you can pick up the slope/curvature better." There was also a comment by one subject who chose the neither (D) option because "both seemed rather arbitrary in relation to the graph, at least in the derivative department."

Comparing the results of questions 5 and 6, which were graph identification questions, shows identical results both in the number choosing the correct graph, and in the distribution of incorrect responses. In question 5, subjects were asked to match a visually presented graph of $e^{-x} \sin x$ to one of five AP auditory graphs. These graphs included the derivative indicators. In question 6, subjects were asked to match a visually presented graph of $e^{-x} \cos x$, 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 with both methods.

Comments about the AP graphs in question 5 indicated that some subjects found the choices indistinguishable. There were statements of "I started to choose E or D, but really I didn't like any of the choices" from a subject choosing None of the Above (F), and "frankly, a-d sounded all the same" from a subject choosing the correct answer (D). One subject who chose incorrectly, noted a disparity between the choice and their reasoning: "I just like the sound of C the best[;] however, listening to the pitches, it almost seems like the two maximums reach the same pitch, but on the graph, the second one is lower."

Several of the subjects who answered question 6 incorrectly provided interesting comments about their choices on the MIDI graphs. One subject who answered incorrectly indicated that "the drums in the background created confusion as to what was going on." Other comments, such as "A and E sounded nearly the same" from one who chose A, and "E seemed the closest, but the derivative portion seemed wrong" from one who chose F, indicated that several subjects almost chose the correct answer E. One subject who gave an incorrect choice of D noted that "the tempo of the drum was most clear in describing the slope of the line, as was the change in sound describing the negative values of the curve." These comments may demonstrate the effect of attempting a comparative study on auditory graphs without having a training tutorial such as the one in the Main Auditory Graph test.

For questions 7 and 8, which also involved graph identification, subjects were given an auditory graph and were asked to choose between several visual graphs or a "None of the Above" choice. Question 7 asked subjects to match an AP graph of cos x to one of five visual graphs. This question had a correct response rate of 75%. Question 8 matched a MIDI graph of sin x to one of five visual graphs and had a correct response rate of 92%.

In question 7, one subject who answered incorrectly mentioned a difficulty in identifying the starting of the sound. Question 8 would have had a 100% correct score, but the one subject who chose F instead of the correct answer B mentioned that the graph "seemed to mostly fit B, but I don't think the derivative was correct." The greater response rate on question 8 than on question 7 may have been a reflection of subjects gaining experience since the two questions were similar. Having a random assignment of which type of graph is encountered first would reduce this type of ambiguity.

The last question asked subjects to rate different auditory representations of the graph of $x \sin x$ on a Likert scale of 1 to 5, where 1 was bad, 2 was poor, 3 was neutral, 4 was good, and 5 was great. The results are given in Table 10.1, but are a bit vague due to the high standard deviations. All rankings should be viewed as essentially equivalent as the averages were all within the smallest standard deviation. ANOVA analysis of the average Likert scores from question 9 shows no significant difference between the methods at the $\alpha = 0.05$ level ($F = 0.82 < F_{\text{critical}} = 2.35$, $P = 0.53$). All the average scores were between 3 and 4 indicating that the methods could still be greatly improved.

Table 10.2 Ranking of Preferred Graph Types

Rank	Average Rating	Graph Type
1	3.83	MIDI with 0
2 (tie)	3.75	MIDI plain, AP plain
3	3.41	AP with derivative
4	3.33	MIDI with 0 and derivative
5	3.03	MIDI with derivative

Several subjects provided general comments on what they found helpful or annoying. 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 loading." The subject is referring to the fact that the computer had a somewhat noisy fan.

10.7. 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 Main 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 users' 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. However, there were also indications that by the end of the test, some of the distinct display techniques became bothersome. In question 9 one 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.

There are several reasons why the results of Auditory Preference Pilot test are unable to be used to make a determination of which auditory graphing techniques are ultimately preferred. These reasons include the small sample size, the desire to test auditory graph options that had not been implemented in the AuditoPlot controls, and the limited number of test questions.

The use of the AudioPlot controls for graph generation has many powerful advantages. Once the control is loaded on a remote computer, many complex auditory graphs can be produced with little more than embedded commands in a Web page. The use of these controls eliminates the need for pre-produced graphs, and creates a dynamic display where users can provide a more thorough investigation of the graph than from passive listening. The use of Visual Basic to create the ActiveX controls can result in short development times when adding features. Disadvantages of the ActiveX control system are potential security risks for users, the potential for missing support files (.dll files) on user computers, and the limitation to a single platform and virtual limitation to a single Web browser. The use of the JAVA language to create the auditory graphs is a possible candidate to remove the limitations of ActiveX controls.

Future studies will need carefully constructed questions as well as many graphing variables in order to provide definitive answers. A longer set of questions, with repetition of graph types to provide multiple comparisons is highly desirable as subjects tend to change their views about which styles are favored as they gain experience with the graphs. 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.

11. CONCLUSION

11.1. Summary of Conclusions of Test Results

11.1.1. Triangle Pilot

The Triangle pilot test was useful for gaining experience in question development. It also provided a method to gauge the potential of an auditory graph display to be used in a testing environment. From the results of this test, the initial auditory graphing technique was modified in two significant ways. First, in order to accentuate the curvature, a derivative tick mark was added. The tick mark was represented by a drum beat where the tempo of the drum represented the magnitude of the graph's slope. Second, the data were mapped to a chromatic scale rather than the previous linear scaling. The tonal quality of the sound was also modified, but this was a result of using MIDI to implement the sound files rather than from research findings. The testing method was modified after analyzing the results from the Triangle Pilot. Testing was changed from a guided interview method to a Web based test so that subjects would be less influenced by time or environmental conditions. Although the Triangle Pilot demonstrated that there was a difference of 34% between the Sound and Visual groups, the group sizes were far too small for meaningful results.

11.1.2. Web Pilot

The Web Pilot test was important for gauging student participation in a self-guided test. Participation 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 Sound and Visual groups for this test was 25%.

11.1.3. Main Auditory Graph Test

The Main Auditory Graph test effectively compared the performances of several groups of subjects when answering graph based questions. Unfortunately, because of poor correlation of the Pre-test with the Main test, the Pre-test was not a reliable indicator of subjects' performance on the test and was therefore dropped from the analysis. 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 Main Auditory Graph test's 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. However, Split-half analysis did show poor correlation between the Math section and Physics section tests indicating that these sub-tests were not equivalent tests. If a subject performed well in the Math section, there was no guarantee that he or she would perform well in the Physics section.

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 Main 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 group's score.

11.1.4. 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.

11.2. 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 Main 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 effect graph comprehension. These points demonstrate many areas open for future research.

11.3. 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 Triangle Calculator is a scientific graphing calculator for Windows'95. This program is an updated version of the DOS Triangle program used in the Triangle Pilot study and 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. The Auditory Preference Pilot test respondents indicated that this feature, while useful, became annoying after repeated listening.

In addition, the Triangle 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.

11.4. 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. The Main Auditory Graph test demonstrated that subjects do not have to be sighted to accomplish this feat.

Auditory graphs hold great promise as a display technique. The Auditory Preference Pilot test demonstrated some of the many areas that future research can be focused on to provide for even more effective displays. Finally, here are two last quotes from subjects. The first is from the Main Auditory Graph test:

"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."

The last comment is from the Auditory Preference Test demonstrating the accomplishment of the previous subjects' request:

"Again, I really like the negative value changing tone. It really helped to see the graph with my eyes closed."

REFERENCES

[Bar86] Barclay, W. (1986). Graphing misconceptions and possible remedies using microcomputer-based labs. Paper presented at the 7th National Educational Computing Conference, University of San Diego, San Diego, CA.

[Bei94] Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. American Journal of Physics, 62, 750-762.

[Bei96] Beichner, R. J. (1996). The impact of video motion analysis on kinematics graph interpretation skills. American Journal of Physics, 64, 1272-1277.

[Ber94a] Berg, C. A. and Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. Journal of Research in Science Teaching, 31, 323-344.

[Ber94b] Berg, C. A. and Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. Science Education, 78, 527-554.

[Bra87] 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.

[Bre90] Brekelmans, M., Wubbels, T., and Créton, H. (1990). A study of student perceptions of physics teacher behavior. Journal of Research in Science Teaching, 27, 335-350.

[Car79] Carmines, E. G. and Zeller, R. A. (1976). "Reliability and validity assessment." Sage University Paper Series n Quantitative Applications in the Social Sciences, 07-017. Beverly Hills. Sage Publications.

[Ceb97] Cebula, D. Personal communication, July 10, 1997.

[Cha80] Champagne, A. B., Klopfer, L. E., and Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. American Journal of Physics, 48, 1074-1079.

[Cle82] Clement, J. (1982). Students' preconceptions on introductory mechanics. American Journal of Physics, 50, 66-71.

[Coh94] Cohn, E. and Cohn, S. (1994). Graphs and learning in principles of economics. Research on Economics Education, 84, 197-200.

[Dis96a] Di Stefano, R. (1996). The IUPP evaluation: What we were trying to learn and how we were trying to learn it. American Journal of Physics, 64, 49–57.

[Dis96b] Di Stefano, R. (1996). Preliminary IUPP results: Student reactions to in-class demonstrations and to the presentation of coherent themes. American Journal of Physics, 64, 58-68.

[Dil86] Dillon, R. F. and Sternberg, R. J. (1986). Cognition and Instruction. New York. Academic Press, Inc. 241-247.

[Ehr95] Ehrlich, R. (1995). Giving a quiz every lecture. The Physics Teacher, 33, 378-9.

[Eri80] Ericsson, K. A. and Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 215-251.

[Esc97] Escalada, L. T. and Zollman, D. A. (1997). An investigation on the effects of using interactive digital video in a physics classroom on student learning attitudes. Journal of Research in Science Teaching, 34, 467-489.

[Flo92] 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.

[Flo93] 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.

[Flo95] 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.

[Flo97] 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.

[Fry90] Frysinger, Stephen P. (1990). Applied research in auditory data representation. SPIE - Extracting Meaning from Complex Data: Processing, Display, Interaction, 1259. 130-139.

[Gar96] Gardner, J., Lundquist, R., and Sahyun, S. (1996). TRIANGLE: A practical application of non-speech audio for imparting information. Proceedings of ICAD 96 International Conference on Auditory Display. http:// www.santafe.edu/~icad/ICAD96/proc96/gardner5.htm

[Gar98] Gardner, J. and Bulatov, V. (1998) Non-visual access to non textual information through dotsplus and accessible VRML Proceedings of the 15th IFIP World Computer Congress. http://dots.physics.orst.edu/publications/dots_vrml.html

[Gal96] Gall, M., Borg, W., and Gall, J. (1996). Educational Research: An Introduction, 6th Ed. Longman Publishers. USA

[Gil94] Gillan, D. J. and Lewis, R. (1994). A componential model of human interaction with graphs: 1. linear regression modeling. Human Factors, 36,419-440.

[Gla93] Glasson, G. E. and Lalik, R. V. (1993). Reinterpreting the learning cycle from a social constructivist perspective: A qualitative study of teachers' beliefs and practices. Journal of Research in Science Teaching, 30, 187–207.

[Gol95] Goldberg, F. and Bendall, S. (1995). Making the invisible visible: A teaching/learning environment that builds on a new view of the physics learner. American Journal of Physics, 63, 978-991.

[Gra96] Grayson, D. J. and McDermott, L. C. (1996). Use of the computer for research on student thinking in physics. American Journal of Physics, 64, 557-565.

[Hal93] Halliday, D., Resnick, R., and Walker, J. (1993) Fundamentals of Physics, Fourth Edition. John Wiley. New York

[Ham96] Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. American Journal of Physics, 64, 1316-1325.

[Hay96] Hayes, B. (1996). Speaking of mathematics. American Scientist, 84, 110-113.

[Hes92a] Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force Concept Inventory. The Physics Teacher, 30, 141-158.

[Hes92b] Hestenes, D. and Wells, M. (1992). A mechanics baseline test. The Physics Teacher, 30, 159-166.

[Huc96] Huck, S. W. and Cormier, W. H. (1996). Reading Statistics and Research. Harper Collins Pub. New York.

[Huf95] Huffman, D., and Heller, P. (1995). What does the Force Concept Inventory actually measure? The Physics Teacher, 33, 138-143.

[Ica96] ICAD '96: Proceedings of the Third International Conference on Auditory Display. http://www.santafe.edu/~icad/ICAD96/proc96/index.htm

[Kas95] Kashy, E., Gaff, S. J., *et al.* (1995). Conceptual questions in computer-assisted assignments. American Journal of Physics, 63, 1000-1005.

[Ken97] Kennedy, J. M. (1997). How the blind draw. Scientific American, 76–81.

[Ker73] Kerlinger, F. N. (1973). Foundations of Behavioral Research, second ed. Holt, Rinehart and Winston, Inc. New York.

[Kir68] Kirk. R. E. (1968). Experimental Design: Procedures for the Behavioral Sciences. Wadsworth Pub. Co. 99-130, 522-555.

[Kni95] Knight, R. D. (1995). The vector knowledge of beginning physics students. The Physics Teacher, 33, 74-78.

[Kra94] Kramer, G. Ed. (1994). Auditory display, SFI studies in the sciences of complexity, Proc. Vol. XVIII, Addison-Wesley.

[Law87] Lawson, R. A. and McDermott, L. C. (1987). Student understanding of the work-energy and impulse-momentum theorems. American Journal of Physics, 55, 811-817.

[Lei90] Leinhardt, G., Zaslavsky, O., and Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. Review of Educational Research, 60, 1-64.

[Lig81] Light, P. H. and Humphreys, J. (1981). Internal spatial relationships in young children's drawings. Journal of Experimental Child Psychology, 31, 521-530.

[Lin87] 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.

[Lun90] Lunney, D. and Morrison R. C., (1990). Auditory presentation of experimental data. Extracting Meaning from Complex Data: Processing, Display, Interaction, Edward J. Farrell, Editor, Proc. SPIE 1259, 140

[Man85] Mansur, D. L., Blattner, M. M. and Joy, K. I. (1985). Sound-graphs: a numerical data analysis method for the blind. Proceedings of the Eighteenth Annual Hawaii International Conference on System Sciences, 198-203

[May94] Mayer, R. E. and Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. Journal of Educational Psychology, 86, 389-401.

[Mcd87] 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.

[Mcd92a] McDermott, L. C. and Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. American Journal of Physics, 60, 994-1003.

[Mcd92b] McDermott, L. C. and Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies. American Journal of Physics, 60, 1003-1013.

[Mil74] Milroy, R. and Poulton, E. C. (1974). Labeling graphs for improved reading speed. IEEE Transactions on Professional communication, PC–22, 30-33.

[Min95] Minghim, R. and Forrest, A. R. (1995). An illustrated analysis of sonifircation for scientific visualisation. Proceeding of the 6th IEEE Visualisation '95 Conference. 110-117.

[Mok87] 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.

[Moo86] Moore, V. (1986). The relationship between children's drawings and preferences for alternative depictions of a familiar object. Journal of Experimental Child Psychology, 42, 187-198.

[Nis77] Nisbett, R. E. and DeCamp W. T. (1977). Telling more than we can know: verbal reports on mental processes. Psychological Review, 84, 231-259.

[Nun78] Nunnally, J. C. (1978). Psychometric Theory. McGraw-Hill. New York.

[Oha85] O'Hanian, H. C. (1985). Physics. W.W. Norton and Company. New York.

[Pet82] Peters, P. C. (1982). Even honors students have conceptual difficulties with physics. American Journal of Physics, 50, 501-508.

[Pol54] Pollack, I. and Ficks, L. (1954). Information of elemantary multidimensional auditory displays. Journal of the Acoustical Society of America, 26, 155-158.

[Pri74] 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.

[Red94] Redish, E. F. (1994). The implications of cognitive studies for teaching physics. American Journal of Physics, 62, 796-803.

[Rief95] Reif, F. (1995). Millikan Lecture 1994: Understanding and teaching important scientific thought processes. American Journal of Physics, 63, 17-32.

[Rieb90] Rieber, L. P. (1990). Using computer animated graphics in science instruction with children. Journal of Educational Psychology, 82, 135–140.

[Rot97] Roth, W-M. and McGinn, M. K. (1997). Graphing: cognitive ability or practice? Science Education, 81, 91-106.

[Rot96] Roth, W-M. and Tobin, K. (1996). Staging Aristotle and natural observation against Galileo and (stacked) scientific experiment of physics lectures as rhetorical events. Journal of Research in Science Teaching, 33, 135-157.

[Sei91] Seigler, R. (1991). Children's Tinking, 2nd Ed. Prentice Hall. New Jersey. 17-55

[Sne89] Snedecor, G. and Cochran, W. (1989). Statistical Methods, 8th Ed. Iowa State University Press.

[Sou99] SoundMachine (1999). http:// www.anutech.com.au/tprogman/SoundMachine_WWW/welcome.html

[Ste97] Stevens, R. D., Edwards, A. D. N., and Harling, P. A. (1997). Access to mathematics for visually disabled students through multimodal interaction. Human-Computer Interaction, 12, 47-92.

[Ton99] Tontata (1999). MIDIGraphy. http://ux01.so-net.ne.jp/~mmaeda/indexe.html

[Tho90] 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.

[Tro80] Trowbridge, D. E. and McDermott, L. C. (1980). Investigation of student understanding of the concept of velocity in one dimension. American Journal of Physics, 48, 1020-1028.

[Tro81] Trowbridge, D. E. and McDermott, L. C. (1981). Investigation of student understanding of the concept of acceleration in one dimension. American Journal of Physics, 49, 242-253.

[Tru96] Trumper, R. (1996). A cross-college age study about physics students' conceptions of force in pre-service training for high school teachers. Physics Education, 31, 227-236.

[Tuf90] Tufte, E. (1990). Envisioning Information. Cheshire Conn. Graphics Press.

[Tur96] 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.

[Van91] Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. American Journal of Physics, 59, 891-897.

[Ver45] Vernon, M. D. (1945). Learning from graphical material. British Journal of Psychology, 36. 145-158.

[Ver52a] Vernon, M. D. (1952). The use and value of graphical methods of presenting quantitative data. Occupational Psychology, 22-34.

[Ver52b] Vernon, M. D. (1952). the use and value of graphical methods of graphical material with a written text. Occupational Psychology, 96–100.

[Wai92] Wainer, H. (1992). Understanding graphs and tables. Educational Researcher, 14-23.

[Wav89] Wavering, M. J. (1989). Logical reasoning necessary to make line graphs. Journal of Research in Science Teaching, 26, 373-379.

[Wild89] Wildbur, P. (1989). Information Graphics. New York: Van Nostrand Reinhold Co.

[Wils96] Wilson, Cathern M. (1996). Listen: A data sonification toolkit. M.S. Thesis, Dept. of Computer Science, University of California, Santa Cruz.

APPENDICES

Appendix A Materials Relating to the Triangle Pilot Test

A.1. Overview

This appendix contains material used in the Triangle Pilot test. These materials include the Informed Consent form, the Background Survey, the Pre-test, Main test, the answer sheet used to record responses, and Triangle screen images demonstrating the testing environment. Finally, there is a summary of the results from this pilot test.

A.2. Informed Consent Form

Physics Department, Oregon State University

Title of Project: A Comparison Between Auditory and Visual Graphing Methods.

Investigators: Steven Sahyun, Graduate Research Assistant, Physics Department
John Gardner, Professor, Physics Department

This purpose of this study is to determine whether auditory graphing (data representation using sound) is comparable to visually displayed graphs. In particular, this study will be examining how well conclusions can be drawn from auditory graphs vs. visually displayed graphs.

This study involves three parts, a survey, a pre-test, and a main test. The survey is to find out information on factors which may affect your knowledge of graphical material. The pre-test is to find out information on basic graph interpretation skills. The main test will be similar to the pre-test except that there will be more questions. The main test will involve viewing a computer monitor, and possibly listening to sounds generated by the computer. If there are sounds involved, you will be able control the sound volume to a level that you are comfortable with.

Any information obtained from this study will be kept confidential. A code number will be used to identify any test results or other information that is provided. The only people who will have access to this information will be the investigators and no names will be used in any data summaries or publications. For subjects volunteering from a physics course, participation in this study will result in credit towards one laboratory class, as determined by your instructor, but results from the study will not be used to determine credit.

Participation in this study is voluntary and you may either refuse to participate or withdraw from the study at any time. You may stop the study at any time or take a break. However, only completed tests will be used in the study and receive full credit.

If you have any questions about the research study and/or specific procedures, please contact Steven Sahyun, Physics Department, 301 Weniger Hall; the phone number is 737-1712. Any other questions should be directed to Mary Nunn, Sponsored Programs Officer, OSU Research Office, 737-0670.

My signature below indicates that I have read and that I understand the procedures described above and give my informed and voluntary consent to participated in this study. I understand that I will receive a signed copy of this consent form.

____________________	____________________	____________________
Signature of subject (or subject's legally authorized representative	Name of Subject	Date signed

____________________		____________________
Signature of Investigator		Date signed

A.3. Survey

Survey: Background Information:

1. Gender: M F

2. Age:

3. Have you taken a high school physics course? Y N

4. Number of years of college level physics: 0 1 2 3 4 5+

5. Have you taken courses other than physics where graphed data was important?
   Y N
   If yes, please describe:

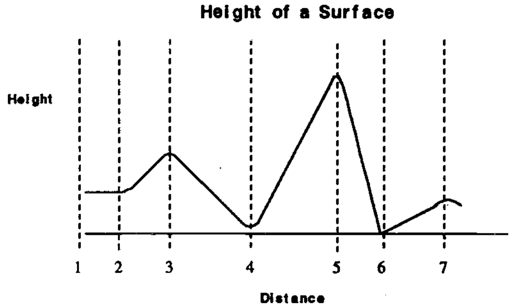
6. Have you learned graphing techniques other than from academic settings?

7. If you have had any musical training, please describe instrument or field and length of time of study:

8. Do you have any difficulties (i.e. vision or hearing) that may affect your ability to receive information?

A.4. Pre-test

Questions 9 - 12 refer to the following graph describing the profile of a surface; the y-axis represents height and the x-axis represents distance.



9. How many relative maxima (locations where a point has a greater y-axis value than the points to either side) are there?

0 1 2 3 4 5 6 7

10. The absolute maximum (greatest y value) is at which point?

1 2 3 4 5 6 7

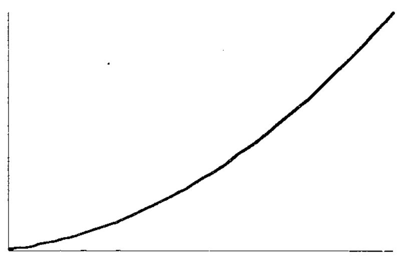
11. The absolute minimum (lowest y value) is at which point?

1 2 3 4 5 6 7

12. The largest magnitude slope (greatest change in y for a change in x) occurs where?

A)	Between points 1 and 2.	D)	Between points 4 and 5.
B)	Between points 2 and 3.	E)	Between points 5 and 6.
C)	Between points 3 and 4.	F)	Between points 6 and 7.

13. The following graph is of an object's motion. The y-axis represents the object's distance, and the x-axis represents time.

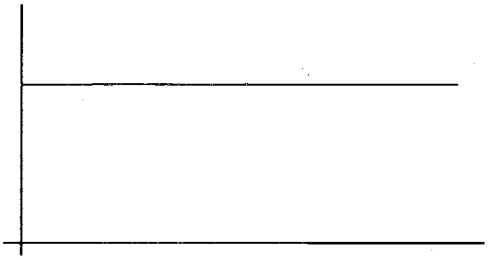


Which sentence is the best interpretation of this graph(please circle answer)?
A) The object is moving with a constant, non-zero acceleration.
B) The object does not move.
C) The object is moving with a linearly increasing velocity.
D) The object is moving with a constant velocity.
E) The object is moving with a decreasing acceleration.

A.5. Main Test

Question 1:

This is a graph of an object's motion. The y-axis represents the object's distance, and the x-axis represents time.

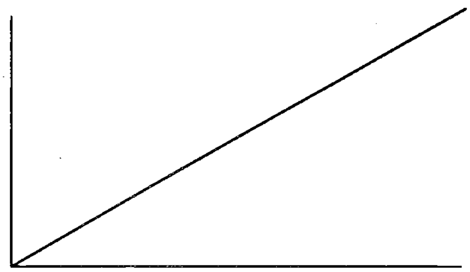


Which sentence is the best interpretation of this graph?

A) The object is moving with a constant non-zero acceleration.
B) The object does not move.
C) The object is moving with a uniformly increasing velocity.
D) The object is moving with a constant velocity.
E) The object is moving with a uniformly increasing acceleration.

Question 2:

This is a graph of an object's motion. The y-axis represents the object's distance, and the x-axis represents time.

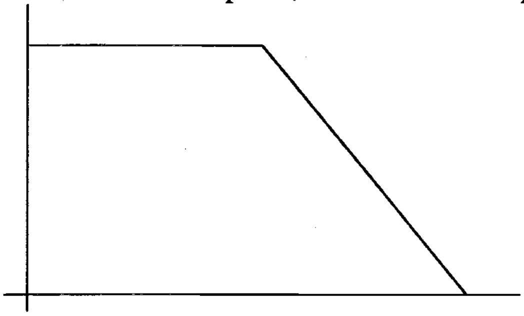


Which sentence is the best interpretation of this graph?

A) The object is moving with a constant non-zero acceleration.
B) The object does not move.
C) The object is moving with a uniformly increasing velocity.
D) The object is moving with a constant velocity.
E) The object is moving with a uniformly increasing acceleration.

Question 3:

This is a graph of an object's motion. The y-axis represents the object's displacement from a reference point, and the x-axis represents time.

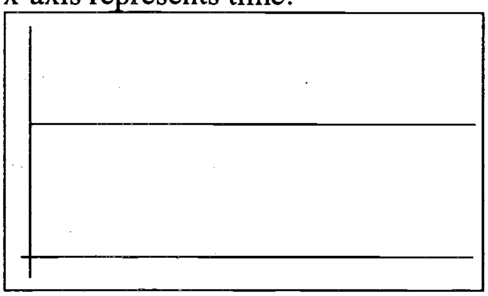


Which sentence is the best interpretation?

A) The object rolls along a flat surface. Then it rolls down a hill, and then finally stops.
B) The object doesn't move at first. Then it rolls down a hill and finally stops.
C) The object is moving at a constant velocity. Then it slows down and stops.
D) The object travels along a flat area, moves down a hill toward the reference point and then keeps moving.
E) The object doesn't move at first. Then it moves towards the reference point and then finally stops.

Question 4:

This is a graph of an object's motion. The y-axis represents the object's velocity, and the x-axis represents time.

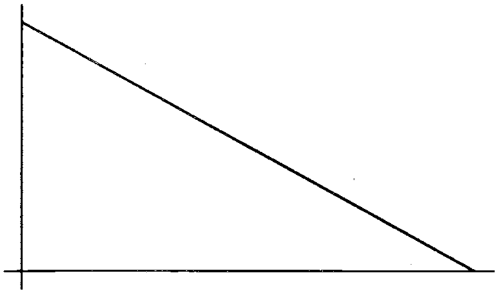


Which sentence is the best interpretation of the object's motion?

A) The object is moving with constant acceleration.
B) The object is moving with uniformly decreasing acceleration.
C) The object is moving with uniformly increasing velocity.
D) The object is moving at a constant velocity.
E) The object does not move.

Question 5:

This is a graph of an object's motion. The y-axis represents the object's velocity, and the x-axis represents time.

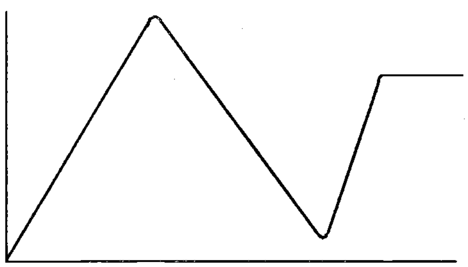


Which sentence is the best interpretation of the object's motion?

A) The object is moving with a constant acceleration.
B) The object is moving with a decreasing acceleration.
C) The object is moving with uniformly increasing velocity.
D) The object is moving at a constant velocity.
E) The object does not move.

Question 6:

This is a graph of an object's motion. The y-axis represents the object's velocity, and the x-axis represents time.

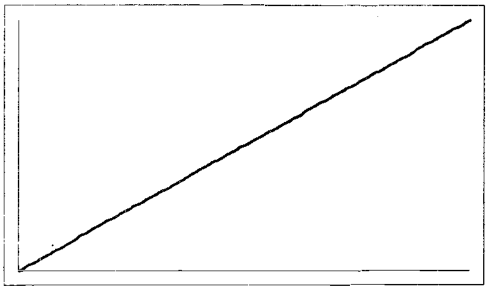


Locate the part of the graph when the acceleration is greatest.

A) Before the maximum velocity.
B) Between the maximum velocity and the next (occurring at a later time) minimum velocity.
C) Between the local minimum and the constant velocity region.
D) During the constant velocity region.
E) The acceleration is never greatest.

Question 7:

The following graph concerns the kinetic energy of a falling ball. The y-axis represents the kinetic energy and the x-axis represents the distance that the ball has fallen.

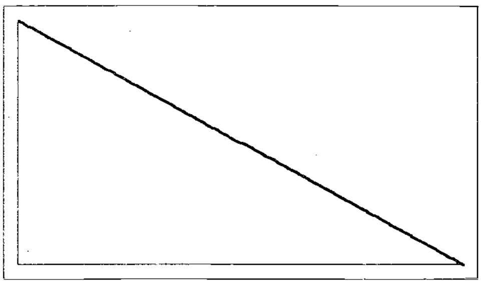


Which sentence is the best interpretation of the graph?

A) The kinetic energy of the ball remains constant.
B) The kinetic energy is increasing linearly with distance.
C) The kinetic energy increases as the square of the distance.
D) The total energy of the ball remains constant.
E) The kinetic energy is decreasing linearly with distance.

Question 8:

The following graph concerns the gravitational potential energy of a falling ball. The y-axis represents the potential energy and the x-axis represents the distance that the ball has fallen.

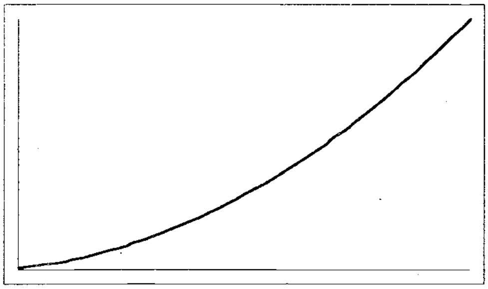


Which sentence is the best interpretation of the graph?

A) The gravitational potential energy of the ball remains constant.
B) The total energy of the ball remains constant.
C) The gravitational potential energy is increasing linearly with distance.
D) The gravitational potential energy increases as the square of the distance.
E) The gravitational potential energy is decreasing linearly with distance.

Question 9:

The following graph is for the elastic potential energy of a spring. The y-axis represents the potential energy of the spring, and the x-axis represents the distance that the spring is compressed.

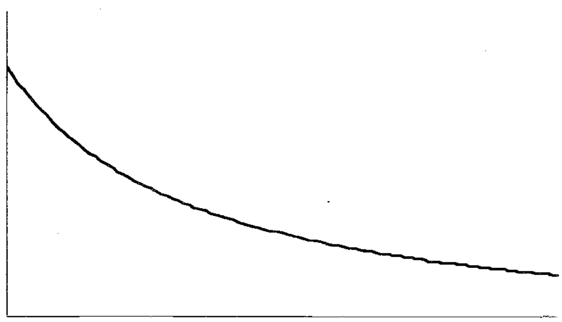


Which sentence is the best interpretation of the graph?

A) The elastic potential energy increases as the square of the distance that the spring is compressed.
B) The elastic potential energy is increasing linearly with the distance the spring is compressed.
C) The elastic potential energy of the spring remains constant.
D) The total energy of the spring remains constant.
E) The elastic potential energy is decreasing.

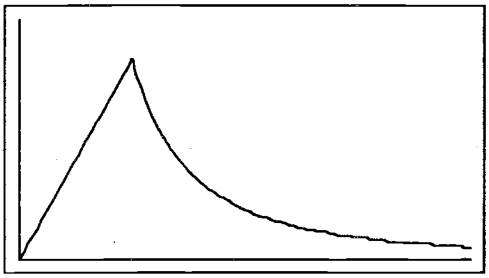
Question 10:

The following graph relates frequency of light to its wavelength. The y-axis represents frequency of light and the x-axis represents the wavelength.



Which sentence is the best interpretation from the graph?
A) The frequency increases linearly with wavelength.
B) The frequency decreases linearly with wavelength.
C) The frequency is proportional to 1 / wavelength.
D) The frequency is constant with the wavelength.
E) The frequency is not related to the wavelength.

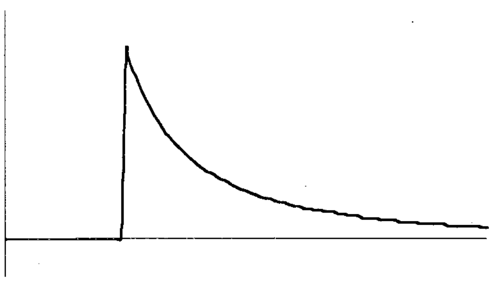
Question 11:
The following graph refers to the gravitational force produced by a solid, uniformly dense sphere acting on a test mass m. The y-axis represents the gravitational force, and the x-axis represents the distance from the center of the sphere.



Which sentence is the best interpretation from the graph?
A) The gravitational force increases linearly with distance.
B) The gravitational force decreases linearly with distance.
C) The gravitational force first increases linearly, then decreases linearly.
D) The gravitational force first increases linearly, then is proportional to 1 / (distance)^2.
E) The gravitational force is not related to the distance.

11b. What does the peak on the graph represent?

Question 12:
The following graph refers to the electric field produced by a charged spherical shell. The y-axis represents the Electric Field, and the x-axis represents the distance from the center of the sphere.

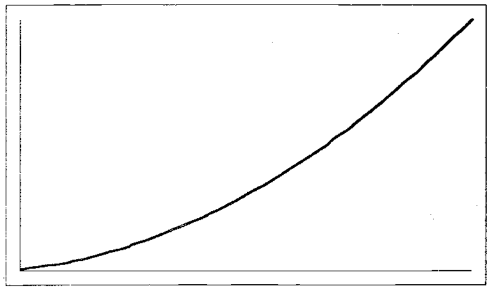


Which sentence is the best interpretation from the graph?
A) The electric field first increases linearly, then decreases linearly.
B) The electric field is zero, then is proportional to 1 / (distance)^2.
C) The electric field increases linearly with distance.
D) The electric field decreases linearly with distance.
E) The electric field is not related to the distance.

12b. What does the peak on the graph represent

Question 13:

The following graph represents the deflection of an electron traveling through an electric field produced by two parallel plates. The electron enters with an initial velocity v_0 and travels from the left to the right. The y-axis represents the distance in the y direction that the electron has traveled, and the x-axis represents the distance traveled in the x direction.



Which sentence best describes the path of the electron?
A) The electron passes through undeflected.
B) The electron is deflected by a linearly increasing distance while between the plates.
C) The electron is deflected proportional to $(\text{distance})^2$ while between the plates.
D) The electron is deflected by a linearly decreasing distance while between the plates.
E) The electron does not pass between the plates.

Question 14:

When water depth is greater than three wavelengths, the speed of ocean waves in water is approximated by the following graph, where the y-axis represents the velocity of ocean waves and the x-axis represents the wavelength.

Which sentence is the best interpretation from the graph?
A) The speed of ocean waves is proportional to √(wavelength).
B) The speed of ocean waves increases linearly with the wavelength.
C) The speed of ocean waves is proportional to (wavelength)2.
D) The speed of ocean waves decreases linearly with the wavelength.
E) The speed is not dependent on the wavelength.

A.6. Answer Sheet

Please mark your answers of the graph test on this page. For multiple-choice questions, please circle the appropriate letter, for short answer questions (11b and 12b) please write your answer on the space provided. If you have any additional comments about the question, please write a note in the area to the right of the answer.

	Answer:					Additional comments:
1.	A	B	C	D	E	
2.	A	B	C	D	E	
3.	A	B	C	D	E	
4.	A	B	C	D	E	
5.	A	B	C	D	E	
6.	A	B	C	D	E	
7.	A	B	C	D	E	
8.	A	B	C	D	E	
9.	A	B	C	D	E	
10.	A	B	C	D	E	
11.	A	B	C	D	E	
11b.						
12.	A	B	C	D	E	
12b.						
13.	A	B	C	D	E	
14.	A	B	C	D	E	

A.7. Screen Image of the TRIANGLE program

This is an image of the screen that was seen by the subjects during the Triangle pilot test. The Sound group had the lower right corner of the screen covered so that they were unable to view the graph.

```
 F10: Menu                          Triangle              Context:(2) sampleq.ext
1. Below is a graph of an object's motion, with the y-axis representing
velocity and the x-axis representing time.

Which sentence is the best interpretation of the graph:

A) The object is accelerating at a constant, positive rate.
B) The object is accelerating at a constant, negative rate.
C) The object is accelerating at an increasing rate.
D) The object is moving at a constant velocity.
E) The object does not move.

- Plot Description ----------------------------

 Plot Cursor  X: 10.000          Y: 100.000

 Plot Scale   X min: 0.00        X max: 10.00
              Y min: -12.00      Y max: 112.00

 Current Plot:  Data Set
```

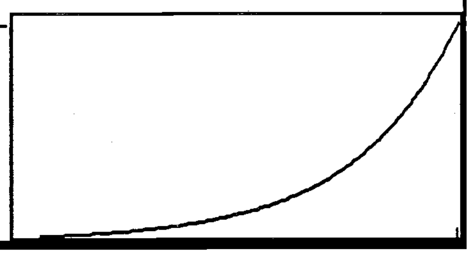


Table A.8.1 Results of the Triangle Pilot Test

Summary		All Subjects					%	Printed Graph					%	Sound Only					%	Sound and Vision					%	Vision Only					%
Gender:	M:	11						2						2						3						2					
	F:	3						1						1						0						1					
H.S. Physics	Y:	10						2						3						2						1					
	N:	4						1						0						1						2					
Pretest:	Ans.	A	B	C	D	E		A	B	C	D	E		A	B	C	D	E		A	B	C	D	E		A	B	C	D	E	
Q 13	E	0	0	0	2	10	83%	0	0	0	1	2	67%	0	0	0	1	2	67%	0	0	0	0	3	100%	0	0	0	0	3	100%
Q 14	A	4	0	3	5	0	33%	0	0	0	3	0	0%	3	0	0	0	0	100%	0	0	2	1	0	0%	1	0	1	1	0	33%
Main Test	Ans.																														
Q 1	B	3	4	0	5	0	33%	0	2	0	1	0	67%	1	0	0	2	0	0%	2	0	0	1	0	0%	0	2	0	1	0	67%
Q 2	D	0	0	6	3	3	25%	0	0	1	1	1	33%	0	0	2	0	1	0%	0	0	1	1	1	33%	0	0	2	1	0	33%
Q 3	E	2	3	4	2	1	8%	0	1	1	1	0	0%	1	0	1	1	0	0%	1	2	0	0	0	0%	0	0	2	0	1	33%
Q 4	D	2	0	0	7	3	58%	2	0	0	0	1	0%	0	0	0	2	1	67%	0	0	0	3	0	100%	0	0	0	2	1	67%
Q 5	A	2	10	0	0	0	17%	0	3	0	0	0	0%	1	2	0	0	0	33%	0	3	0	0	0	0%	1	2	0	0	0	33%
Q 6	C	3	2	7	0	0	58%	2	0	1	0	0	33%	0	1	2	0	0	67%	1	0	2	0	0	67%	0	1	2	0	0	67%
Q 7	B	1	9	2	0	0	75%	0	2	1	0	0	67%	1	1	1	0	0	33%	0	3	0	0	0	100%	0	3	0	0	0	100%
Q 8	E	0	0	0	1	11	92%	0	0	0	1	2	67%	0	0	0	0	3	100%	0	0	0	0	3	100%	0	0	0	0	3	100%
Q 9	A	9	3	0	0	0	75%	3	0	0	0	0	100%	1	2	0	0	0	33%	3	0	0	0	0	100%	2	0	0	0	0	100%
Q 10	C	0	5	5	0	2	42%	0	1	2	0	0	67%	0	3	0	0	0	0%	0	0	1	0	2	33%	0	1	2	0	0	67%
Q 11	D	0	0	2	9	1	75%	0	0	0	3	0	100%	0	0	1	2	0	67%	0	0	0	2	1	67%	0	0	1	2	0	67%
Q 12	B	0	9	0	0	3	75%	0	2	0	0	1	67%	0	2	0	0	1	67%	0	2	0	0	1	67%	0	3	0	0	0	100%
Q 13	C	0	6	6	0	0	50%	0	0	3	0	0	100%	0	3	0	0	0	0%	0	1	2	0	0	67%	0	2	1	0	0	33%
Q 14	A	7	4	1	0	0	58%	3	0	0	0	0	100%	0	3	0	0	0	0%	2	0	1	0	0	67%	2	1	0	0	0	67%
Average	Pretest						58%						33%						83%						50%						67%
	Main Test						53%						57%						33%						57%						67%

A.8. Results of the Triangle Pilot Test

Appendix B Material Relating to the Web Pilot Test

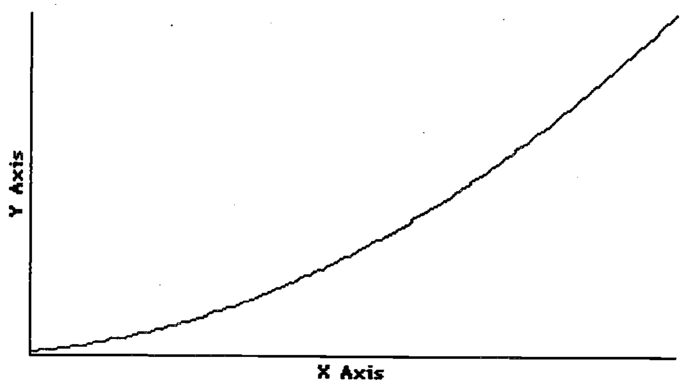
B.1. Overview

This appendix contains material used in the Web Pilot test. These materials include the Introductory page and Informed Consent form, the Background Survey, the Pre-test, and Main test Web pages demonstrating the testing environment. The test questions were similar to those used in the Triangle Pilot test shown in Appendix A. There is a tabulated summary of the results obtained from the Web Pilot test in section B6. Finally, there is a brief analysis of the distribution of test scores.

B.2. Introductory Page and Informed Consent Form

Welcome

The object of this test is to compare students' ability to answer questions from data contained in graphical formats. Two types of graphs are utilized, pictorial graphs which almost everyone is familiar with, and sonified data graphs (data represented by sound), which is much less common. For the pictorial graphs, you will need a web browser such as Netscape or Explorer which can display .gif images. An example of such an image is the following:



The sonified data will be available in several sound formats: QuickTime, MIDI, and .wav files. The QuickTime .mov files are embedded in the pages and appear as a gray bar with controls. After these bars are links to midi and .wav files for downloading. QuickTime is the recommended format as it will be displayed on each page and is most convenient to use. The midi files are small (2 Kb) while the equivalent .wav files are large (about 130 Kb) and take a longer time to load. To hear the sound graphs you will need a multi-media computer capable of generating the sounds, such as a PC with a sound card, or a Mac; and the appropriate software. To listen to the QuickTime files, you will need Apple's QuickTime plug-in for your web browser. The following is a sample sound graph. **Please make sure you can hear this graph before taking the test.**

 MIDI WAV

If you do not see a bar above this line, and if QuickTime is installed, there may have a conflict with is your browser, if this is the case try reloading the page. If you do see the control bar, click on the "play" arrow to listen to the graph. The slider bar gives an indication of how much time has elapsed, you can move the slider to any point in the graph and play from that point.

To listen to the graph as a midi file, click on the MIDI link above; your browser should be configured to automatically start a midi player to listen to the file. You can also download the file by shift-clicking on the link. To listen to the graph as a .wav file, click on the WAV link above; again, your browser should be configured to automatically start a sound player after the file has downloaded. If you can hear the QuickTime graph then you do not need to use these links.

The Y axis of the sound graph is represented by pitch and the X axis is represented by time. The second derivative, or the rate at which the graph is increasing or decreasing is represented by the background "clicks." Play the graph several times and match the sound graph to the picture above.

Go to the Sonitypes Page.

The Sonitypes Page contains a more detailed explanation and basic examples of the sound graphs. Please follow this link to become more acquainted with the basic graph types that will be used in the test. Return here after listening to the graphs.

Statement of Informed Consent

Physics Department, Oregon State University

Title of Project: Comparison Between Auditory and Visual Graphing Methods.

Investigators:
Steven Sahyun, Graduate Research Assistant, Physics Department.
John Gardner, Professor, Physics Department.

This purpose of this study, as stated above, is to determine whether auditory graphing (data representation using sound) is comparable to visually displayed graphs. In particular, this study will be examining how well conclusions can be drawn from auditory graphs vs. visually displayed graphs.

How the test will work:

The format for this test is divided into several parts. First subjects will be asked to give their name and a school code which is used to identify which testing group the student is from.

Next, a demographic survey will be presented to the subject for completion. This is to receive some indication of the subject's background and training. A Pre-test follows the survey and consists of five questions about two different graphs.

Finally, the subject is given a series of questions relating to physics and graphs. The questions are to be answered from the information contained within the graphs, although the subject matter is drawn from material that students are exposed to during a first-year general physics course at a typical college or university. Subjects will be randomly assigned to one of three groups. Those using picture graphs, those using sound graphs, and those using both sound and picture graphs (please listen to the graphs before answering the questions!) There will be 14 questions in the main test. **Only data from subjects who answer all 14 test questions will be used.**

It should be noted that all responses are being transferred on the Internet, and are not encrypted. However, reasonable attempts at confidentiality are made in that the subject's name will not appear with, or be stored with, any of their responses. Names will not be used in any publications or presentations of the data obtained.

It should also be noted that in cases where student subjects are taking this test for extra credit in a course, a list of which students have taken the test will be forwarded to the respective school's instructor so that those subjects may receive credit. Results of the study will not be used to determine credit. Also, a summary listing of the average responses to the test questions will be available to the instructor.

Participation in this study is voluntary and you may either refuse to participate or withdraw from the study at any time, although full participation is greatly appreciated.

If you have any questions about the research study and/or specific procedures, please contact:
Steven Sahyun
Physics Department
Oregon State University
301 Weniger Hall
Corvallis OR, 97330
USA
The phone number is (541) 737-1712. Any other questions should be directed to Mary Nunn, Sponsored Programs Officer, OSU Research Office, Oregon State University, Corvallis, OR, 97331. The phone number is: (541) 737-0670.

After reviewing and agreeing to the above procedure,

Start the test.

Questions about this test? Send me e-mail:
sahyuns@ucs.orst.edu

Last modified October 12, 1997.

B.3. Survey

The survey questions were similar to those used in the Triangle Pilot, but with slight modifications as noted in chapter 7. The survey utilized radio buttons and text box areas for selecting and typing responses. A screen shot of the survey is displayed in Figure B.3.1.

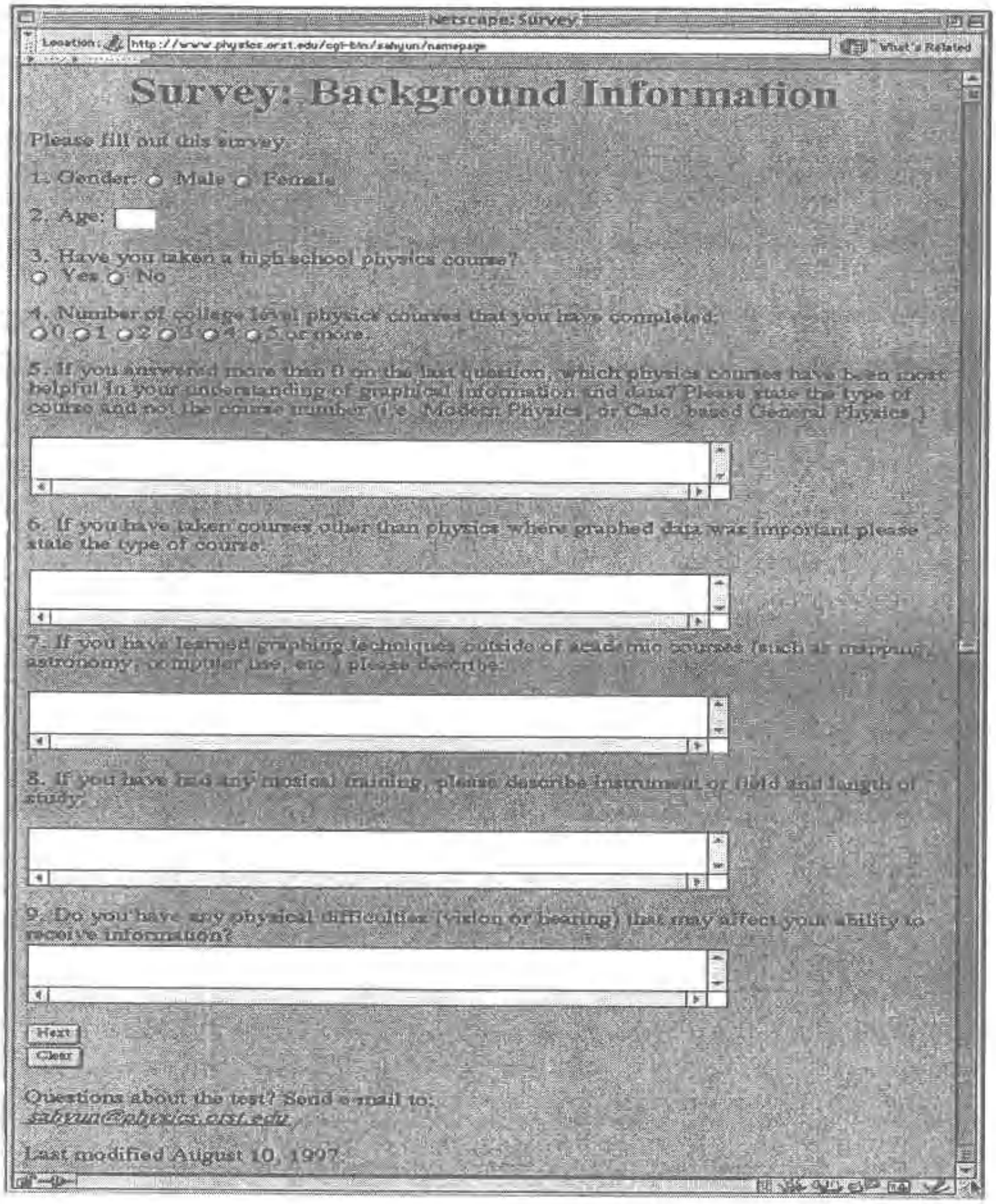


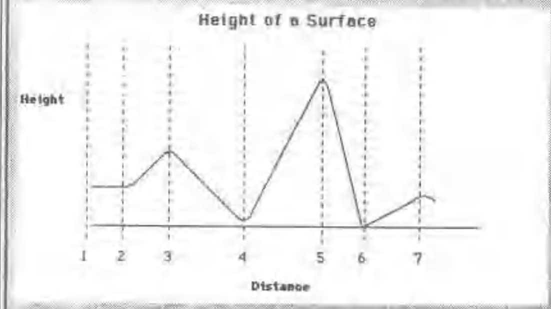
Figure B.3.1. Screen Image of the Web Pilot Survey

B.4. Pre-test

The Web Pilot's pre-test text and images were the same as in the Triangle Pilot, but the layout of the Web page contained radio buttons for answer selection. The pre-test is displayed in Figure B.4.1.

Pretest

The next five questions refer to the following graph describing the profile of a surface: the y-axis represents height and the x-axis represents distance.



1. How many relative maxima (locations where a point has a greater y-axis value than the points to either side) are there?

○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7

2. The absolute maximum (greatest y value) is at which point?

○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7

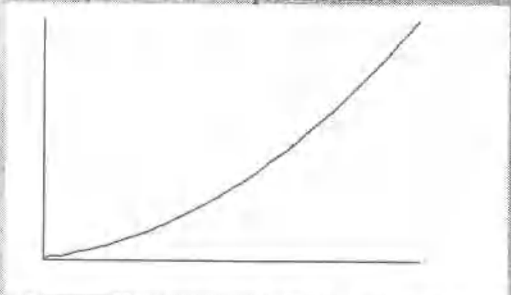
3. The absolute minimum (lowest y value) is at which point?

○0 ○1 ○2 ○3 ○4 ○5 ○6 ○7

4. The Largest magnitude slope (greatest change in y for a change in x) occurs where?

○ Between points 1 and 2. ○ Between points 2 and 3.
○ Between points 3 and 4. ○ Between points 4 and 5.
○ Between points 5 and 6. ○ Between points 6 and 7.

5. The following graph is of an object's motion. The y-axis represents the objects distance, and the x-axis represents time.



Which sentence is the best interpretation of the graph?

○ A: The object is moving with a constant, non-zero acceleration.
○ B: The object does not move.
○ C: The object is moving with a decreasing velocity.
○ D: The object is moving with a constant velocity.
○ E: The object is moving with a decreasing acceleration.

[Done]
[Clear]

Questions about the test? Send e-mail to:
sabyun@physics.orst.edu

Last modified July 20, 1997.

Figure B.4.1 Screen image of the Pre-test.

B.5. Main Test

Figure B.5.1 is a screen image of a typical question page presented to the subjects. The subjects chose their answer via radio button selections. They were required to select one choice before the next question would be displayed. The questions were similar to those displayed in Appendix A but with the modifications as noted in Chapter 7.

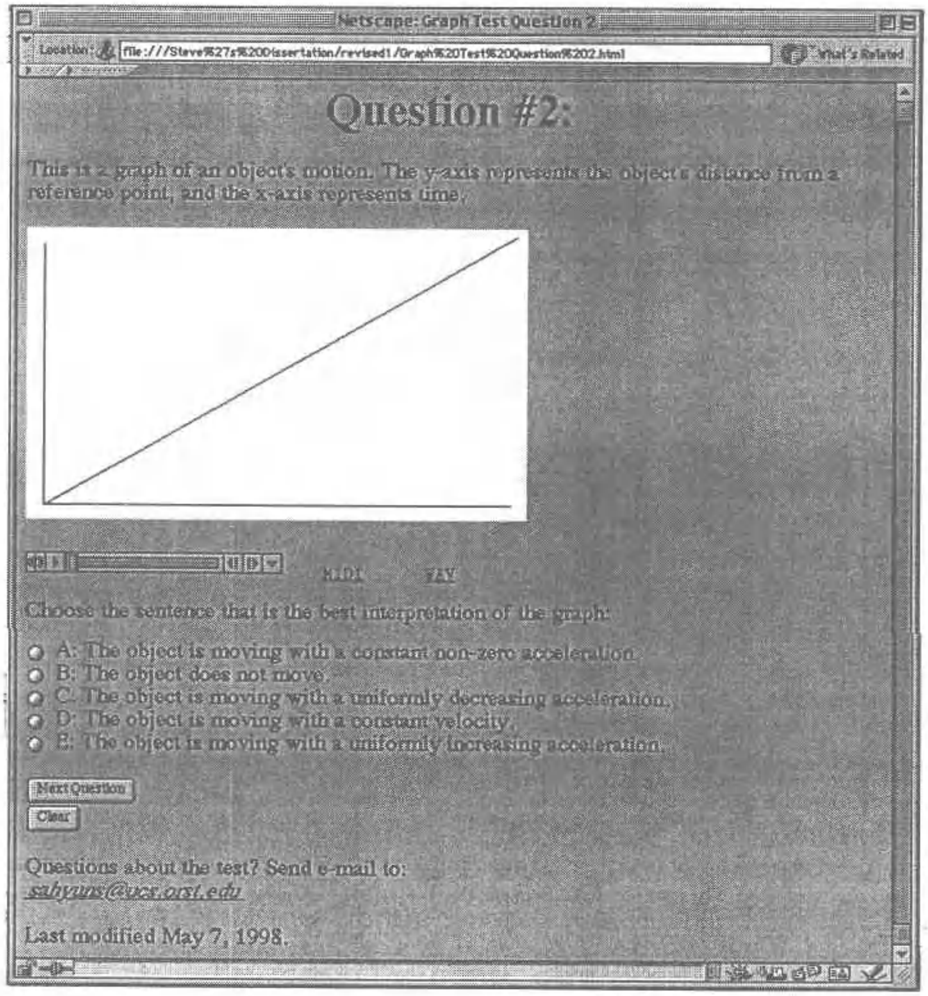


Figure B.5.1 Screen image from the Web Pilot Main Test

B.6. Summary of Results from the Web Pilot Test

Tables B.6.1 and B.6.2 are summaries of the results obtained from the Web Pilot test. The tables are divided into four columns: All, Vision, Both, and Sound. The All category represents the average of all subjects taking the test, Vision represents the subjects given the test with visually presented graphs, Both represents the group given the test with both visual and auditory graphs, and Sound represents the subjects given the test with only auditory graphs.

Table B.6.1 contains the average scores for the test per group, as well as demographic information obtained from the survey. This information included the percentage of females taking the test, the average age, whether or not they had taken a physics course in high school, and whether or not they had taken previous physics courses.

Table B.6.1 . Summary of Results from the Web Pilot Test.

	Group			
	All	Vision	Both	Sound
Pre-test				
Avg.	4.14	4.27	4.00	4.23
std. dev.	0.94	0.89	0.98	0.96
Main test				
Avg.	8.04	9.29	9.18	5.65
std. dev.	3.55	3.21	3.26	2.91
Number of Subjects.	221	72	74	75
Gender %F	54%	60%	43%	59%
Avg. Age	21	21	21	21
H.S. Phys.	59%	63%	61%	53%
Coll. Phys.	16%	21%	15%	13%

Table B.6.2 Summary of Results from the Web Pilot Test, continued.

Question	All	% correct Vision	Both	Sound
Pre-test				
P1	74%	77%	71%	78%
P2	97%	97%	97%	96%
P3	97%	99%	96%	95%
P4	80%	83%	78%	81%
P5	65%	71%	58%	73%
Main Test				
M1	63%	68%	58%	30%
M2	66%	67%	65%	36%
M3	56%	59%	53%	22%
M4	73%	71%	75%	57%
M5	10%	11%	10%	16%
M6	67%	68%	65%	23%
M7	76%	81%	71%	64%
M8	86%	84%	89%	73%
M9	71%	69%	74%	54%
M10	70%	69%	71%	41%
M11	76%	71%	81%	41%
M12	73%	71%	81%	41%
M13	68%	67%	69%	42%
M14	68%	67%	69%	31%
Avg. Time (min.)	14.18	13.28	15.12	16.93

B.7. Histogram and Normal Distribution of Data.

One of the assumptions made when analyzing the data was that it follows a normal distribution population. The normal population is modeled by the Gaussian curve given by:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2} \tag{B.7.1}$$

where X is the average value of the data, and σ is the standard deviation.

Figure B.7.1 compares a histogram of the distribution of the total number of correct responses for all subjects to the Gaussian ideal. Obviously the distributions do not match well, but this can be partially attributed to the low scores from the Sound group as well as the small number of questions. In addition, the fit depends greatly on how the scores are grouped. Figure B.7.1 displays the data when grouped by the number of correct answers.

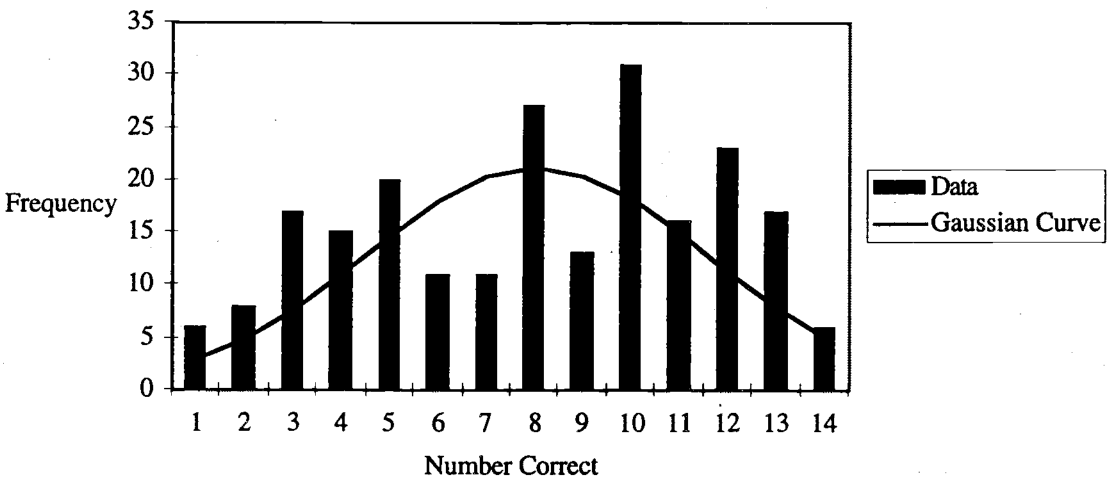


Figure B.7.1. Histogram of Web Pilot Data for All Subjects

A method for comparison of data distributions to the ideal Gaussian distribution is with a Normal Probability plot. In this graph the data is ordered from smallest to largest and then plotted with the points' value for the y axis, and the probability that an observation from a normal distribution is smaller than the data point for the x axis value. [Sne89 p. 59] Ideally, the data points should lie on a line with a slope of $1/\sqrt{n}$. Sometimes the values are plotted as the standard normal deviate, in which case the x axis has units of the standard deviation and the slope is σ.

Figure B.7.2 shows that the distribution deviates from the normal. The fact that the distribution does not completely follow that of the normal may introduce some inaccuracies in the comparative tests as the $F_{critical}$ and $t_{critical}$ values assume normal distributions.

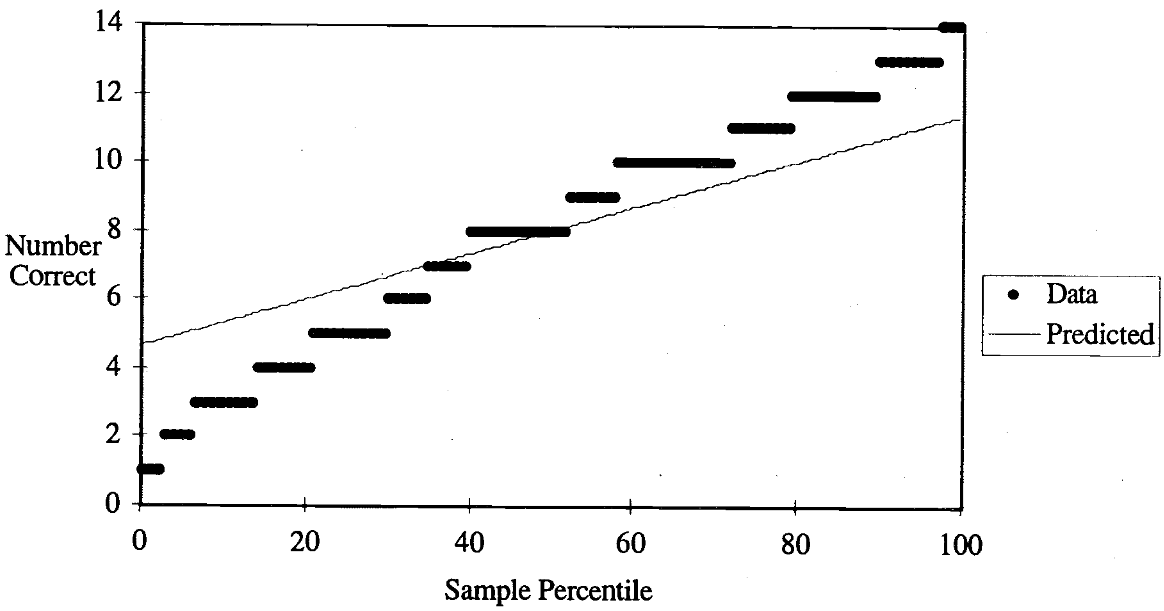


Figure B.7.2. Normal Probability Plot for All Subjects

Appendix C Material Relating to the Main Auditory Graph Test

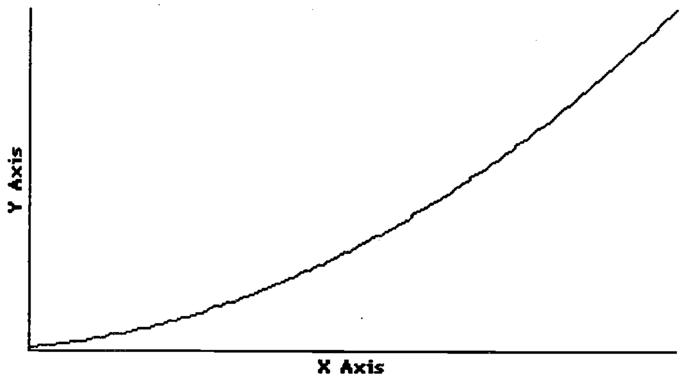
C.1. Overview

This appendix contains material used in the Main Auditory Graph test. These materials include the Introductory page and Informed Consent form, the Introduction to Auditory Graphs Page, a sample page from the Main test and the questions and graphs used in the test. Following the test questions, there is a summary of the results obtained from the Main Auditory Graph test. Finally, there is a brief analysis of the distribution of test scores.

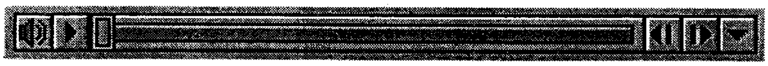
C.2. Introductory page and Informed Consent Form

Welcome

The object of this test is to compare students' ability to answer questions from data contained in graphical formats. Two types of graphs are utilized, pictorial graphs which almost everyone is familiar with, and sonified data graphs (data represented by sound), which is much less common. For the pictorial graphs, you will need a web browser such as Netscape or Explorer which can display .gif images. An example of such an image is the following:



The sonified data will be available in several sound formats: QuickTime, MIDI, and .wav files. The QuickTime .mov files are embedded in the pages and appear as a gray bar with controls. After these bars are links to midi and .wav files for downloading. QuickTime is the recommended format as it will be displayed on each page and is most convenient to use. The midi files are small (2 Kb) while the equivalent .wav files are large (about 130 Kb) and take a longer time to load. To hear the sound graphs you will need a multi-media computer capable of generating the sounds, such as a PC with a sound card, or a Mac; and the appropriate software. To listen to the QuickTime files, you will need Apple's QuickTime plug-in for your web browser. The following is a sample sound graph. Please make sure you can hear this graph before taking the test.


MIDI WAV

If you do not see a bar above this line, and if QuickTime is installed, there may have a conflict with is your browser, if this is the case try reloading the page. If you do see the control bar, click on the "play" arrow to listen to the graph. The slider bar gives an indication of how much time has elapsed, you can move the slider to any point in the graph and play from that point.

To listen to the graph as a midi file, click on the MIDI link above; your browser should be configured to automatically start a midi player to listen to the file. You can also download the file by shift-clicking on the link. To listen to the graph as a .wav file, click on the WAV link above; again, your browser should be configured to automatically start a sound player after the file has downloaded. If you can hear the QuickTime graph then you do not need to use these links.

The Y axis of the sound graph is represented by pitch and the X axis is represented by time. The second derivative, or the rate at which the graph is increasing or decreasing is represented by the background "clicks." Play the graph now to be sure your computer is configured correctly for these sounds. In the following pages, there will be detailed explanations and basic examples on how to listen to these graphs.

Before proceeding, please take a moment to read the following document.

Statement of Informed Consent

Physics Department, Oregon State University

Title of Project: Comparison Between Auditory and Visual Graphing Methods.

Investigators:
Steven Sahyun, Graduate Research Assistant, Physics Department.
John Gardner, Professor, Physics Department.

This purpose of this study, as stated above, is to determine whether auditory graphing (data representation using sound) is comparable to visually displayed graphs. In particular, this study will be examining how well conclusions can be drawn from auditory graphs vs. visually displayed graphs.

How the test will work:

The format for this test is divided into several parts. First, an introduction to auditory graphs is given, with samples and questions about these graphs. This is to provide a common basis of understanding.

Subjects will then be asked to give their name and a school code which is used to identify which testing group the student is from.

Next, a demographic survey will be presented to the subject for completion. This is to receive some indication of the subject's background and training. A Pre-test follows the survey and consists of five questions about two different graphs.

Finally, subjects are given a series of 17 graph questions relating to mathematics, and 17 questions relating to physics graphs. The questions are to be answered from the information contained within the graphs, although the subject matter is drawn from material that students are exposed to during a first year general physics course at a typical college or university.

Subjects will be randomly assigned to one of three groups. Those using picture graphs, those using sound graphs, and those using both sound and picture graphs (please listen to the graphs before answering the questions!)

Only data from subjects who answer all 34 test questions will be used.

It should be noted that all responses are being transferred on the Internet, and are not encrypted. However, reasonable attempts at confidentiality are made in that the subject's name will not appear with, or be stored with, any of their responses. Names will not be used in any publications or presentations of the data obtained.

It should also be noted that in cases where student subjects are taking this test for extra credit in a course, a list of which students have taken the test will be forwarded to the respective school's instructor so that those subjects may receive credit. Results of the study will not be used to determine credit, only the fact that the test has been taken. Also, a summary listing of the average responses to the test questions will be available to the instructor.

The test is estimated to take about 30 - 40 minutes.

Participation in this study is voluntary and you may either refuse to participate or withdraw from the study at any time, although full participation is greatly appreciated. You may take a break at any time, just be sure not to loose the web page question that you are on. If there should be a technical problem (crash) during the test, you will need to start over.

If you have any questions about the research study and/or specific procedures, please contact:
Steven Sahyun
Physics Department
Oregon State University
301 Weniger Hall
Corvallis OR, 97330
USA

The phone number is (541) 737-1712. Any other questions should be directed to Mary Nunn, Sponsored Programs Officer, OSU Research Office, Oregon State University, Corvallis, OR, 97331. The phone number is: (541) 737-0670.

After reviewing the above statement, please click on the link below:

I agree to this test.

Questions about this test? Send me e-mail:
sahyuns@ucs.orst.edu

Last modified April 28, 1998.

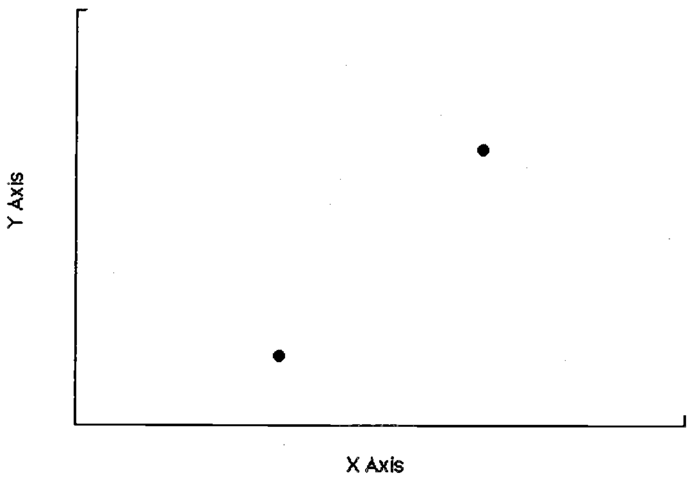
C.3. Introduction to Auditory Graphs

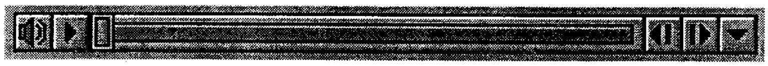
Before taking the test, subjects were given a web page introducing auditory graphs. This web page served as a training method for the auditory graphs. To provide equivalency between groups, all subjects were presented with this page.

Introduction to Auditory Graphs

Before the test begins, a short explanation about Auditory graphs is necessary for those who will be listening to the graphs instead of seeing them. **It is a random process as to who gets which graph type**, so everyone should understand how to listen to these graphs.

The basic auditory graph involves mapping the Y axis data to pitch, and the X axis data to time. So the greater the Y value, the higher the frequency of the sound, and the greater the X value the later the sound will be played. As an example, in the following graph there are two (X,Y) data points: (1,1) and (2,4). The (1,1) point can be heard first, and has a low tone, the (2,4) point is played second and has a higher tone.

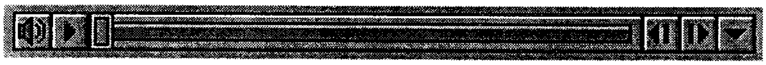




MIDI WAV

Please play this graph now.

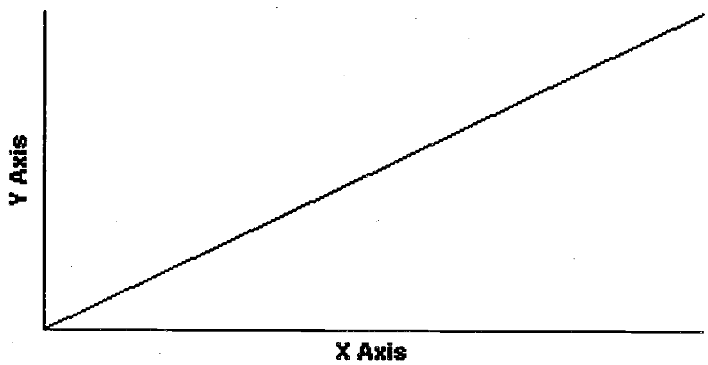
Note: The concept of 0 is difficult to represent in an auditory graph. We have used the following tone to represent 0 in all of the graphs in this test:

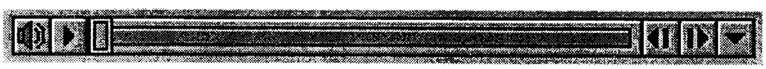


MIDI WAV

For most of the graphs that you will encounter, the lowest tone played will represent 0.

A series of points would be played as a series of tones. The following graph gives an example of a series of points that increase in both X and Y values:





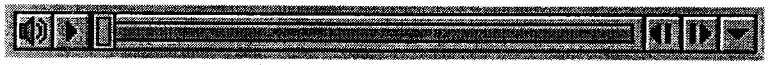
MIDI WAV

Please play this graph now.

In previous studies, it was noticed that it is difficult to tell when a graph is linearly increasing (i.e. $Y = X$), vs. when there is some curvature to the graph (i.e. $Y = X^2$). For this reason, a sound to alert the listener to the slope and curvature of a graph has been added. This sound is heard as a series of "drum" beats.

The slope of a graph is defined as the rise/run or $\frac{dY}{dX}$. The greater the slope, the more rapid the beat.

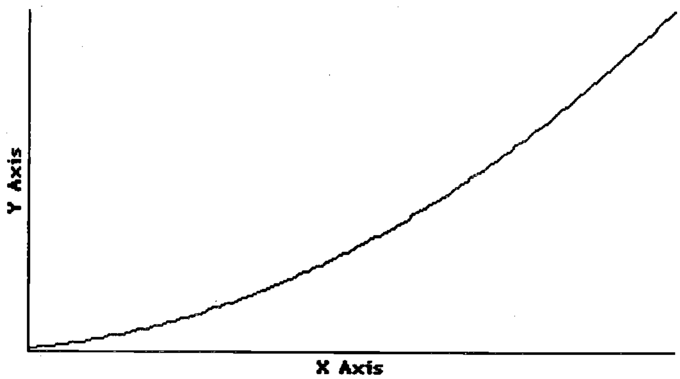
Please listen to the following graphs to determine which one has the greatest slope.

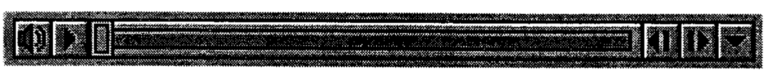
1:  MIDI WAV

2:  MIDI WAV

(The second graph has a greater slope)

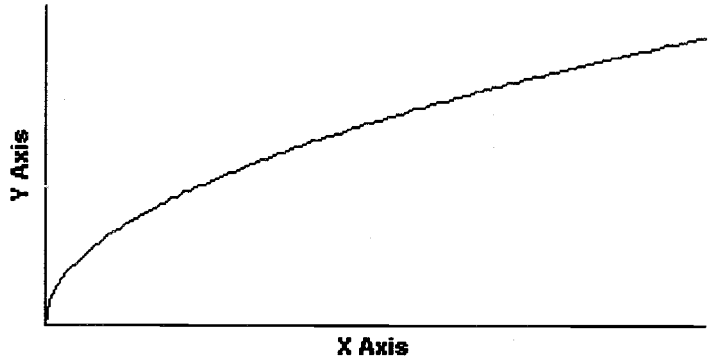
The pitch of the drum beat indicates the curvature of the graph. The curvature is defined as the change in the slope, or $\frac{d^2Y}{dX^2}$. When the curvature is positive ($\frac{d^2Y}{dX^2} > 0$), as it is for the graph of $Y = X^2$, the graph is bowl shaped, and is represented with a low pitched drum beat. This graph looks and sounds like the following:

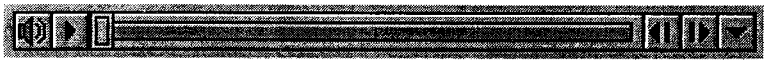


 MIDI WAV

Please play this graph now.

When the curvature is negative ($\frac{d^2Y}{dX^2} < 0$) as in $Y = \sqrt{X}$, the graph is hat, or hill shaped. This type of graph has a high pitched drum beat. This graph looks and sounds like the following:



 MIDI WAV

Please play this graph now.

When there is no curvature ($\frac{d^2Y}{dX^2} = 0$), as was seen in the linear graphs above, (remember, the graph can still have a non-zero slope) the pitch of the drum beat is between those of the positive or negative curvature graphs.

If you would like to see and hear more examples, please go to the **Sonitype page**.

If this method of auditory graphing is clear, you're ready to

Start the Test

C.4. Pre-test and Survey

The Pre-test and Survey questions and layout were identical to those used in the Web Pilot test. For descriptions and screen images, see Appendix B, sections 3 and 4.

C.5. Main Auditory Graph Test

The question layout for the Main test questions was similar to the layout for the Web Pilot test, although an extra sound bar was included for use as a reference to 0. Figure C.5.1 is a screen image of a typical question page presented to the subjects. The questions of the test and their related graphs follow the figure.

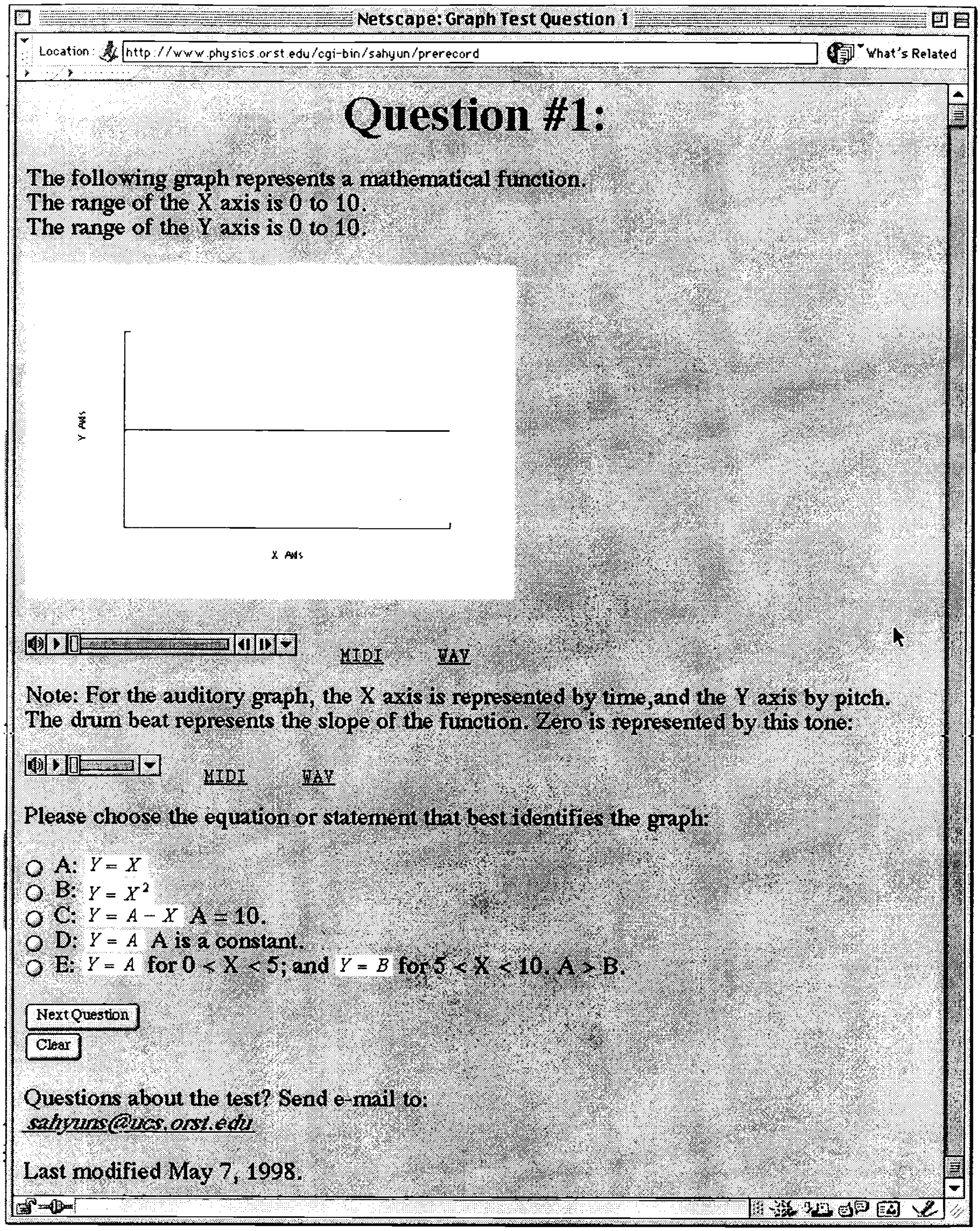
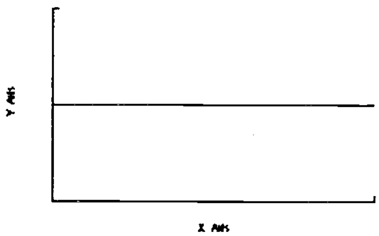


Figure C.5.1. Question Layout for Main Auditory Graph Test

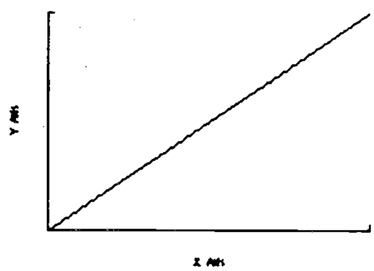
Question 1:

The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



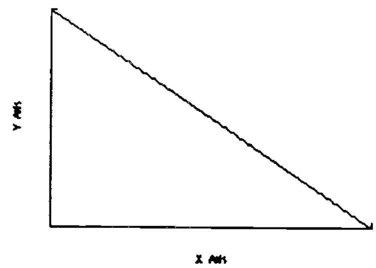
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = A - X$ A = 10.
D: $Y = A$ A is a constant.
E: $Y = A$ for $0 < X < 5$; and $Y = B$ for $5 < X < 10$. $A > B$.
Answer is: D

Question 2:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



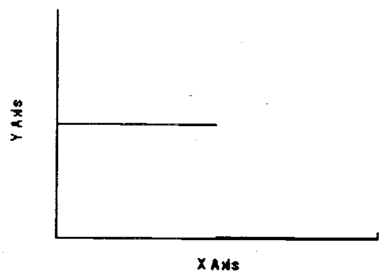
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = A - X$ A = 10.
D: $Y = A$ A is a constant.
E: $Y = A$ for $0 < X < 5$; and $Y = B$ for $5 < X < 10$. $A > B$.
Answer is: A

Question 3:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



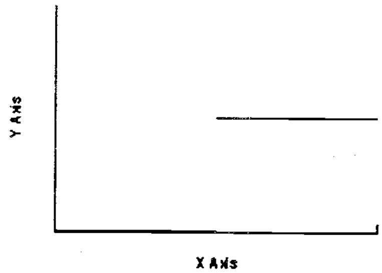
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = A - X$ A = 10.
D: $Y = A$ A is a constant.
E: $Y = A$ for $0 < X < 5$; and $Y = B$ for $5 < X < 10$. A > B.
Answer is: C

Question 4:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



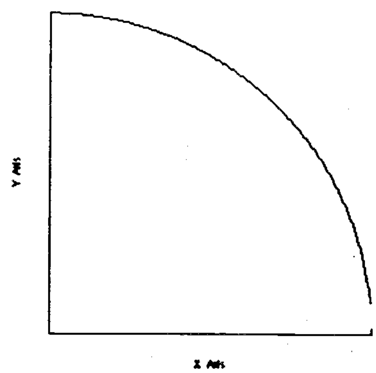
Please choose the equation or statement that best identifies the graph:
A: Y = X.
B: Y = A. A is a constant.
C: Y = A - X. A = 10.
D: Y = A for $0 < X < 5$; and Y = B for $5 < X < 10$. A > B.
E: Y = A for $0 < X < 5$; and Y = B for $5 < X < 10$. A < B.
Answer is: D

Question 5:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



Please choose the equation or statement that best identifies the graph:
A: Y = X.
B: Y = A. A is a constant.
C: Y = A - X. A = 10.
D: Y = A for 0 < X < 5; and Y = B for 5 < X < 10. A > B.
E: Y = A for 0 < X < 5; and Y = B for 5 < X < 10. A < B.
Answer is: E

Question 6:
The following graph represents a mathematical function.
The range of the X axis is 0 to 1.
The range of the Y axis is 0 to 1.



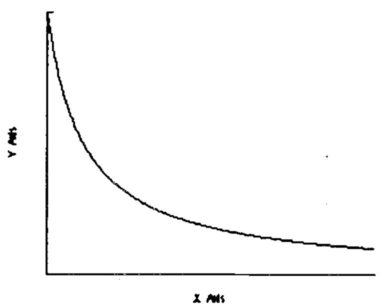
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = -X^2$
D: $Y = \sqrt{1 - X^2}$
E: $Y = \sqrt{X}$
Answer is: D

Question 7:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.

The range of the Y axis is 0 to 100.

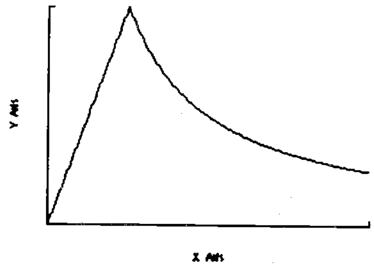
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = -X^2$
D: $Y = \sqrt{1 - X^2}$
E: $Y = \sqrt{X}$
Answer is: B

Question 8:
The following graph represents a mathematical function.
The range of the X axis is 1 to 10.
The range of the Y axis is 0 to 1.



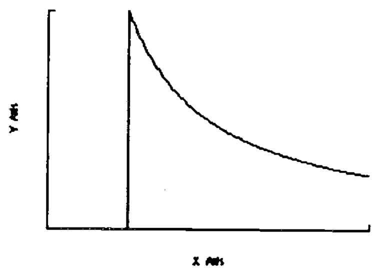
Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = \frac{1}{X}$
D: $Y = \sqrt{1 - X^2}$
E: $Y = \sqrt{X}$
Answer is: C

Question 9:
The following graph represents a mathematical function.
The range of the X axis is 0 to 4.
The range of the Y axis is 0 to 1.



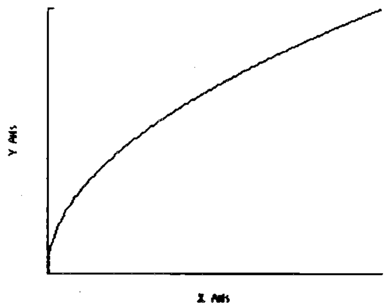
Please choose the equation or statement that best identifies the graph:
A: Y = X.
B: Y = 0 for 0 < X < 1; and Y = X for 1 < X < 4.
C: Y = X for 0 < X < 1; and Y = 1 - X for 1 < X < 4.
D: Y = 0 for 0 < X < 1; and Y = 1/X for 1 < X < 4.
E: Y = X for 0 < X < 1; and Y = 1/X for 1 < X < 4.
Answer is: E

Question 10:
The following graph represents a mathematical function.
The range of the X axis is 0 to 4.
The range of the Y axis is 0 to 1.



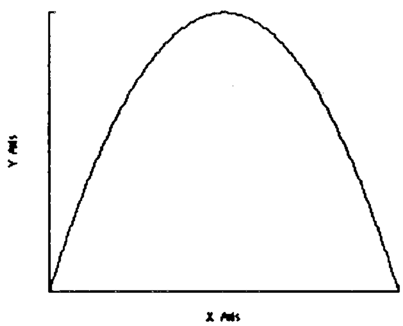
Please choose the equation or statement that best identifies the graph:
A: Y = X.
B: Y = 0 for 0 < X < 1; and Y = X for 1 < X < 4.
C: Y = X for 0 < X < 1; and Y = 1 - X for 1 < X < 4.
D: Y = 0 for 0 < X < 1; and Y = 1/X for 1 < X < 4.
E: Y = X for 0 < X < 1; and Y = 1/X for 1 < X < 4.
Answer is: D

Question 11:
The following graph represents a mathematical function.
The range of the X axis is 0 to 100.
The range of the Y axis is 0 to 10.



Please choose the equation or statement that best identifies the graph:
A: $Y = X$
B: $Y = X^2$
C: $Y = {}^{1}\!/_{X}$
D: $Y = \sqrt{1 - X^2}$
E: $Y = \sqrt{X}$
Answer is: E

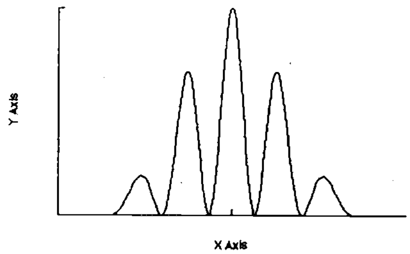
Question 12:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 125.



Please choose the equation or statement that best identifies the graph:
A: $Y = AX - \frac{1}{2}BX^2$
B: $Y = e^{-(5-X)^2}$
C: $Y = \sin^2(X)$
D: $Y = \sqrt{1 - X^2}$
E: $Y = X + \sin(X)$
Answer is: A

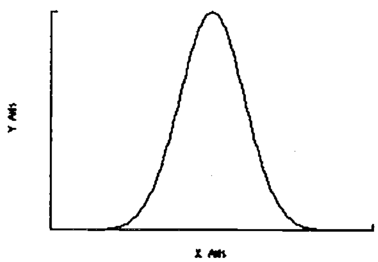
Question 13:

The following graph represents a function with one or more maxima.



Which statement best describes this graph?
A: This graph contains 5 peaks of equal magnitudes.
B: This graph contains 5 peaks of unequal magnitudes.
C: This graph contains 5 peaks of random magnitudes.
D: This graph contains 10 peaks of random magnitudes.
E: This graph contains only 1 peak.
Answer is: B

Question 14:
The following graph represents a mathematical function.
The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 1.



Please choose the equation or statement that best identifies the graph:

A: $Y = AX - \frac{1}{2}BX^2$

B: $Y = e^{-(5-X)^2}$

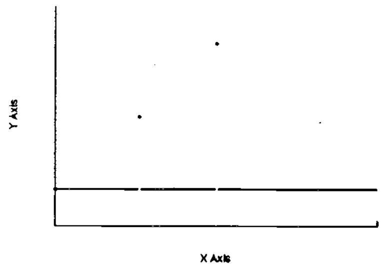
C: $Y = \sin^2(X)$

D: $Y = \sqrt{1-X^2}$

E: $Y = X + \sin(X)$

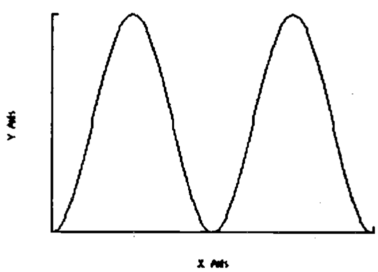
Answer is: B

Question 15:
The following graph represents a function with one or more discontinuities.



Which statement best describes this graph?
A: This graph contains 2 points of discontinuity whose function values are of unequal magnitudes.
B: This graph contains 2 points of discontinuity whose function values are of equal magnitudes.
C: This graph contains 2 points of discontinuity at which the function is undefined.
D: This graph contains 4 points of discontinuity whose function values are of unequal magnitudes.
E: This graph contains 4 points of discontinuity whose function values are of equal magnitudes.
Answer is: A

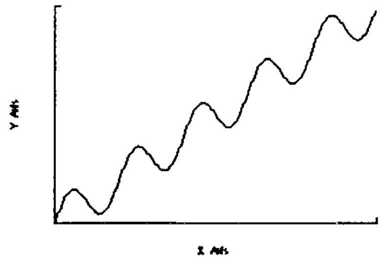
Question 16:
The following graph represents a mathematical function.
The range of the X axis is 0 to 2 Pi (= 6.3).
The range of the Y axis is 0 to 1.



Please choose the equation or statement that best identifies the graph:
A: $Y = AX - \frac{1}{2}BX^2$
B: $Y = e^{-(5-X)^2}$
C: $Y = \sin^2(X)$
D: $Y = \sqrt{1 - X^2}$
E: $Y = X + \sin(X)$
Answer is: C

Question 17:
The following graph represents a mathematical function.

The range of the X axis is 0 to 10.
The range of the Y axis is 0 to 10.



Please choose the equation or statement that best identifies the graph:

A: $Y = AX - \frac{1}{2}BX^2$

B: $Y = e^{-(5-X)^2}$

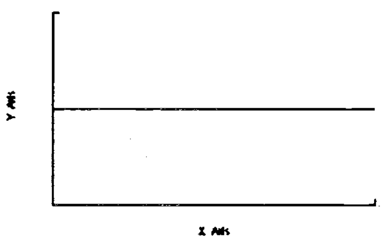
C: $Y = \sin^2(X)$

D: $Y = \sqrt{1 - X^2}$

E: $Y = X + \sin(X)$

Answer is: E

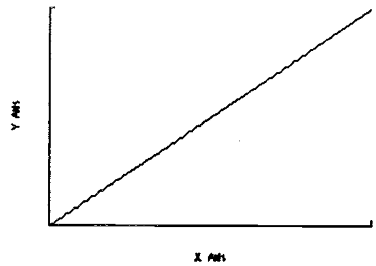
Question 18:
This is a graph of the motion of an object.
The X axis represents time, and has a range of 0 to 10 seconds.
The Y axis represents the object's distance from a reference point, and has a range of 0 to10 m.



Choose the sentence that is the best interpretation of the graph:
A: The object is moving with a constant non-zero linear acceleration.
B: The object does not move.
C: The object is moving with an acceleration whose magnitude is decreasing.
D: The object is moving with a constant non-zero linear velocity.
E: The object is moving with an acceleration whose magnitude is increasing.
Answer is: B

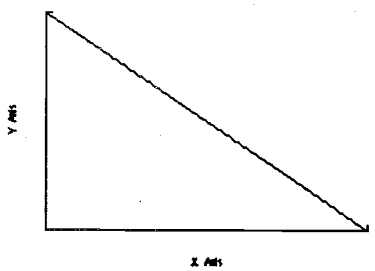
Question 19:
This is a graph of the motion of an object.
The X axis represents time, and has a range of 0 to 10 seconds.

The Y axis represents the object's velocity, and has a range of 0 to 10 m/s.



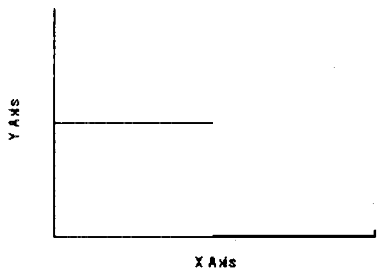
Choose the sentence that is the best interpretation of the graph:
A: The object is moving with a constant, non-zero acceleration.
B: The object is moving with an acceleration whose magnitude is decreasing.
C: The object is moving with an acceleration whose magnitude is increasing.
D: The object is moving with a constant velocity.
E: The object is moving with a decreasing velocity.
Answer is: A

Question 20:
This is a graph of the motion of an object.
The X axis represents time, and has a range of 0 to 10 seconds.
The Y axis represents the object's velocity, and has a range of 0 to 10 m/s.



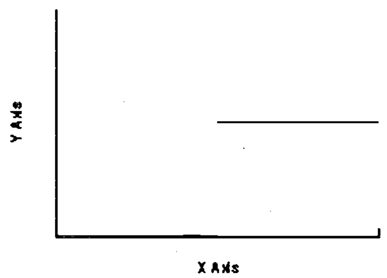
Choose the sentence that is the best interpretation of the graph:
A: The object is moving with a constant, non-zero acceleration.
B: The object is moving with an acceleration whose magnitude is decreasing.
C: The object is moving with an acceleration whose magnitude is increasing.
D: The object is moving with a constant velocity.
E: The object is moving with an increasing velocity.
Answer is: A

Question 21:
A force is applied to a 1 Kg object according to the graph below. The object moves without friction.
The X axis represents time; the range of the X axis is 0 to 2 seconds.
The Y axis represents the applied force; the range of the Y axis is 0 to 2 Newtons.



Choose the best completion to the following statement:
At 2 seconds, the object
A: is moving with an increasing velocity.
B: is moving with a constant velocity.
C: is moving with a decreasing velocity.
D: is not moving.
E: is moving backwards.
Answer is: B

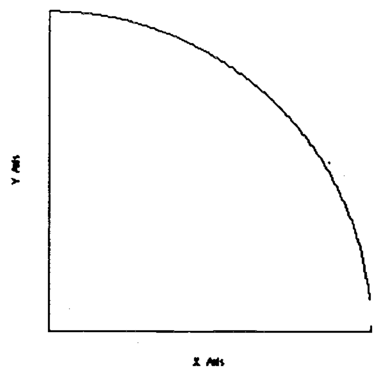
Question 22:
The following graph represents the current in a simple circuit consisting of a wire connecting a battery, a switch, and a resistor.
The X axis represents time; the range of the X axis is 0 to 2 seconds.
The Y axis represents the current in the wire; the range of the Y axis is 0 to 2 Amps.



Choose the best statement about the graph:
This graph shows
A: the current steadily increasing from 0 to 2 Amps.
B: the current steadily decreasing from 2 to 0 Amps.
C: that the current suddenly changes from 0 to 2 Amps.
D: that the current suddenly changes from 2 to 0 Amps.
E: that the current does not change.
Answer is: C

Question 23:
The following graph represents the position trajectory of an object traveling with constant speed. The object is moving from a minimum X value to a maximum X value.
The range of the X axis is 0 to 1 meter.

The range of the Y axis is 0 to 1 meter.



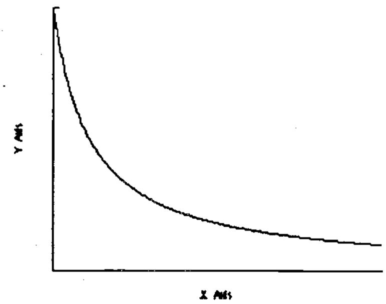
Which statement best describes the acceleration of the object?
A: The acceleration of the object is 0.
B: The acceleration of the object is parallel to the direction of travel.
C: The acceleration of the object is perpendicular to the direction of travel.
Answer is: C

Question 24:
The following graph is for the elastic potential energy of a material.
The X axis represents the distance that the spring is compressed.
The Y axis represents the potential energy of the spring.

Choose the sentence that is the best interpretation of the graph:
A: The elastic potential energy increases as the square of the distance that the material is compressed.
B: The elastic potential energy is increasing linearly with the distance the material is compressed.
C: The elastic potential energy is a non-zero constant as the material is compressed.
D: The elastic potential energy is decreasing.
E: The elastic potential energy of the material is 0.
Answer is: A

Question 25:
The following graph represents an ideal gas in a container.
The X axis represents volume and has a range of 1 to 10 L^3.

The Y axis represents pressure and has a range of 0 to 1 Atm.



Please choose the statement that best describes the relationship between pressure and volume:
A: The pressure is constant as the volume changes.
B: The pressure is not related to the volume.
C: The pressure increases linearly as the volume increases.
D: The pressure decreases linearly as the volume increases.
E: The pressure decreases as 1/volume.
Answer is: E

Question 26:
The following graph refers to the gravitational force produced by a sphere of mass M, acting on a test object of mass m.
The X axis represents distance, and has a range of 0 to 4 R (R = radius of sphere.)
The Y axis represents the gravitational force, and has a range of 0 to $1\frac{GMm}{R^2}$.

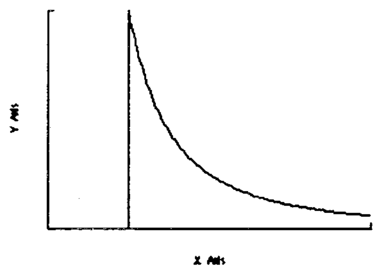
Choose the sentence that is the best interpretation of the graph.
A: The gravitational force first increases as $(\text{distance})^2$ until distance = R, then decreases linearly.
B: The gravitational force first increases as $(\text{distance})^2$ until distance = R, then is proportional to $\frac{1}{(\text{distance})^2}$.
C: The gravitational force first increases linearly until distance = R, then decreases linearly.

D: The gravitational force first increases linearly until distance = R, then is proportional to $\frac{1}{(\text{distance})^2}$.

E: The gravitational force is not related to the distance.
Answer is: D

Question 27:
The following graph refers to the electric field produced by a charged sphere.
The X axis represents distance, and has a range of 0 to 4 R (R = radius of the spherical shell.)

The Y axis represents the electric field, and has a range of 0 to $1\frac{Q}{4\pi R^2\varepsilon_0}$.



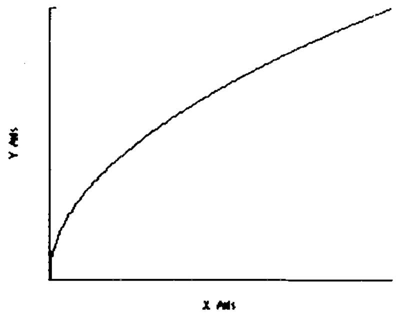
Choose the sentence that is the best interpretation of the graph.
A: The electric field is 0 until distance = R, then decreases linearly from a maximum value.

B: The electric field is 0 until distance = R, then is proportional to $\frac{1}{(\text{distance})^2}$.

C: The electric field increases linearly until distance = R, then decreases linearly.
D: The electric field increases linearly until distance = R, then is proportional to $\frac{1}{(\text{distance})^2}$.

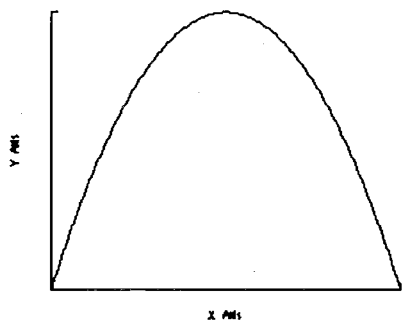
E: The electric field is not related to the distance.
Answer is: B

Question 28:
When the ocean depth is greater than three wavelengths, the speed of waves in the ocean is approximated by the following graph.
The X axis represents the wavelength, and has a range of 0 to 1000 m.
The Y axis represents the velocity of ocean waves, and has a range of 0 to 40 m/s.



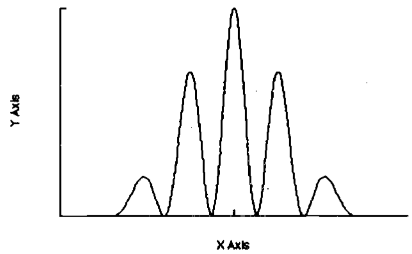
Choose the sentence that is the best interpretation of the graph.
A: The speed of ocean waves is proportional to the wavelength.
B: The speed of ocean waves decreases with increasing wavelength.
C: The speed of ocean waves is proportional to the square root of the wavelength.
D: The speed of ocean waves is proportional to $(\text{wavelength})^2$.
E: The speed is not dependent on the wavelength.
Answer is: C

Question 29:
The following graph shows the trajectory of a projectile with a constant velocity component in the X direction.
The X axis represents distance, and has a range of 0 to 10 m.
The Y axis represents height, and has a range of 0 to 125 m.



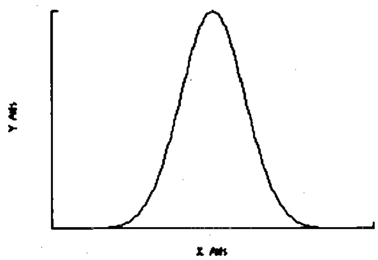
Choose the sentence that is the best interpretation of the graph.
A: At the graph's maximum Y value, the magnitude of the projectile's velocity component in the Y direction is 0.
B: At the graph's maximum Y value, the magnitude of the projectile's velocity component in the Y direction is greater than 0.
C: Just before the graph's maximum X value, the magnitude of the projectile's velocity component in the Y direction is 0.
D: At the graph's maximum Y value, the projectile's acceleration is 0.
E: At the graph's maximum Y value, the projectile's distance is 0.
Answer is: A

Question 30:
The following graph shows the pattern of light intensity projected onto a screen from a monochromatic light source.
The X axis represents distance on the screen, and has a range of -1 to 1 mm.
The Y axis represents relative light intensity, and has a range of 0 to 1.



This pattern represents light that:
A: has passed through a single slit aperture.
B: has passed through a double slit aperture.
C: has passed through a diffraction grating.
D: is produced by a light beam with a single central maximum intensity.
E: displays the effects of edge diffraction from a semi-infinite screen.
Answer is: B

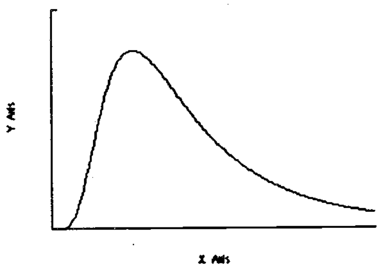
Question 31:
The following graph shows the pattern of light intensity projected onto a screen from a monochromatic light source.
The X axis represents distance on the screen, and has a range of -1 to 1 mm.
The Y axis represents relative light intensity, and has a range of 0 to 1.



This pattern represents light that:
A: has passed through a single slit aperture.
B: has passed through a double slit aperture.
C: has passed through a diffraction grating.
D: is produced by a light beam with a single central maximum intensity.
E: displays the effects of edge diffraction from a semi-infinite screen.
Answer is: D

Question 32:

The following graph shows the light intensity produced by a black body object, such as the Sun, with a temperature at 5000 K.
The X axis represents the wavelength of light, and has a range of 0 to 2000 nm.
The Y axis represents relative light intensity, and has a range of 0 to 1.

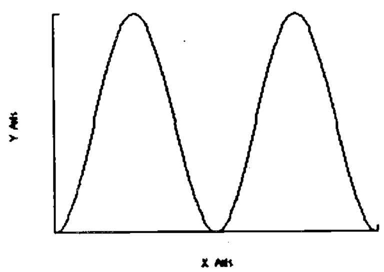


Please choose the sentence that best describes the graph.
A: There is a constant distribution of light intensity vs. wavelength
B: The maximum intensity occurs at approximately 500 nm.
C: The maximum intensity occurs at approximately 1000 nm.
D: The maximum intensity occurs at approximately 1500 nm.
E: The intensity is increasing throughout this range.
Answer is: B

Question 33:
In an AC (alternating current) circuit, the instantaneous electric power dissipated by a resistor is given in the following graph.

The X axis represents time, and has a range of 0 to $\frac{2\pi}{\omega}$ Seconds. Where w is the frequency of the AC.

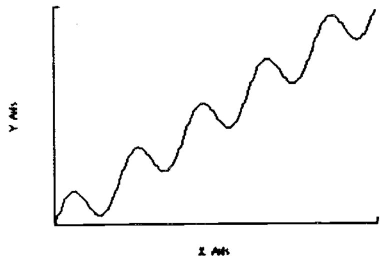
The Y axis represents power dissipated, and has a range of 0 to $1\frac{\mathcal{E}_{max}^{2}}{R}$. E_{max} is the maximum EMF Voltage amplitude.



Please choose the sentence that best describes the graph.
A: The instantaneous power dissipated is a non-zero constant in time.
B: The instantaneous power dissipated is always decreasing with time.
C: The instantaneous power dissipated is always increasing with time.
D: The instantaneous power dissipated is 0 at specific points in time.
E: The instantaneous power dissipated is always 0.

Answer is: D

Question 34:
A test mass is suspended by springs on a cart. Assume that the mass of the cart is much greater than that of the test mass. The following graph describes the motion of the test mass in the Y direction.
The X axis represents time, and has a range of 0 to 10 Seconds.
The Y axis represents distance that the mass has traveled, and has a range of 0 to 1 m.



Please choose the sentence that best describes the conditions to produce the motion portrayed in the graph. All velocities are in the lab rest frame.
A: Only the cart was given an initial velocity in the Y direction.
B: Only the test mass was given an initial velocity in the Y direction.
C: The cart was given an initial velocity in the Y direction while the test mass was given an initial velocity in the - Y direction.
D: The cart and test mass are given the same initial velocity in the Y direction.
E: The test mass is given twice the initial velocity of the cart in the Y direction.
Answer is: E

C.6. Summary Table of Results from the Main Test

Table C.6.1 and Table C.6.2 are summaries of the results obtained from the Main Auditory Graph test. The tables are divided into six columns: All, Vision, Both, Sound, Blind, and Grad. The All category represents the average of the scores from the Vision, Both, and Sound groups. Vision represents the subjects given the test with visually presented graphs, Both represents the group given the test with both visual and auditory graphs, and Sound represents the subjects given the test with only auditory graphs. The Blind group had subjects who were blind and took the test. Only one of these five subjects answered the Pre-test questions. The Grad group represents 6 physics graduate students who took the test with only auditory graphs.

Table C.6.1 Summary of results for the Main Auditory Graph Test. Part 1.

Group	All (B, S, V)	Visual	Both	Sound	Grad	Blind
# Subjects	231	76	81	74	6	5
% Female	48%	41%	59%	43%	17%	20%
Avg Age	22	22	22	22	29	38
% Correct Answers:						
Pre-Test						
pt1	78%	72%	83%	78%	100%	20%*
pt2	96%	95%	96%	97%	83%	20%*
pt3	92%	93%	89%	93%	83%	20%*
pt4	79%	67%	81%	89%	100%	20%*
pt5	63%	55%	68%	66%	100%	20%*
Main Test						
mt1	68%	84%	64%	57%	100%	100%
mt2	71%	86%	79%	46%	83%	100%
mt3	80%	82%	80%	77%	100%	100%
mt4	77%	78%	79%	76%	100%	100%
mt5	77%	80%	83%	69%	100%	100%
mt6	42%	61%	42%	24%	67%	80%
mt7	78%	83%	81%	70%	100%	100%
mt8	74%	76%	78%	68%	100%	80%
mt9	60%	66%	68%	45%	100%	60%
mt10	65%	62%	75%	55%	83%	80%
mt11	51%	58%	58%	36%	100%	80%
mt12	30%	38%	38%	14%	67%	40%
mt13	61%	68%	65%	49%	83%	80%
mt14	22%	28%	17%	22%	83%	40%
mt15	63%	66%	63%	59%	83%	80%
mt16	47%	49%	46%	46%	100%	80%
mt17	28%	30%	26%	28%	100%	80%
mt18	49%	57%	51%	41%	83%	100%
mt19	37%	42%	35%	35%	33%	60%
mt20	23%	21%	28%	19%	83%	80%
mt21	36%	41%	41%	26%	83%	20%
mt22	81%	76%	81%	84%	100%	100%
mt23	55%	62%	44%	58%	50%	80%
mt24	71%	72%	79%	62%	100%	100%
mt25	60%	70%	62%	49%	100%	100%
mt26	52%	58%	62%	36%	100%	60%
mt27	59%	63%	64%	50%	100%	100%
mt28	47%	54%	58%	27%	100%	100%
mt29	44%	54%	43%	35%	83%	40%
mt30	49%	45%	48%	54%	17%	80%
mt31	38%	37%	36%	42%	83%	60%
mt32	68%	72%	70%	62%	100%	80%
mt33	69%	72%	70%	64%	100%	100%
mt34	19%	18%	16%	23%	17%	0%
Avg Time (min.)	29.37	23.56	30.72	33.86	39.74	2488.98

* = only 1 Blind subject answered the pre-test

Table C.6.2 Summary of results for the Main Auditory Graph Test. Part 2.

Avg. # Correct	All	Visual	Both	Sound	Grad	Blind
Pre-Test	4.08	3.83	4.17	4.24	4.67	1.00*
Std. Dev.	0.65	0.86	0.52	0.63	0.46	0.00*
Main Test	18.53	20.08	19.32	16.07	28.83	26.40
Std. Dev.	5.93	6.03	5.46	5.59	3.31	3.71
Math Section	9.95	10.93	10.43	8.41	15.50	13.80
Std. Dev.	3.40	3.39	2.98	3.35	1.76	1.92
Physics Section	8.58	9.14	8.89	7.66	13.33	12.60
Std. Dev.	3.18	3.41	3.15	2.78	1.75	2.07
r = Math/Physics correlation	0.31	0.27	0.44	0.17	0.31	0.27
Split Half Reliability	0.47	0.43	0.61	0.29	0.47	0.42
	Average	# Subjects				
Sound with Music Trianing	16.91	47				
Sound w/o Music Trianing	14.59	27				

* = only 1 Blind subject answered the pre-test

C.7. Histograms and Tests of Normal Distribution of Data

In Chapter 9, the data were assumed to follow a normal distribution so that *t*-tests could be applied. The following charts show that the data is a fair, but not perfect approximation to the ideal Normal population distribution. In this test, there seems to be a few more outlying data points than would normally be expected. Appendix B.7 has a discussion on the techniques used for the Normal Probability graphs.

Figure C.7.1 portrays a histogram of the distribution of scores on the Main test. The distribution closely, but not perfectly follows the Gaussian ideal. These results are greatly improved over those of the Web Pilot Test. 0 displays the Normal Probability plot of the distribution. As can be readily seen, the slope of the best fit (0.20) is not in perfect agreement with the $1/\sqrt{n} = 1/\sqrt{231} = 0.07$ slope of the predicted ideal. A normal population would have a 0.20 slope if the sample size were 25 rather than the 231 used in this study. Separating the subject groups showed similar results.

Thus, the data does not follow a perfect normal population distribution which may introduce some inconsistent results in the tests applied in Chapter 9. However, it should be noted that the data is not radically different from a normal population, so the general analysis should be satisfactory.

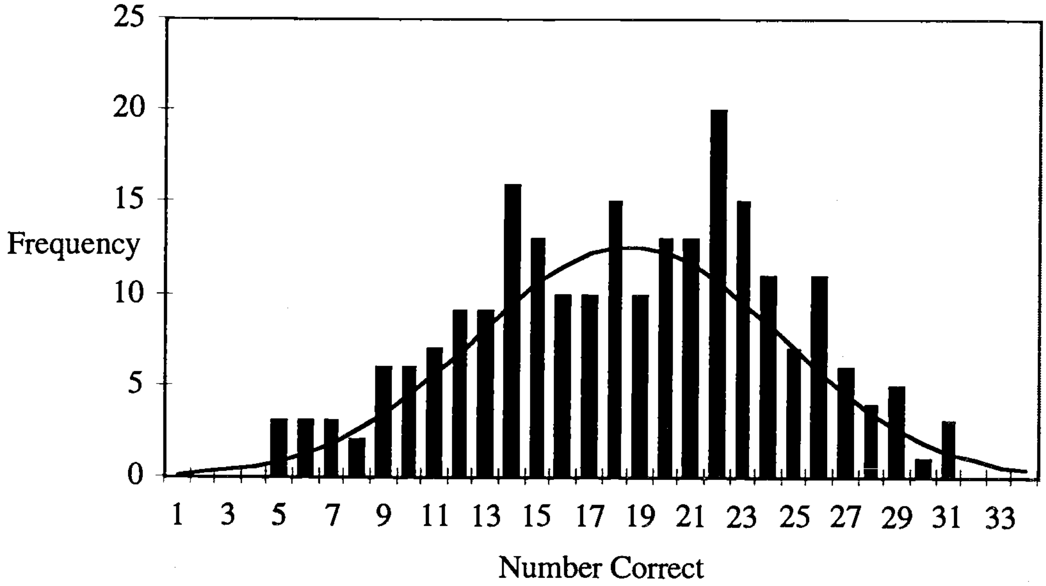


Figure C.7.1. Comparison of All subjects (S, B, and V groups) to the Gaussian Ideal.

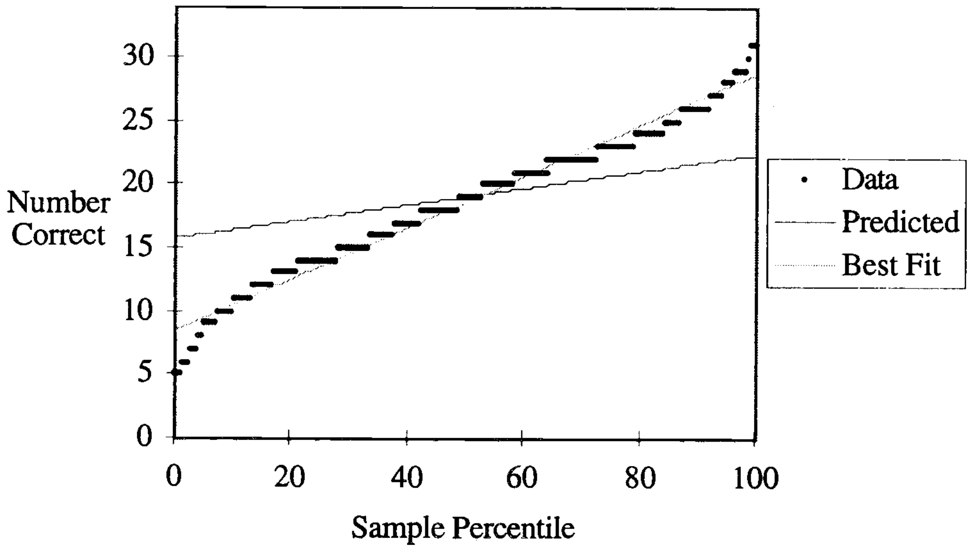


Figure C.7.2. Normal Probability distribution for All (S, V, and B groups).

Appendix D Material Relating to the Auditory Preference Pilot Test

D.1. Overview

This appendix contains material used in the Auditory Preference Pilot test. These materials include the Introductory page, Informed Consent form, and the question pages. There is also a summary of the results obtained from this test as well as a listing of the subject's text responses for each of the questions.

D.2. Introductory page and Informed Consent Form

Welcome

The object of this test is to compare preferences between different types of auditory representations of graphical information. Two types of graphs are utilized, pictorial graphs which almost everyone is familiar with, and sonified data graphs (data represented by sound), which is much less common.

In general, the Y axis of the pictured graph will be represented by pitch and the X axis by time. Additionally, some of the auditory graphs may contain characteristics relating to first and second derivative values as well as negative quantities.

As this is a test about preferences, many questions do not have a right or wrong answer, simply respond as to which choice seems to work best.

To listen to the graphs when they are presented, click on the button marked "Play" next to the answer choice. You can play a graph as many times as you like. Answers will be of a multiple-choice format, with a text area for additional comments.

Hardware and software requirements for this test are a sound capable computer, Windows95, and Internet Explorer 4.0 or greater. Some of the sounds used in this test utilize Microsoft's ActiveX controls. Due to the nature of these controls, some pages may take a minute or two to load. Please be patient.

Since some of the ActiveX controls are not registered, the View -> Internet Options -> Security setting will need to be set to Low. Neither Oregon State University, nor the investigators assume any responsibility for the actions of these controls. **Use at your own risk.**

Before proceeding, please take a moment to read the following document.

Statement of Informed Consent

Physics Department, Oregon State University

Title of Project: Comparison Between Auditory and Visual Graphing Methods.

Investigators:
 Steven Sahyun, Graduate Research Assistant, Physics Department.
 John Gardner, Professor, Physics Department.

This purpose of this study, as stated above, is to determine preferences between auditory graphing (data representation using sound) methods.

How the test will work:

Subjects will be given a log-in page that records their name. Names are not stored the same computer file as responses.

Subjects will then be given a series of 9 pages containing questions in multiple-choice and Likert scale preference formats. Also, some questions will ask for identification of displayed graphs.

Only data from subjects who answer all 9 question pages will be used.

It should be noted that all responses are being transferred on the Internet, and are not encrypted. However, reasonable attempts at confidentiality are made in that the subject's name will not appear with, or be stored with, any of their responses. Names will not be used in any publications or presentations of the data obtained.

It should also be noted that in cases where student subjects are taking this test for extra credit in a course, a list of which students have taken the test will be forwarded to the respective school's instructor so that those subjects may receive credit. Results of the study will not be used to determine credit, only the fact that the test has been taken. Also, a summary listing of the average responses to the test questions will be available to the instructor.

The test is estimated to take about 10 minutes.

Participation in this study is voluntary and you may either refuse to participate or withdraw from the study at any time, although full participation is greatly appreciated. You may take a break at any time, just be sure not to loose the web page question that you are on. If there should be a technical problem (crash) during the test, you will need to start over.

If you have any questions about the research study and/or specific procedures, please contact:
Steven Sahyun
Physics Department
Oregon State University
301 Weniger Hall
Corvallis OR, 97330
USA
The phone number is (541) 737-1712. Any other questions should be directed to Mary Nunn, Sponsored Programs Officer, OSU Research Office, Oregon State University, Corvallis, OR, 97331. The phone number is: (541) 737-0670.

After reviewing the above statement, please click on the link below:

I agree to this test.

Questions about this test? Send me e-mail:
sahyuns@ucs.orst.edu

Last modified November 10, 1998.

D.3. Test Questions

The questions used in the Auditory Preference Pilot Test were displayed as a series of 9 Web pages. Each question displayed a series of choices with radio style button selectors, as well as a text entry field so that they could record comments relating to their choice selection. Figure D.3.1 demonstrates a typical question. The text of the different questions, as well as an explanation of which sound graphs were being compared follows the figure.

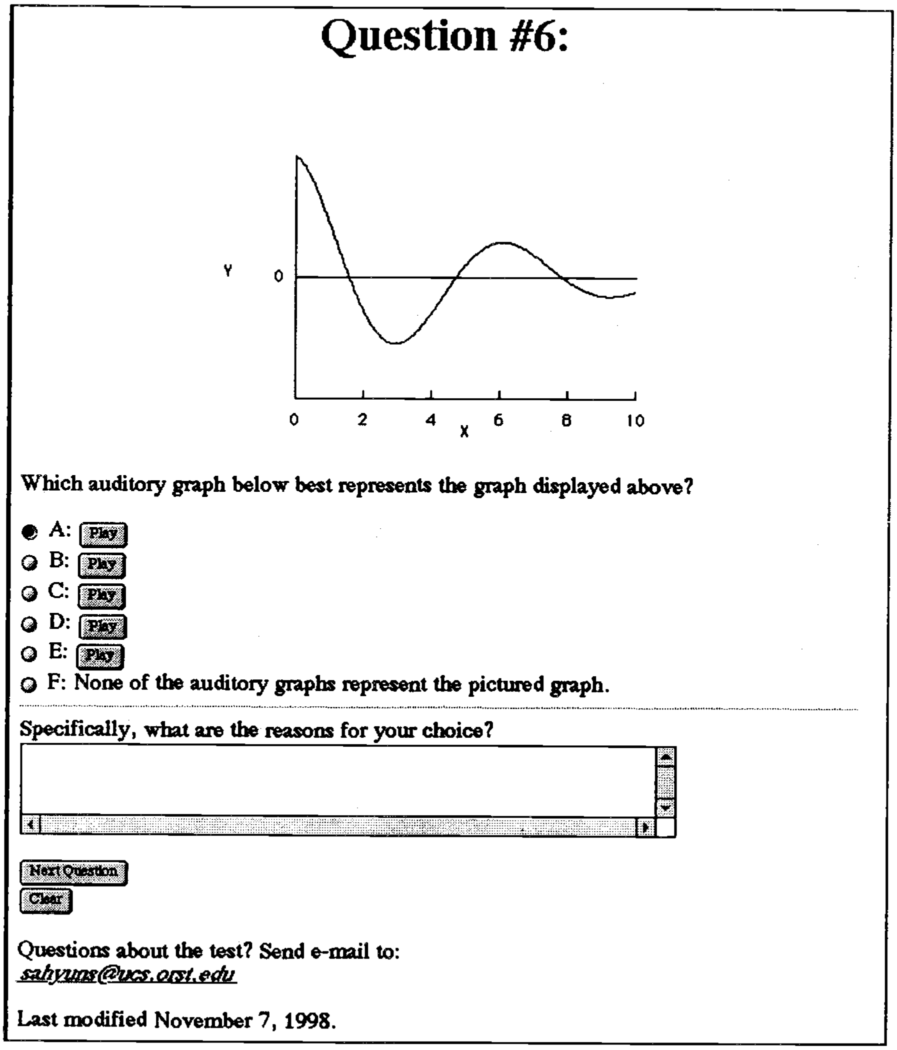
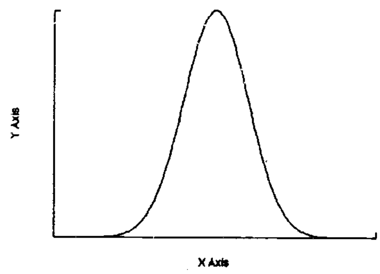


Figure D.3.1. Typical Question Layout for the Auditory Preference Pilot Test

Question #1:



Which of the auditory graphs below, if any, do you prefer as best matching the graph displayed above?

In these auditory graphs, the y axis is represented by pitch and the x axis with time.

⊗A: Play
⊗B: Play
⊗C: Both represent the graph equally well.
⊗D: Neither represents the graph well.

Specifically, what are the reasons for your choice?

Next Question
Clear

Comment on Question 1:

This question compared two Auditory representations of a Gaussian curve using only pitch to represent the curve. Choice A used an ActiveX control producing a smooth, continuously varying tone, while Choice B used the MIDI piano instrument with notes representing data points. This question was intended to check how many people preferred the smooth tone versus the more staccato sounding MIDI.

Question 2 used the same graph as in question 1, but the text read as:

In the auditory graphs below, the y axis is represented by the pitch of the piano tone, and the x axis is represented by time. Additionally, the first derivative (slope) is represented by the frequency of a drum beat. The second derivative (curvature) is represented by the pitch of the drum beat.

Which auditory graph do you prefer as best matching the graph displayed above?
⊗A: Play

⊗B: Play
⊗C: Both represent the graph equally well.
⊗D: Neither represents the graph well.

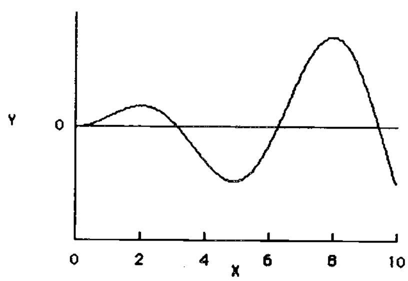
Specifically, what are the reasons for your choice?

Next Question
Clear

Comment on question 2:

This question compared two MIDI graphs, where the difference was the pitch of a tick mark (drum-beat) representing the first (slope) and second (curvature) derivatives. Choice A used the mapping of a low tone for the tick mark pitch when the curvature was positive, and a high tone when it was negative. Choice B used the reverse mapping in that the positive curvature had a high pitch tone and negative had a low pitch.

Question #3:



In the auditory graphs below, the y axis is represented as pitch and the x axis as time.

Which auditory graph do you prefer as best matching the graph displayed above?

⊗A: Play
⊗B: Play
⊗C: Both represent the graph equally well.
⊗D: Neither represents the graph well.

Specifically, what are the reasons for your choice?

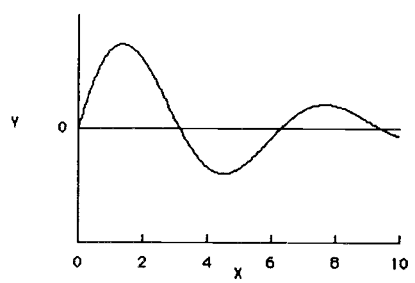
Next Question
Clear

Comment on Question 3:

This question used two MIDI auditory graphs to determine it there was a preference for indicating when the graph's y axis values were negative. Choice B had an instrument change from piano to harpsichord for negative values, while Choice A had no change.

Question 4 had the same graph and similar wording to question 3, but compared subjects' preference for the ActiveX smooth tone and tick mark beats having high pitch for positive values and low pitch for negative values (Choice A) to a MIDI graph with drum-beats (high pitch for negative values, low pitch for positive ones) and an instrument change for negative *y* axis values (Choice B). A change in the tone quality representing negative *y* axis values was not used with the ActiveX graph as this feature was not available at the time of testing.

Question #5:



Which auditory graph below best represents the graph displayed above?

⊗A: Play
⊗B: Play
⊗C: Play
⊗D: Play
⊗E: Play
⊗F: None of the choices represent the displayed picture.

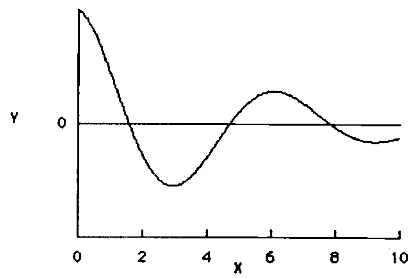
Specifically, what are the reasons for your choice?

Next Question
Clear

Comments on question 5:

This question was intended to determine how effectively subjects could match a visually presented graph to an ActiveX graph with the tick mark beats having high pitch for positive values and low pitch for negative values. The five graph choices were representations of (A) sin x, (B) cos x, (C) x sin x, (D) e^{-x} sin x, and (E) e^{-x} cos x.

Question #6:



Which auditory graph below best represents the graph displayed above?

⊗A: Play
⊗B: Play
⊗C: Play
⊗D: Play
⊗E: Play
⊗F: None of the auditory graphs represent the pictured graph.

Specifically, what are the reasons for your choice?

Next Question
Clear

Comments on question 6:

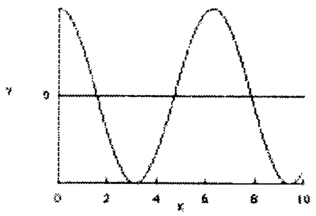
This question was similar to question 5, except that it used MIDI auditory graphs and the pitch of the drum was high for negative curvature and low for positive curvature. The five graph choices were representations of (A) sin x , (B) cos x, (C) x sin x, (D) e^{-x} sin x, and (E) e^{-x} cos x. The point of questions 5 and 6 was to compare the two auditory methods for subjects' ability to correctly match a visual graph to one of several auditory graphs.

Question #7:

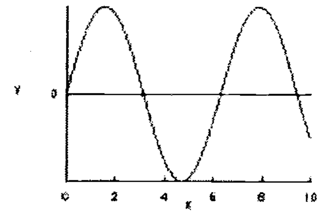
Play

Which graph pictured below best represents the auditory graph above?

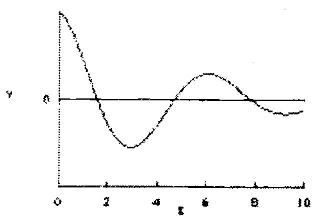
⊗A:



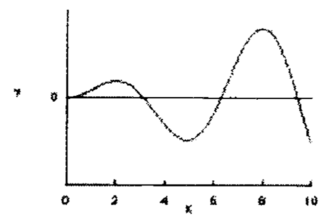
⊗B:



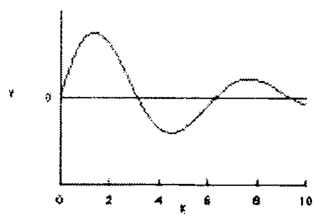
⊗C:



⊗D:



⊗E:



⊗F: None of the pictured choices represent the sound.

Specifically, what are the reasons for your choice?

Next Question
Clear

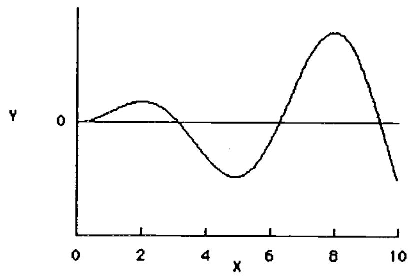
Comments on question 7:

Question 7 was similar to those of 5 and 6, however, in this question, the subject was asked to match an ActiveX auditory plot of cos x with derivative tick-mark beats (high pitch for positive values and low pitch for negative values) to one of the five visually presented graphs.

Question #8:

Question 8 was presented identically to question 7. The auditory graph for this question was a MIDI representation of sin x with derivative (high for negative curvature and low for positive curvature) and an instrument change for negative values. The point of questions 7 and 8 was to compare the two auditory methods for subjects' ability to correctly match an auditory graph to one of several visual graphs.

Question #9:



The following auditory graphs represent the graph pictured above. For each of the auditory graphs below, please give a rating on how well you feel that the sound graph represents the graph pictured above.
The scale is 1 to 5, where 1 is bad, 2 is poor, 3 is neutral, 4 is good, and 5 great.

A: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play
B: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play
C: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play
D: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play
E: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play
F: ⊗1 ⊗2 ⊗3 ⊗4 ⊗5 Play

Specifically, what are the reasons for your choices?

End
Clear

Comments on question 9:

This last question was designed to find subjects' overall preference for the different auditory graphing methods that were used. This question allowed subjects to indicate which method was their overall preference and how well they felt that the auditory graphs acted as a display method.

D.4. Summary of the Results from the Auditory Preference Pilot Test

Table D.4.1 lists the results from the test questions. Tables D.4..2-10 list subject text responses.

Table D.4.1 . Summary of Results from the Auditory Prefernce Pilot Test.

Total #	12					
	A	B	C	D	E	F
1. Gaussian curve, A = ActiveX, B=MIDI, C=both, D=neither	33%	58%	8%	0%		
2. Gaussian curve, drum beat for deriv, pitch for second deriv, A=low +, high -; B = high +, low -, C = both good, D=neither good	33%	17%	33%	17%		
3. x sin$x$. A=no change at 0, B= instrument change at 0, C both good, D neither good	50%	42%	0%	8%		
4. x sin$x$. A=ActiveX with deriv., B=MIDI with deriv and pitch change a 0; C = both, D=neither	33%	50%	0%	17%		
5. Match e^{-x} sin x, graph (d), ActiveX sounds. a= sin x, b= cos x,, c= x sin x, d= e^{-x} sin x,, e= e^{-x} cos x, f=none	17%	0%	17%	58%	0%	8%
6. Match e^{-x} cos x, graph (e), MIDI sounds. a=sin x, b=cos x, c=x sin x, d= e^{-x} sin x,, e= e^{-x} cos x,, f=none	8%	0%	0%	17%	58%	17%
7. Match ActiveX sound of cox(x) to picture graph (a). a=cos x, b=sin x, c=x sin x, d= e^{-x} cos x,, e= e^{-x} sin x, f=none	75%	0%	0%	17%	8%	0%
8. Match MIDI sound of sin(x) to picture graph (b). a=cos x, b=sin x, c=xsin(x), d= e^{-x} cos x,, e= e^{-x} sin x,, f=none	0%	92%	0%	0%	0%	8%

9. Likert preference of xsin(x) graph with different sound represntations 1- 5, 1 is bad, 2 is poor, 3 is neutral, 4 is good, and 5 great.

	Average	Std. Dev.
a. MIDI	3.75	0.87
b. ActiveX	3.75	1.14
c. MIDI, dx	3.08	0.79
d.ActiveX, dx	3.42	1.08
e. MIDI, 0	3.83	1.34
f. MIDI, dx, 0	3.33	1.44
Average a-f:	3.53	1.11

Table D.4.2 . Text responses for Question 1.

Subject	Comment
35	I was more familiar with sound A as a representative of motion going up and down.
36	It was harder for me to picture the second part of the curve with A. B just seemed clearer.
40	I seem to make the connection better for the higher pitches
41	The greater difference between the maximum and minimum tones made the graph easier to visualize
42	Sound A was a continuous sound, and was difficult for me to hear the steepness of the graph. Sound B seemed to help in this instance, but still needed more discontinued sounds. Also, with sound A, it was difficult to distinquish when it was at the top of the graph. I heard it better in B.
44	A sounded more continuous
46	B is like to going up the hill and seems that a car has going up to a hill and gives more gas.
47	Selection A was less dramatic than B.
48	It sounds more gradual and note quite as steep as the other
49	Both the sounds seemed rise and fall about how the visual graph. The only real difference was that they had different time scales.

Table D.4.3 . Text responses for Question 2.

Subject	Answer	Comment
35	C	I had a hard time listening and interpreting the drum beats when they were at a low pitch. I wasn't really sure of the frequency, however, I chose C because the piano and high pitch drum beat were representative of the graph for both A and B.
40	A	b is very confusing
41	C	neither changed my perception of the graph much
42	D	I actually had a hard time first distinquishing the difference between the two sounds, then I did not think either one represented the graph. The pitch of the piano tone did not seem to be as drastic as the graph previously.
44	D	The beats were confusing, if they could represent time intervals, they may be more useful.
46	C	However, they need a little bit dramatical sound when they reached at the top of the graph, I think.
47	A	The high pitch at the top of the curve best described the peak of the curve.
48	B	I could hear the first/second derivatives better in this one
49	B	The freqency of the drum seemed to match what I would expect for the dirivative.

Table D.4.4 . Text responses for Question 3

Subject	Answer	Comment
35	B	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.
36	B	I liked hearing the differece between positive and negative.
40	A	b involves too many options for the ear to play with. I did not feel that I needed a different sound for negative values.
41	A	The change from something sounding like a piano to something like an harpsichord broke up the graph a little too much
42	A	I wasn;t sure, but sound B seemed like it went down in volume as the graph decreased. OTher than that, they sounded the same.
44	D	Although A represented it more closely, the intervals for the maxima didn't seem to agree with the sound
46	B	B expressed better on the ups and downs. The sound of downs is easily recognized, but that of ups is little bit.
47	B	The change in sound when the curve went into negative values was very helpful.
48	B	Definitely because the change as you cross into negative, very helpful
49	A	Although both seem to discribe the graph correctly, I didn't like to sound when the function went negative.

Table D.4.5 . Text responses for Question 4.

Subject	Answer	Comment
35	B	The distinct sounds in B were much more clear than in A. In fact I think I just started to learn how to interpret the drum beats.
36	B	With A I couldn't hear a lot of the drum. It may be something I could learn to listen for with enough use but one first listen, I thought B was more clear.
40	A	a would be better if the drum pitch had those high harmonics for positive values instead of the negative ones.
41	D	too much information at once
42	B	In sound B, the slopes were more apparent, however, I think there needs to be more drum beats (lacking in A) Also, doesn't the time and change in pitch indicate the slope of the curve? The drum beats really don;t seem to help me.
44	A	dipicted amplitude and interval the best
46	A	A sound is good to me. I recognized that piano and drum beat are little bit hard to make a picture in my mind. Sharp pitch is better to me, but this one also needs some different sound to express the ups and downs.
47	B	Te combination of the negative value change in sound and the high pitch at the peak of the curve.
48	B	this agin, for negative change and you can pick up the slope/curvature better
49	D	Both seemed rather arbitrary in relation to the graph, at least in the dirivative department.

Table D.4.6. Text responses for Question 5.

Subject	Answer	Comment
35	F	I started to choose E or D, but really I didn't like any of the choices. I thought that the pitch did not go high enough at the beginning of the graph or was not low enought at the end of the graph. Also, the long drawn out sound at the end created a sen
36	D	With D, I found it easier to distinguish the X ad Y (the legnth vs. the height of the curve). B and E I would have had reversed around the y axis. A I would not have pictured both curves about the same size.
41	D	sounded the closest
42	C	B and E sounded the opposite of what the graph illustrated (if a higher pitch is up, lower down) I just like the sound of C the best, however, listening to the pitches, it almost seems like the two maximums reach the same pitch, but on the graph, the second one is lower.
46	D	D sound represent the distance as time goes.
47	D	The change in sound as the curve went into the negative region of the y- axis.
48	D	But frankly, a-d sounded all the same
49	C	I think that A,C, and D seemed to do the graph justice but I think that C was the most clear

Table D.4.7 . Text responses for Question 6.

Subject	Answer	Comment
35	E	E was the only one I could trace mentally on the graph.
36	E	B, C, and D sounded upside down. E was the clearest. A had too much to listen to.
41	D	sounded closest although the drums in the background created confusion as to what was going on there
42	A	B,c,d sounded opposite to the graph and A and E sounded nearly the same.
46	E	Among them, E is the best, but I think if not considering emotion of people, drum and piano sound are hardly recognized by distance and ups and downs.
47	D	The tempo of the drum was most clear in describing the slope of the line, as was the the change in sound describing the negative values ofthe curve.
48	E	sounded not as choppy, not so harsh on the change over to negative
49	F	E seemed the closest, but the dirrivative portion seemed wrong.

Table D.4.8 . Text Responses for Question 7.

Subject	Answer	Comment
35	A	It whining sound was well represented both up and down. The drum beats added something that I could also verify my choice with.
36	A	It sounded like it started high, went down and raised just as high before going just as low down ad then having a slight upturn at the end.
42	A	The time between the minimums and maximums sounded equally apart.
46	D	I thought sharp sound would be better drum and piano sound, but in this question sharp sound do not well represent the starting point. I mean in C and D I cannot really identify the starting sound.
47	D	The pitch started high and progressively got lower. As the curve passed into the negative I heard the distinct sound. The duration of the sound also lent itself to curve D.
48	A	Because the min/max sounded the same pitch
49	A	Its the only one that seemed to fit, right down to the bit of the next period.

Table D.4.9 . Text Responses for Question 8.

Subject	Comment
35	I could hear the piano pitch go up and down and I listened to see if the pitch both times was equal. It seemed equal to me, so I picked B.
36	Went high, low, high, low with drums concentrated at the shifts from high to low.
42	I was going by the time between the original incline, to the dip and the incline again.
46	Drum and piano sound have a good effect to represent the ups and downs. If we know about which sound (drum or piano) is for ups or downs, I think I can identify these graphs.
47	Symmetrical pitch and rate for both sides of the curve.
48	Because of the apparant high rise in the beginning
49	That seemed to mostly fit B, but I don't think the derivative was correct.

Table D.4.10 . Text Responses for Question 9.

Subject	Comment
35	By this time, I have a better understanding of how the sounds work together. They all represented the graph well, it just depended on if one was interested in slope and curvature. I think the person taking this test shoud get some warm up graphs to under
36	I like hearig 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 loading.
40	a is simple and pleasing to the ear. d is good and contains more information than (a) and would be easy to understand with limited explanation.
42	A:it went really slow, B:the sound was a bit annoying, but like the pace, C:I am starting to find the drum beats to be annoying, D:had an annoying pitch along with the drum beats. E:I am starting to like the discrete noises of the piano over time without the drum beats. This still may have been to slow, but I liked it the best, F:Drum beats and slow.
47	E and F were the best because they told you when the curve went into the negative region of the y-axis.
48	Again, I really like the negative value changing tone. It really helped to see the graph with my eyes closed
49	The drum used to specify the derivatives in C and F I think are rather hard to identify.

Appendix E C++ code for the DataReader program

The C++ code in this appendix relates to the DataReader program which read in an x, y data set file, converted the data to a chromatic scale, added drum beats, and changed the instrument for negative values. It then wrote the converted data to an SLG text file. There are two files, the first is the DataPane header file, and the second is the DataReader program.

E.1. DataPane.h

```
//© 1997 Steven Sahyun 9/22/97
//DataPane.h
#pragma once
#include <LPane.h>
#include <LCommander.h>
class  DataPane : public LPane, public LCommander
{
   public:
      DataPane( LStream *inStream );
      static  DataPane*  CreateDataPaneStream( LStream *inStream );
  //    static  void WriteData( int max, float *xptr, float *yptr  );
   private:
              double x[500];
              double y[500];
              int i, maxvalue;
              char fileName[64];
              double dataarray[1000];
              double *dataarray_ptr;
   protected:
      virtual void  DrawSelf( void );
                  virtual    void    ClickSelf(   const   SMouseDownEvent
&inMouseDown );
              //virtual void  SaveSelf(void);
};
```

E.2. Data Reader.cp

```
// © 1997 Steven Sahyun 9/22/97 Oregon State University
// sahyuns@ucs.orst.edu
// This code may be used with written permission from the author
// for non-commercial applications.
// written for CodeWarrior Pro 3 compiler
//DataPane.cp
#include "DataPane.h"
#include <fstream.h>
#include <string.h>
```

```
#include <ctype.h>
#include <iomanip.h>

//prototype
void itoa(int n, char s[]);
void ftoa(float n, char s[]);
void reverse(char s[]);
void convertPascalStr (Str63, char *);
void   StandardPutFile   (Str255   prompt,   Str255   defaultName,
StandardFileReply *replyPtr);
#define StandardPutFile

char *pitch( int maxvalue, double yvalue, double *yarray_ptr);
float ReadData(double *dataarray, char fileName[64]);
float WriteData(double *dataarray);
typedef char Str25[26];
typedef char Str250[251];

// ==========================================================
//              create a DataPane pane from a PPob resource

DataPane*  DataPane :: CreateDataPaneStream( LStream *inStream )
{
   return ( new DataPane( inStream ) );
}

// ==========================================================
//                      respond to a mouse click on the pane
void  DataPane :: ClickSelf(const SMouseDownEvent &)
{
//temporary solution until I figure out how to access the data from a
//protected class
// really want this in the cmd_Save section of DataReaderPP.cp
            dataarray_ptr = &dataarray[0];
            WriteData(dataarray_ptr);
}

// ==========================================================
//                      the construct-from-stream constructor
DataPane :: DataPane( LStream  *inStream ) : LPane( inStream )
{
            dataarray_ptr = &dataarray[0];
//get file name here so that it will be part of the datapane object
//choose the file to read in
// Data structures and variables for call to StandardGetFile
            StandardFileReply replyStruct;
            SFTypeList typeList; // we're going to allow all file types
            short numTypes = -1; // allow all types of files
            StandardGetFile (nil, numTypes, typeList, &replyStruct);
            if (!replyStruct.sfGood)
            return; // user cancelled but continue anyway

            convertPascalStr (replyStruct.sfFile.name, fileName);

            ReadData( dataarray_ptr, fileName);

            //load data into object's x, y array
            maxvalue = *(dataarray_ptr++);
```

```
            for (i=0; i<maxvalue; i++)
            {
                        x[i] = *(dataarray_ptr++);
            }
            for (i=0; i<maxvalue; i++)
            {
                        y[i] = *(dataarray_ptr++);
            }

}

// ==========================================================
//                          draw the pane's frame and contents
void  DataPane :: DrawSelf( void )
{
//Display the data
   ::TextFont( systemFont );
   ::TextSize( 12 );

             i = 0;
            Str250 textstring;
            Str25 numberstring;
            unsigned char *text;
            MoveTo ( 10,20);
            strcpy (textstring, "This is Data Set: ");
            strcat (textstring, fileName);
text = c2pstr (textstring); //convert C string to a pascal string
//so can write in window
            DrawString( text );

            for (i=0; i< maxvalue; i++)
            {
//generate x string to write to window, this is a bit cumbersome
                        MoveTo ( 5, i*12 + 35);
                        strcpy (textstring, " x[");
                        itoa (i, numberstring);
                        strcat (textstring, numberstring);
                        strcat (textstring, "]: ");
                        ftoa (x[i], numberstring);
                        strcat (textstring, numberstring);
                        text = c2pstr (textstring); //convert C string
//to a pascal string so can write in window
                        DrawString( text );

//generate y string to write to window, this is a bit cumbersome
                        MoveTo ( 100, i*12 + 35);
                        strcpy (textstring, " y[");
                        itoa (i, numberstring);
                        strcat (textstring, numberstring);
                        strcat (textstring, "]: ");
                        ftoa (y[i], numberstring);
            strcat (textstring, numberstring);
            text = c2pstr (textstring); //convert C string to a pascal
//string so can write in window
            DrawString( text );
            }
}
//=========================================================
```

```
//                                                          functions
//this is a function used in getting the file information
void convertPascalStr (Str63 Pascalstring, char * Cstring)
{
            int length;
            int i;
            length = Pascalstring[0]; // get length byte
            for (i = 0; i < length; i++)
                        Cstring[i] = Pascalstring[i+1];
            Cstring[i] = '\0'; // don't forget the terminating null
}
// reverse: to reverse string s in place; goes with itoa, K&R p. 62
void reverse(char s[])
{
            int c, i, j;
            for (i = 0, j = strlen(s)-1; i<j; i++, j--)
            {
                        c = s[i];
                        s[i] = s[j];
                        s[j] = c;
            }
}
// FTOA: convert n to characters in string K&R p. 64
void ftoa(float nflo, char s[])
{
            int i, sign;
            int n;
            i = 0;
            nflo *= 100;// look at only 2 places past decimal
            n = (int) nflo; // chop off rest
            if ((sign = n) < 0) // record sign
                        n = -n;     // make positive
            while (i < 2)
            {
                        s[i++] = n % 10 + '0';
                        n /= 10;
            }
            s[i++] = '.';               // decimal point
            do {        // generate digits in reverse order
                        s[i++] = n % 10 + '0';   // get next digit
                        }
            while (( n /= 10) > 0);             // delete it
            if (sign < 0)
            s[i++]  = '-';
            s[i] = '\0';
            reverse(s);
}
// ITOA: convert n to characters in string K&R p. 64
void itoa(int n, char s[])
{
            int i, sign;
            if ((sign = n) < 0) // record sign
            n = -n;                                  // make positive
            i = 0;
do {
// generate digits in reverse order
s[i++] = n % 10 + '0';   // get next digit
}
```

```
while (( n /= 10) > 0);                    // delete it

            if (sign < 0)
                        s[i++]   = '-';
            s[i] = '\0';
            reverse(s);
}
// Reading Data from a file
//this reads x y data from a given file and puts it into a single
array.
//It returns a pointer to the array.
//the array sructure is # of items in array, x data, y data.
float ReadData(double *dataarray, char fileName[64])
{
            //get data here
            float result = 0, *result_ptr = &result;
            //float dataarray[1000], *dataarray_ptr;
            double x[500], y[500];
            int i, maxvalue;
            ifstream fin (fileName);
            if (!fin)
                        return *dataarray;
            //read the data here
            i = 0;
            while (!fin.fail())
            {
                        fin >> x[i] >> y[i];
                        i++;
            }

            fin.close();
            maxvalue = i;
dataarray[0] = maxvalue - 1;      //   Data Array 0     1    2   .... n
//n+1 n+2 ... 2n
for (i=1; i<maxvalue; i++)          //   value       max  x0  x1 ... xn-1
//y0  y1  ... y2n-2
            {
                        dataarray[i] = x[i-1];
            }
            for (i= 0; i<maxvalue; i++)
            {
                        dataarray[i+ maxvalue] = y[i];
            }
            return *dataarray;
}

// Writing Data to a file
float WriteData (double *dataarray)
{
            double x[500], y[500];
            //char fileName[64];
            int i, maxvalue;
                        //load data into x, y arrays
            maxvalue = *(dataarray++);
            for (i=0; i<maxvalue; i++)
            {
                        x[i] = *(dataarray++);
            }
```

```
            for (i=0; i<maxvalue; i++)
            {
                        y[i] = *(dataarray++);
            }

Str255 prompt = "\pSave file as:"; // \p tells the compiler to
//make a Pascal string
            Str255 default_name = "\pData.slg";
            StandardFileReply replyStruct;
            StandardPutFile (prompt, default_name, &replyStruct);

//          if (!replyStruct.sfGood)
//                      return *dataarray; // user cancelled; don't
save

// trying to put a name for the data set; for some reason the save dlg
//isn't working
//convertPascalStr (replyStruct.sfFile.name, fileName);
            ofstream fout ("Data.slg"); //ofstream fout (fileName);
//          if (!fout)
//                      return *dataarray;

//this section converts the data and outputs it in SLG text MIDI format
            int  bar, beat, durbeat;
            int beatpart, durbeatpart, TIMEBASE = 100;
            float temp;
            float maxx = 0;
            float maxy = 0, yold = y[0];
            float slope1, slope2;
            float curvature;
            float minx = 9999999;
            float miny = 9999999;
            float zeropoint = 0; // in case of negative y values
            int instrument = 4; //electric piano - works well as sound
does //not decay
            int neginstrument = 6; //harpsichord - for negative values
            int ytickinstrument = 47; //melodic tom - seemed to work
best for //sound as is short
// c2 for - curve, c3 for no curvature, c4 for + curve

            char *note; //pointer to start of the note array value for
the x //data set
char dnote[4]; // "" for the derivative's data set
//data is loaded into function's array here, x is for time, y is for
//pitch
for (i = 0; i < maxvalue; i++)
{
temp = x[i];
if (maxx < temp)
maxx = temp;
if (minx > temp)
minx = temp;
temp = y[i];
if (maxy < temp)
maxy = temp;
if (miny > temp)
miny = temp;
}
```

```
// deal with negative y values by y = y+abs(minimum y), then do a
//program change piano/harpsicord around y = abs(min y)
if (miny < 0)
{
for (i = 0; i<maxvalue; i++)
{
y[i] -= miny; // subtract a negative = shift data set //into positive
//territory
}
zeropoint -= miny; //want to set the new 0
maxy -= miny; //shift max and min values by min value
miny -= miny;
}
//header, uses timebase of 100 for easy calculations
fout << "%MIDIFILE\n@MThd\n@FORMAT 0\n@TIMEBASE
"<<TIMEBASE<<"\n@END\n";
fout << ";-----------------------------------------\n@MTrk 1\n";
//instrument is written here
if (y[0] >= zeropoint )
fout << "001|1|000  1:PROG  " << instrument << "\n";        //main data
//tone on midi chan 1
else
fout << "001|1|000  1:PROG  " << neginstrument << "\n";
//if first point is negative switch to negative instrument
fout << "001|1|000  2:PROG  " << ytickinstrument << "\n"; //dx/dy info
on midi chan 2
fout << setiosflags (ios::fixed | ios :: right) << setprecision(0) <<
setfill('0'); // make the output look pretty

// sound is written here, convert into SLG format for times and pitch
// initial slope beat at start
//checking the curvature
slope2 = (y[1] - y[0])/(x[1] - x[0]);
curvature = slope2;
//may need to put an if statement here limiting sensitivity
if  (curvature > 0)  //positive curvature dy2/dx2 > 0; //concave; lower
//value, like water in a bowl
strcpy (dnote, "C3");
else if (curvature < 0)   //negative curvature dy2/dx2 < 0; //convex,
//higher value, like water hitting a hat
strcpy (dnote, "C5");
else
strcpy (dnote, "C4"); //line; no curvature
//write the beat info here
fout << "001|1|000  2:" << dnote << "  127    0|020\n";
//data notes and beat notes
for (i = 0; i < maxvalue; i++)
{
//sets the duration of the note;
if (i+1 == maxvalue)  //need to worry about last data point
{
durbeatpart = (int) (2000*(x[i] - x[i-1])/(maxx -
minx))%TIMEBASE;//2000 is the total time, Timebase is for the beatpart
durbeat = (int) (20*(x[i]- x[i-1])/(maxx - minx))%4;
}
else
{
durbeatpart = (int) (2000*(x[i+1] - x[i])/(maxx -
```

```
minx))%TIMEBASE;//2000 is related to the total time, Timebase is for
the
beatpart and is same as timebase
durbeat = (int) (20*(x[i+1]- x[i])/(maxx - minx))%4;
}
beat = (int) (2000*(x[i]- minx)/(maxx - minx))%(4*TIMEBASE); // for
intervals less than a beat,
//each beat has 4 parts
beatpart = beat%100; //kludge because I couldn't figure out //how to
//add this in the previous line
beat = (int) (20*(x[i]- minx)/(maxx - minx))%4 + 1;
bar = (int) 20*(x[i] - minx)/(maxx - minx)/4 + 1;
note = pitch( maxvalue, y[i], &y[0]); //put the values of //the
returned pitch pointer value into note pointer
// check to see if crossed zeropoint, if so change instrument.
if (y[i] <= zeropoint && y[i+1] > zeropoint)
fout << setw(3) << bar <<"|"<< beat << "|" << setw(3) << beatpart << "
1:PROG  " << instrument << "\n";
//switch to //positive instrument
if (y[i] >= zeropoint && y[i+1] < zeropoint)
fout << setw(3) << bar <<"|"<< beat << "|" << setw(3) << beatpart << "
1:PROG  " << neginstrument << "\n";
//switch to //negative instrument
//check to see if have passed crossed a Cn value, if so, check slope
and
//print out a beat
if (i > 0 && i < maxvalue-1)  // putting in the derivative //beat here,
//don't take 1st or last points
{
 float beatspan = 0.099; // 1/#of notes wanted in 5 //octaves, less a
//fudge factor, best is normally 0.099
 if ((y[i] - yold)/maxy >= beatspan || (y[i] - yold)/maxy <=
(-1*beatspan)) //sounds span 5 octaves. 2 notes per octave, so check if
crossed a differece point (less a little fudge factor)
{
yold = y[i];

//checking the curvature
slope1 = (y[i] - y[i-1])/(x[i] - x[i-1]);
slope2 = (y[i + 1] - y[i])/(x[i + 1] - x[i]);
curvature = slope2 - slope1;
//may need to put an if statement here limiting sensitivity
if  (curvature > 0)  //positive curvature dy2/dx2 > 0;
//concave lower value, like water in a bowl
strcpy (dnote, "C3");
else if (curvature < 0)   //negative curvature //dy2/dx2 < 0;
//convex, higher value, like water hitting a hat
strcpy (dnote, "C5");
else
strcpy (dnote, "C4"); //line; no curvature
//write the beat info here
fout << setw(3) << bar <<"|"<< beat << "|" << setw(3) << beatpart << "
2:" << dnote << "  127     0|020\n";
}
}
//print out x, y data point
fout << setw(3) << bar <<"|"<< beat << "|" << setw(3) << beatpart << "
1:" << note << "  127   ";
```

```
fout << durbeat << "|" << setw(3) << durbeatpart << "\n";
}
fout << "@END\n";
return *dataarray;
}
char *pitch( int maxvalue, double yvalue, double *yarray_ptr)
{
double yarray[500];
static char note[4] = "   ";
char *charvalue = "";
int value, i;
double ymax;
ymax = 0;
for (i = 0; i < maxvalue; i++)
{
yarray[i] = *(yarray_ptr + i);
if (ymax < yarray[i])
ymax = yarray[i];
}
value = (int) (yvalue*60/ymax)%12;
//turn into a switch function for case ...
switch (value)
{
case 13:
strcpy (note, "Z");//kludge:for some reason it skips //the first case
break;
case 0:
strcpy (note, "C");
break;
case 1:
strcpy (note, "C#");
break;
case 2:
strcpy (note, "D");
break;
case 3:
strcpy (note, "D#");
break;
case 4:
strcpy (note, "E");
break;
case 5:
strcpy (note, "F");
break;
case 6:
strcpy (note, "F#");
break;
case 7:
strcpy (note, "G");
break;
case 8:
strcpy (note, "G#");
break;
case 9:
strcpy (note, "A");
break;
case 10:
strcpy (note, "A#");
```

```
break;
case 11:
strcpy (note, "B");
break;
default:
strcpy (note, "C");
}
value = (int) (yvalue*60/ymax)/12;
switch (value)
            {
            case 0:
            strcpy (charvalue, "2");
            break;
            case 1:
            strcpy (charvalue, "3");
            break;
            case 2:
            strcpy (charvalue, "4");
            break;
            case 3:
            strcpy (charvalue, "5");
            break;
            case 4:
            strcpy (charvalue, "6");
            break;
            default:
            strcpy (charvalue, "7");
            }
strcat (note, charvalue);
return note;
}
```

Appendix F PERL Code for Studies Using the World Wide Web

F.1. Overview

The PERL code in this appendix relates to several programs used to generate the Web pages and record the answers provided by the subjects. Namepage recorded a log of subjects who participated in the studies. It also assigned a code number for each subject and randomly chose which graph type group they would be in. Surveyrecord appended subject's text responses to the survey and pre-test pages to a file called "surveylong" and passed the rest of the information to the first test question page it created from a question file. Temprecord generated the question HTML pages from a question file and appended the answers to a file called "finalresult".

F.2. Namepage

This script appends the name of the subject and a code number to a testlog file, and assigns them to one of three graph categories. It then loads and displays the survey Web pages.

```
#Namepage, ©1997 Steven Sahyun, Oregon State University
#!/usr/local/bin/perl
require '../cgi-lib.pl';
&ReadParse(*in);
#Variables that are passed from name page
$name=$in{'name'};
#Limit input size
$max_string_length = 40; #sets the max length for the test string
$test = $name;
$counter = 0;
while ($test ne "")
{
		chop $test;
		$counter++;
}
if ($counter > $max_string_length)
{
		$name = ""; #if length fails, give null value
```

```
}

$school_code=$in{'SCHOOL_CODE'};

#Limit input size
$max_string_length = 4; #sets the max length for the test string
$test = $school_code;
$counter = 0;
while ($test ne "")
{
            chop $test;
            $counter++;
}
if ($counter > $max_string_length)
{
            $school_code = ""; #if length fails, give null value
}

#Page locations
$FILEPAGE="http://www.physics.orst.edu/~sahyun/survey/quest/namepage.ht
ml";
$NEXTFILEPAGE="/usersA/sahyun/public_html/survey/quest/survey.html";
#Get a student code number ...
open(getcode, "testlog");
            while (<getcode>)
            {
                        $lastline = $_;
            }

close(getcode);
#Get the last number of listing (first item in list array) and
increment
chop($lastline);
($oldcode, $oldname, $oldschool) = split(/  /,$lastline);
$code = $oldcode;
$code++;

#Send them back to the page if no selection
if($name eq '' || $school_code eq '')
{
        print "Location: $FILEPAGE\n\n";
}
#Record name and codes and move to the start of test
else
{
            #Record student and code number in testlog if not 000
#           if ($school_code ne '000')
#           {
                        open(OUT, ">>testlog");
            print OUT "\n".$code."*".$name." *".$school_code;
                        close(OUT);
#       }
            open(IN, $NEXTFILEPAGE);
            print "Content-type: text/html\n\n";
            while (<IN>)
            {
            #check to find the form field to add the code number
                        $line = $_;
```

```
                        chop($line);
                        if ($line eq "<!--code-->")
                        {
                        print "<input name=\"code\" type=\"hidden\"
value=\"".$code."\">\n";
                print "<input name=\"composite\" type=\"hidden\"
value=\"*S:\">\n";
#S: is to mark the start of the survey
                        }
                        else
                        {
                                    print $_;
                        }
            }
            close(IN);
}
```

F.3. Surveyrecord

This script appends subject's text responses to the survey and pre-test pages and passes the rest of the information to the first test question page that it creates from a question file.

```
      # Surveyrecord ©1997 Steven Sahyun, Oregon State University
      #!/usr/local/bin/perl
      require '../cgi-lib.pl';

      &ReadParse(*in);
      #Variables that are passed from page (the answer for the
question)
      $code=$in{'code'};
      $composite = $in{'composite'};
      $gender = $in{'gender'};
      $age = $in{'age'};
      #Limit input size
      $max_string_length = 3; #sets the max length for the test string
      $test = $age;
      $counter = 0;

      while ($test ne "")
      {
              chop $test;
```

```
		$counter++;
	}
	if ($counter > $max_string_length)
	{
		$age = "000"; #if length fails, give age a 0 value
	}

	$hsphys = $in{'hsphys'};
	$cphys = $in{'cphys'};

	#free response answers
	$pcours = $in{'physcourses'};
	$ocours = $in{'othercourses'};
	$ograph = $in{'othergraph'};
	$mus = $in{'music'};
	$diff = $in{'difficulty'};
	$composite = $composite." *".$gender." *".$age." *".$hsphys."
*".$cphys." *M:";

	#to record total time to take the test
	$starttime = time;

	#Question locations
	$NEXTFILEPAGE="/usersA/sahyun/public_html/survey/quest/firstq.htm
l";

	#open file to write the longer survey answers
	open(OUT, ">>surveylong");
	print OUT $code." *pcours: *".$pcours." *ocours: *".$ocours."
*ograph: *".$ograph." *mus: *".$mus." *diff: *".$diff."\n";
	close(OUT);
	#get first question page and write added variables to it
	open(IN, "$NEXTFILEPAGE");
	print "Content-type: text/html\n\n";
	while (<IN>)
	{
	#check to find the form field to add the code number and
composite score
		$line = $_;
```

```
		chop($line);
				if ($line eq "<!--code-->")
		{
	print "<input name=\"code\" type=\"hidden\"
value=\"".$code."\">\n";
	print "<input name=\"starttime\"
type=\"hidden\"value=\"".$starttime."\">\n";
	print "<input name=\"composite\" type=\"hidden\"
value=\"".$composite."\">\n";
		}
		else
		{
				print $_;
		}
	}
	close(IN);
```

F.4. Temprecord

This script generates the question pages from a question file. The question file is a HTML fragment that contains the text of the question, the answers, and <!—code --> and <!—graph--> text. The script inserts the appropriate graph and codes from previous answers, adds a header and footer to create a Web page for the subject.

```
#Temprecord © 1997 Steven Sahyun, Oregon State University
#!/usr/local/bin/perl
require '../cgi-lib.pl';
&ReadParse(*in);
#Variables that are passed from page (the answer for the question)
$code=$in{'code'};
$composite=$in{'composite'};
$number=$in{'number'};
$answer=$in{'answer'};
$starttime=$in{'starttime'};

#Number of questions in the test
$numquest = 14;

#Question locations
$FILEPAGE="/usersA/sahyun/public_html/survey/quest/mtq".$number.".html"
;
$NEXTFILEPAGE="/usersA/sahyun/public_html/survey/quest/mtq".($number+1)
.".html";
```

```
$thankyou="http://www.physics.orst.edu/~sahyun/survey/quest/thankyou.ht
ml";

#Send them back to the page if no selection
if($answer eq '')
{
        open(IN, "$FILEPAGE");
        &output;
            close(IN);
}
#Record answer and move to the next question
else
{
#If anwered last question then record final answer and go to Thankyou
page
            if($number == $numquest)
            {
                        $etime = time - $starttime;
                        $composite = $composite." *".$number."
*".$answer." *etime: *".$etime;
                        open(OUT, ">>finalresult");
                        print OUT $code."  ".$composite."\n";
                        close(OUT);
                        print "Location: ".$thankyou."\n\n";
            }
#otherwise move on to the next question
            else
            {
            open(IN, $NEXTFILEPAGE);
                        $composite = $composite." *".$number."
*".$answer;
                        $number++;
            &output;
            close(IN);
            }
}

sub output
{
print "Content-type: text/html\n\n";
print "<html>\n";
print "<h1><TITLE>Graph Test Question ".$number."</TITLE></h1>\n";
print "<body>\n";
print "<head>\n";
print "<!-- Author: Steve Sahyun (sahyun\@physics.orst.edu) -->\n";
print "</head>\n";
print "<h1><center>Question \#".$number.": </center></h1>\n";
print "<p>\n";

while (<IN>)
{
   $line = $_;
   chop($line);
   if ($line eq "<!--graph-->")
   {
print "<p>\n";
print "<center>\n";
print "<img
```

```
src=\"http://www.physics.orst.edu/~sahyun/survey/quest/mtq".$number."pi
c.gif\" align=bottom alt=\"Graph for question ".$number."\">\n";
print "</center>\n";
print "<p>\n";
print "<center>\n";
print "<EMBED
SRC=\"http://www.physics.orst.edu/~sahyun/survey/quest/mtq".$number."sn
d.mov\" HEIGHT=24 WIDTH=200 CONTROLLER=TRUE LOOP=FALSE
AUTOPLAY=FALSE>\n";
print "</center>\n";
print "<p>\n";
   }
   elsif ($line eq "<!--code-->")
   {
print "<br>\n";
print "<FORM METHOD=\"POST\"
#ACTION=\"http://www.physics.orst.edu/cgi-bin/sahyun/temprecord\">\n";
print "<input name=\"code\" type=\"hidden\" value=\"".$code."\">\n";
print "<input name=\"starttime\" type=\"hidden\"
value=\"".$starttime."\">\n";
print "<input name=\"composite\" type=\"hidden\"
value=\"".$composite."\">\n";
print "<input name=\"number\" type=\"hidden\"
value=\"".$number."\">\n";
   }
   else
   {
            print $_;
   }
}
print "<p>\n";
print "</form>\n";
print "<p>\n";
print "Questions about the test?  Send e-mail to: <a
href=\"mailto:sahyun\@physics.orst.edu\">\n";
print "<address>sahyun\@physics.orst.edu</address></a>\n";
print "<p>\n";
print "Last modified September 11, 1997.\n";
print "</body>\n";
print "</html>\n";
}
      #The End!!!
```