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 or 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)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF APPENDICES

LIST OF APPENDICES (continued)

LIST OF FIGURES

Figure | Page

LIST OF APPENDIX FIGURES

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF APPENDIX TABLES

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

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \qquad (2.1)$$

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

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

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

$$s = \sqrt{\frac{\sum\left(X_i - \bar{X}\right)^2}{n-1}}. \tag{2.3}$$

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

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

where n = number of samples, μ = mean of a data group, and $\bar{X}$ 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. 83].

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

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

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

$$\frac{x - x_{\min}}{x_{\max} - x_{\min}} = \frac{n - n_{\min}}{n_{\max} - n_{\min}} \qquad (2.10)$$

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

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

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

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

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.