Effects of Digital Audio Quality on Students' Performance in LAN Delivered English Listening Comprehension Tests

A dissertation presented to

the faculty of

the College of Education of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Xiangui Yang

March 2009

This dissertation titled

Effects of Digital Audio Quality on Students' Performance in LAN Delivered English Listening Comprehension Tests

by

XIANGUI YANG

has been approved for

the Department of Educational Studies

and the College of Education by

Teresa J. Franklin

Associate Professor of Educational Studies

Renée A. Middleton

Dean, College of Education

ABSTRACT

YANG, XIANGUI, Ph.D., March 2009, Curriculum and Instruction, Instructional Technology

Effects of Digital Audio Quality on Students' Performance in LAN Delivered English Listening Comprehension Tests (110 pp.)

Director of Dissertation: Teresa J. Franklin

Web-based language listening tests require digitized and compressed audio files which can be stored on computer storage media and then delivered through the Internet, intranet or local area network (LAN). For digital audio, higher sound quality is achieved at the expense of bigger data size and greater bit rate since higher sound quality requires higher sample rate and higher sample depth. This causes a dilemma for testers in choosing the proper sample rate and sample depth for digital audio recording for the purpose of web-based English listening comprehension tests. While higher quality is desirable for better test validity and reliability, lower quality is preferred so that the audio can be transferred smoothly through the Internet, intranet or LAN. This study examined the effects of digital audio quality on students' performance in a LAN delivered English listening comprehension test created with the World Wide Web (WWW) technology. Based on the data collected through the English listening comprehension test participated by 335 EFL students in a LAN environment, the researcher compared the effects of three different sample rates (44 kHz, 22 kHz and 11 kHz) and two different sample depths (16 bit and 8 bit). Using the statistical method of 2 x 3 ANOVA, the researcher has found that students tested with 44 kHz audio and 22 kHz audio performed significantly better than those

tested with 11 kHz audio, but students tested with 16 bit audio did not perform significantly better than those tested with 8 bit audio. No interaction effect was found between sample rate and sample depth. The researcher concluded that 22 kHz and 8 bit should be set as the standard for digital audio for the purpose of web-based English listening comprehension tests and all equipment involved in such a test must be qualified for effectively recording, transmitting or rendering digital audio of 22 kHz and 8 bit.

Approved: _____

Teresa J. Franklin

Associate Professor of Educational Studies

ACKNOWLEDGMENTS

First of all I would like to extend a most gracious thank you to all the members of my dissertation committee. To Dr. Teresa Franklin, chair of my dissertation committee, thank you for your wise guidance and heart-warming encouragement throughout my entire graduate experience. Thank you for your kind support and generous help during all my moments of difficulties. Thank you also for your valuable mentorship and advice that have empowered me to complete this dissertation. To Dr. George Johnson, I am very grateful for your advice that has greatly improved the design of this study. Thank you for teaching me the knowledge of statistics that has become the engine of this study, and thank you for spending your valuable time helping me with the statistic analyses. To Dr. Gregory Kessler, thank you so much for enlightening me with your rich knowledge in the field of computer aided language learning, and thank you for directing me to the rich source of literature. Your insightful comments and questions have held me to a high research standard. To Dr. Albert Akyeampong, I am very grateful that you could join my dissertation committee on such a short notice. Thank you for your advice on the structure of the dissertation, and thank you for your advice on the data collection plan.

I am very thankful to my friend Yihong Zhang, who has helped me to arrange the data collection process and provided me with descriptions of the equipment used for this study. I am also grateful to Yue Gao for his valuable and timely support. This study could not have been possible without their generous help.

My special thanks go to my father-in-law, Yongsheng Xiao, and mother-in-law, Changfen Na, who came thousands of miles away from China to take care of my children

so that I can concentrate on my dissertation. Thank them for their loving care when I fell ill, and thank them for their gentle encouragement during my time of frustration.

I would also like to thank all the members of my family. During the days that I had to devote all my time to this study, my wife, Mingli Xiao, took the burden of the entire house work while working full time as an instructional designer. Her love, patience and support have helped me to pass the most difficult times through out the research process. I am so grateful to my three children, Grace, Daniel and Evan. They give me reason and strength to achieve what I have achieved today.

TABLE OF CONTENTS

Page

Abstract…….…….....3

Acknowledgements……..….....5

Chapter One: Introduction 13

Background of the Study 13

Statement of Problems.....14

Purpose of the Study 17

Research Questions.....18

Research Hypotheses 19

Significance of the Study.....20

Limitations and Delimitations of the Study.....22

Definitions of Terms 23

Organization of the Study…….....24

Chapter Two: Literature Review 26

History of Computerized Testing.....26

Advantages of Computerized Testing.....27

Disadvantages s of Computerized Testing.....29

Design and Development of Web-based Testing……..30

Comparability of Paper Based Tests and Computer Based Tests…….31

Currently Technology That Supports Web-based Testing……32

Studies on Audio Technology……..…….....33

The Nature and Characteristics of Sound……...…..…33

Analog Audio vs. Digital Audio……...…...…34

Sample Rate and Sample Depth of Digital Audio……...…….....34

Human Perception of Sound…….....35

Digital Audio Compression……...……..35

Computer Assisted Language Learning and Language Testing…….....36

Technology in Computerized Language Listening Assessment…….…….37

Summery Literature Review……..…….…….38

Chapter Three: Methodology.....40

Population and Sample……...…….....41

Instrument Used for the Study……...…….....44

Equipment Used for the Study……...…..……47

The Audio Cassette Tape Player……...……..47

The Audio Digitizing Computer…….……..…..48

The Test Server……...…….....…..48

The Computer Labs……...…….……49

Data Collection Methods…….....…..50

Operational Definitions of the Variables…….....51

Independent Variables……..…….....51

Dependent Variables……..…….....51

Methods for Data Analysis..……...…....52

Methods for Answering Research Questions……..……..52

Chapter Four: Findings and Discussion.....54

The Data Collection Procedure.....54

The Pilot Study…….....…..55

The Actual Test for Data Collection…….....58

The Data Analysis and Results……..……...…..60

Description of Data…….……..……..60

Assumption of Testing…….……..……62

Hypothesis Testing Results…….……..…….....…67

Answers to Research Questions……..……..……..….73

Additional Findings……..……...…..……75

Chapter Five: Conclusions…….....77

Implications of the Research.....77

Limitations and Weaknesses……..……....….80

Recommendations for Future Research…….….83

Summery and Conclusion…….…….……84

References…….....87

Appendix A: Sample Pages of Listening Comprehension Section of

CET Band 4…….....….....96

Appendix B: English Listening Comprehension Test as the Instrument for

This Study……..…….…..….103

Appendix C: Letter of Permission for Use of CET Test Items as the Instrument

of This Study…….…….....…107

Appendix D: Contact Information of Researcher and Research Assistant…….....…108

Appendix E: IRB Approval Letter…….…….…….109

Appendix F: IRB Amendment Approval Letter……...…….....…110

LIST OF TABLES

Page

Table 4.1: Results of Pilot Test Participated in by Twelve English Majors…….57

Table 4.2: Descriptive Statistics of Participants' Test Scores……..…61

Table 4.3: Kolmogorov-Smirnov Test for Normality Assumption of Test Scores…..65

Table 4.4: Results of Levene's Test of Equality of Error Variances…….....67

Table 4.5: Two-way ANOVA: Sample Rate by Sample Depth on Test Score…..….70

Table 4.6: Tukey's test for Comparison among Three Levels of Sample Rates.....72

LIST OF FIGURES

Page

Figure 4.1: Histogram of the Distribution of Participants' Test Scores……..….63

Figure 4.2: Boxplots of the Distribution of Participants' Scores in Each Cell for the Two-way ANOVA Design……..…64

Figure 4.3: Histograms of the Distribution of Participants' Scores in Each Cell for the Two-way ANOVA Design…..…….…..66

Figure 4.4: Line Graph of Sample Rate and Sample Depth for the Test Score…..…..71

CHAPTER ONE: INTRODUCTION

Background of the Study

With the advent of modern computer technology and the development of distance education, computerized assessment has become increasingly important to educators. As Rabinowitz and Brandt (2001) have stated, computer technology has become essential in today's society and its great potential for improving assessment could make computerized testing the main stream of the future test administration. Computerized testing comes in several different forms, which include Computer Based Testing (CBT), Computer Adaptive Testing (CAT), and Web-based Testing (WBT). CBT and CAT have different underlying theoretical foundations (Mason, Patry & Bernstein, 2001). The regular CBT was simply a paper-and-pencil test carried out on a computer, in which assessment items for students were identical (Mason et al., 2001). Conversely, CAT has been developed on the basis of the Item-Response Theory (IRT) and has the ability to adapt the test to the individual test taker by selecting the next item for the test taker in accordance with his or her responses to the previous item (van der Linden & Glas, 2000).

Web-based testing, or WBT, is the version of CBT or CAT that is built using the language for the web so that it can be delivered over computer networks and taken using web browsers on individual computers (Roever, 2001). Although based on the technology for the World Wide Web (WWW), WBT is suitable for delivery not only on the Internet but also on the intranet or the local area network (LAN) provided that http service is available and, if necessary, media streaming service and database system are installed on the server. WBTs can be accessed at test taker's own choice of time and

location (Roever, 2001). Today, because of the rapid development of computer and network technology, the designs of many assessment systems are oriented toward WBT, offering the possibility of radically changing the ways students take tests and promoting the use of large-scale assessment (Hamilton, Klein & Lorie, 2000).

In the field of language education, CBT, CAT and WBT (also known as Computer Assisted Language Testing, or CALT), are playing an increasingly important role (Wang, 2003). The use of computers allows language educators to administer large-scale language tests and obtain the test results immediately, and the newest format, Web-based Language Testing (WBLT), allows a learner to take a language test using the Internet from almost anywhere in the world at anytime (Roever, 2001). Many professional testing agencies have been aggressively striving to migrate their tests online. The Educational Testing Services (ETS), for example, has successfully administered the Test of English as a Foreign Language (TOEFL) through the use of computers for several years and has now successfully adopted a web-based version of this important test, the TOEFL iBT (ETS, 2006). It is evident now that web-based language testing should be considered an integrated part of language education, especially distance language education. If successfully developed, it can serve as a powerful tool to enhance the implementation of distance language teaching and learning.

Statement of the Problems

Computerized testing, especially web-based testing is still relatively new and many questions still exist with respect to its characteristics and its comparability to traditional paper-based testing (Wang, 2003). The same is true in the field of computer

assisted language assessment where some problems still exist (Chapelle, 2001). Many people who are influenced by computer assisted language testing have expressed the concern that computer delivered tests may be less valid than paper based tests (Chapelle, 2001). Concerning language listening assessment, there are such questions as whether video input and aural input involve the same abilities from test takers, or whether we can interpret scores from a video based listening test in the same way as an audio based listening test (Alderson & Bachman, 2006).

This study examines the particular issue of digital audio in web-based English listening comprehension tests. Unlike traditional paper and pencil based tests, where audio is recorded and delivered in analog format, web-based language listening tests require digitized and compressed audio files which can be stored on computer storage media, e.g., a hard drive, and then delivered through the Internet, intranet or LAN. Analog audio is typically recorded on a cassette tape and then played back using a tape player to test takers either through speakers or through headphones. Since no digital data storage or data transmission is involved in this process, audio can be made as high quality as possible. In contrast to analog audio, digital audio has a unique property with respect to the relationship between sound quality and data size. That is, the higher its sound quality, the bigger the data size, since higher sound quality is achieved at the cost of higher sample rate and sample depth (Scanlan, 2002), and higher sample rate and sample depth proportionally increase the cost of storage space (Watkinson, 1994). Web-based language testing, similar to traditional language testing, requires audio quality to be as high as possible in order to achieve desired test validity. Yet, on the other hand, it

requires the audio data to be as small as they can be to achieve smooth data transmission through a computer network. This has created a great dilemma for web-based English listening comprehension test designers and administrators (hereafter referred to as testers). Should testers keep the audio quality as high as possible and rethink the principles for web-based language assessment? Should they wait for better computer and Internet or LAN technology to be developed before considering moving language listening comprehension assessment online? Or, should they sacrifice audio quality to accommodate the current design and the current ability of computer and network systems? If testers decide to go for high audio quality, to what extent can they increase the sample rate so as not to waste the computer and network resource, since the human ears are most sensitive to frequencies between about 2 kHz and 5 kHz and can hear nothing beyond 20 kHz (Watkinson, 1994)? This question is even more difficult to answer if testers take sample depth into consideration as well. Although in general, higher bit depth means better sound, a 24 bit digital audio may not be necessary for web-based listening tests since higher sample depth mostly improves audio quality by increasing the dynamic range and improving signal to noise ratio (SNR) (Bartlett & Bartlett, 2002), which may not be as important for human speech as for instrumental music (Liang, 2001). On the other hand, if testers would prefer smaller data size for easy storage and smooth transfer, to what extent can they suppress sample rate and sample depth so that the audio, even though low in quality, still meets the request for good validity of the test? These issues are further complicated by the fact that students in an English as a Foreign Language (EFL) setting, especially in China, are mostly trained with

grammar translation and audioligual methods, and that an audio presentation comprehensible for a native speaker may not sound clear at all to a test taker.

Purpose of the Study

The purpose of this study is to examine the effect of digital audio quality on students' performance in a LAN delivered web-based English listening comprehension test. The test was created using the language of the World Wide Web (WWW) and was suitable for delivery on the Internet. However, due to the restrictions of the Internet environment at the location where data was collected, this web-based English listening comprehension test was delivered in a LAN environment where common Internet services were installed. This test is referred to as a web-based English listening comprehension test throughout this paper in order to reflect its essence of web-based technology.

The two major factors that determine the quality of digital audio are sample rate and sample depth (Scanlan, 2002). Other factors such as compression rate (Pan, 1993) and noise level (Liang, 2001) also affect digital audio quality. It would be difficult to design one single study that looks into all these factors simultaneously. This study has focused on sample rate and sample depth, which are the two most critical factors that determine the quality of digital audio. By comparing digital audio recorded with different sample rates and sample depths in the context of web-based English listening comprehension tests, the researcher has sought to discover how sample rate and sample depth in digital audio affect test takers' performance, and how these two factors interact in shaping the quality of the tests. Through answering these questions, a best compromise

point between audio quality and data size may be identified and a guideline for controlling digital audio quality may be established for the purpose of web-based English listening comprehension assessment. Although digital audio recorded at low sample rate and low sample depth, 11 kHz and 8 bit in this case, could be "good sounding" to the human ear (Currier, 1996, p.3), whether it is suitable for web-based language listening comprehension assessment still remains unknown. Findings of this study could provide some insight on this issue.

Research Questions

In digital audio, different sample rates produce different audio qualities and have different uses. A sample rate of 8 kHz produces low quality audio and is mostly used in telephone systems (Luther, 1997); 11 kHz is good for sound of 5 kHz and is the lowest sample rate that is suitable for human voice recording (Liang, 2001); 44 kHz results in high quality and is the standard for digital audio CDs (Stockdale, 2002); 96 kHz high quality audio is used for the new DVD audio standard (Plichta & Kornbluh, 2005). Similarly, different sample depths also result in different audio qualities and have different uses. Eight bit audio is low quality and is only used for telephone or for the purpose of saving space on storage media (Liang, 2001); 16 bit is the standard for CD quality audio (Stockdale, 2002); 24 bit is the new high definition standard and is found on Digital Audio Tape (DAT) machines or on DVD-audio disks (Plichta & Kornbluh, 2005). These two factors, however, affect digital audio quality in different ways. Sample rate mainly affects the frequency range of the audio, while sample depth mostly affects the dynamic range and noise level (Bartlett & Bartlett, 2002). Although there have been

studies on how these two factors affect the comprehension of speech content of digital audio (Xu, Thompson & Pfingst, 2005), their impact on students' performance in English listening comprehension tests remains largely unknown. This study will examine the sample rates of 44 kHz, 22 kHz and 11 kHz, and the sample depths of 16 bit and 8 bit, since these sample rates and sample depths are most commonly used in recording digital audio. The following research questions are the focus of this study:

1. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample rates (44 kHz, 22 kHz and 11 kHz)?
2. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample depths (16 bit and 8 bit)?
3. Do sample rate and sample depth interact in affecting students' performance in web-based English listening comprehension tests?

Research Hypotheses

In order to answer the research questions of this study, the following research hypotheses were tested:

HO_1: There is no significant difference of the test scores among test with 44 kHz digital audio, test with 22 kHz digital audio and test with 11 kHz digital audio.

HO_2: There is no significant difference of the test scores between test with 16 bit digital audio and test with 8 bit digital audio.

HO_3: There is no significant relationship between sample rate and sample depth in terms of the test scores.

Significance of the Study

Web-based language testing, especially web-based language listening comprehension testing, is still relatively new to language educators (Wang, 2003). Although some research (Sereci, 2003, Dunkel, 1999) have been conducted on such issues as its development, characteristics and comparability to traditional paper based tests, very little research has been carried out on the quality of digital audio in computerized language tests or web-based language tests. The study of digital audio quality for computerized language assessment, especially web-based language listening comprehension assessment, is highly significant in several aspects.

Computerized language assessment is now a reality to all language learners, language educators and applied linguists (Chapell, 2001), and web-based assessment allows existing computer labs of an educational institute to become testing centers (Bridgeman, Lennon & Jacknthal, 2003). This situation, however, has been complicated by the uneven distribution of computer capability or the difference in network speed in different geographical locations. It is quite likely that test designers and administrators in different areas use audio of different standards to suit their own technological conditions or financial situations. This may seem to be reasonable and acceptable for low stakes or localized assessment activities. However, this is not acceptable for WBT in its true meaning since the primary purposes and advantages of web-based language assessment have been large-scale and globalization (Roever, 2001). To make sure that no test taker

will be disadvantaged by a particular configuration of the audio used in a web-based language listening test, standards will have to be set for the factors that affect digital audio quality that in turn may affect test validity. The findings of this study may serve as a guideline to establish these standards.

Although in general higher audio quality is always preferred over lower audio quality for the purpose of language listening comprehension assessment, the cost and effort associated with high quality audio may not always pay off since the human ear is mostly sensitive to frequencies between 2 kHz and 5 kHz and most listeners cannot even tell if the highest frequency in sound is 20 kHz or 16 kHz (Watkinson, 1994). Extremely high quality audio, such as audio recorded at 96 kHz of sample rate and 24 bit of sample depth may well pass the saturation point of the human hearing ability and thus cause waste of computer resource and money. On the other hand, there might be a point on the continuum of audio quality where, even comparatively low, test validity is perfectly maintained without excessive cost or high demand on computer and network capability. This study sought to identify such a point for the purpose of web-based English listening comprehension test.

Findings of this study may also provide some reference for selecting computer hardware and accessories for the purpose of web-based English listening comprehension tests. Some computer hardware components (such as a sound card) and some computer accessories (such as headphones or audio cables) have certain frequency response properties and dynamic range properties. Knowing what sample rate and sample depth are suitable for web-based English listening comprehension tests means knowing what

specifications these types of hardware or accessories must have to be qualified for use in the testing practice.

Limitations and Delimitations of the Study

As mentioned earlier, the quality of digital audio is determined by many factors such as sample rate, sample depth, compression level and noise (Scanlan, 2002; Pan, 1993; Liang, 2001) and an experimental design incorporating all these features at deferent levels will not be feasible. The major limitation of this study is its inability to examine the whole picture of digital audio. The result of this study will only provide ideas on how sample rate and sample depth in digital audio may affect web-based language listening comprehension test. Other studies similar to this one are still needed to provide a comprehensive understanding of digital audio in web-based English listening comprehension tests. Another limitation of this study lies in the fact that the signal source for this experiment comes from prerecorded tape materials. Although the recordings are carried professionally on high performance cassette tapes, their quality does not fully represent what is achievable in ideal recording environments such as a sound studio. The result of this study could be a valuable guidance for digitizing existing materials recorded on tapes for the purpose of web-based English listening comprehension assessment. Yet, any generalization of these findings to direct studio digital recording should be carried out with caution. Moreover, since this study is conducted using participants and instruments from an EFL environment, its findings will be most suitable for reference in such an environment. Findings may not apply to web-based English assessment in English as Second Language (ESL) settings.

Definition of Terms

Some terms that are important for understanding this study may be unfamiliar to the reader. They are defined and explained as follows:

Audiolingualism: a structure-based language teaching method that emphasizes pattern drills and the practice of listening and speaking using such tools as tape recordings or language labs. In Audioligualism, the teacher is the center of the class, modeling the target language and controlling the direction and pace of students' learning activities. (Richards & Rodgers, 2001).

Bit: a single number having a value of 0 or 1. The word *bit* is used as a short form for *binary digit*, which the computer uses to represent different states. One bit can represent two states, e.g., on and off; two bits can be used to represent four states; three bits represents eight states, and so on (Clifford, 1992; Adobe, 2003).

College English Test (CET): a national large-scale standardized English test for college students in China. Generally referred to as CET 4 and CET 6, College English Test is given twice each year and it plays an important role in college and university English teaching in China (NCETC, 2006).

Computer adaptive testing (CAT): the form of computerized testing that has the ability to select the next item according to the test taker's responses to the previous item (van der Linden & Glas, 2000).

Computer assisted language learning (CALL): language teaching and learning carried out with the aid of computer-based learning activities know as courseware. CALL activities in the U.S. started as early as the 1960s (Chapelle, 2001).

Computer assisted language testing (CALT): language assessment carried out with the aid of computer technology. Examples of CALT include such types of project as computer adaptive reading test, computer assisted writing test, computer assisted listening test and reading test with computer assisted response analysis (Chapelle, 2001)

Computer based testing (CBT): computerized version of a paper-and-pencil test, in which assessment items for all test takers were the same (Mason et al., 2001).

Digital audio: an audio signal represented by a stream of digits, which can be stored and processed by a computer (Luther, 1997).

Grammar translation: a language teaching method that emphasizes detailed analysis of grammar rules and translation skills, with the first language as a reference system (Richards & Rodgers, 2001).

Local area network (LAN): a computer network that does not reach beyond a local area such as within a building. LANs are capable of transmitting data at a much faster rate than the Internet (Wang, 2003).

Signal to noise ratio: the comparison of the level of a signal and the level of a noise voltage (Clifford, 1992).

Web-based testing (WBT): CBT or CAT that is built using the language for the web so that it can be delivered over computer networks and taken using browsers for the World Wide Web on individual computers (Roever, 2001).

Organization of the Study

This dissertation consists of five chapters. Chapter One gives the background of this study, states the problems and introduces the research questions. Chapter Two

presents a review of some literature on computerized testing in general, web-based testing in particular, computer assisted language testing as well as the study on digital audio properties. Chapter Three describes the methods used in the study, including the process of selecting participants, the research instruments, methods for collecting data, and analysis of data. Chapter Four presents an analysis of data and the findings from the analysis. Chapter Five summarizes the study and discusses the implications of the findings, the limitations of the study, as well as possible future research.

CHAPTER TWO: LITERATURE REVIEW

This chapter is a review of the literature in the areas related to web-based English listening comprehension tests. This review covers both computerized testing in general and web-based testing in particular. This review also covers such other topics as computer assisted language learning, computer assisted language testing, and studies on sound properties and digital audio technology.

History of Computerized Testing

Computers have been used for decades for the purpose of testing (Chapelle, 2001). In the 1960s and 1970s, the U.S. Department of Defense saw the potential benefits of computerized testing and began to support a wide range of theoretical research in this field (Dunkel, 1999). In late 1970s and 1980s, some science educators explored the use of microcomputers in attitudinal or achievement assessment in science education (Cole, MacIssac & Cole, 2001; Arons, 1984; Waugh, 1985). Before web-based testing (WBT) was developed, computerized testing consisted mainly of two types—regular computer based testing (CBT) and computer adaptive testing (CAT). These two types of computerized testing are based on different theoretical foundations (Mason, Patry & Bernstein, 2001). The regular CBT was simply a paper-and-pencil test carried out on a computer, in which all test takers were given the same test items (Mason et al. 2001). Conversely, CAT was developed on the basis of the Item-Response Theory (IRT), emphasizing that the probability that a person answers an item correctly was determined only by the person's ability (Hambleton & Swaminathan, 1985). In fact, the notion of adaptive testing can be traced back to the original academic screening test developed by

Binet in 1905 (van der Linden & Glas, 2000), in which test items became increasingly difficult as the test progressed and students' patterns of response were used as stopping rules for ending the test (Sereci, 2003).

The early attempts to build CBT tests relied on large-scale computers and were often not successful and very expensive (Dunkel, 1999). By the early 1980s, personal computers had achieved the power of large-scale computers of earlier years, thus enabling computer aided testing to step out of the stage of theory studies into real world implementation (Dunkel, 1999).

With the development of computer and Internet technology, CBT and CAT, which were once hosted only on individual computers, were upgraded to work on LANs (Wang, 2003), or on the Internet so that users could take them using browsers for the World Wide Web (Roever, 2001). The name Web-based Testing, or WBT, has been used to distinguish this new form of test from the more conventional CBT and CAT (Wang et al., 2004). Compared with CBT on a PC, web-based testing is more flexible and more useful for educational research (Bonham, Beichner, Titus & Martin, 2000). Today, because of the rapid development of Internet technology, many assessment systems are designed for WBT, which will potentially change the ways students take tests and make large-scale assessment possible (Hamilton et al., 2000).

Advantages of Computerized Testing

Computerized testing, either CBT or CAT, offers good testing efficiency, and has such advantages as flexible administration schedules, improved test security, instantaneous grade reporting, and inclusion of multimedia in the assessment (Sereci,

2003). Computer adaptive testing, in particular, enables testers to confidently estimate test takers' performance using fewer items than what may be required on non-adaptive tests (Hamilton et al., 2000). CATs are usually shorter than their paper-and-pencil versions and are usually about half the length of comparable non-adaptive tests (Wainer, 1993). Some other advantages of computerized testing include self-grading, having the ability to incorporate audio and video, supporting test takers with disabilities, as well as immediate feedback (Roever, 2001; Wang, 2003). Web-based testing, WBT, carries almost all the above advantages except a good test security. The biggest advantage of a WBT is its flexibility in terms of time and space, since all that is required to take a WBT is a computer with a web browser and an Internet or LAN connection (Roever, 2001). WBT allows students to choose their own time and place for the test and, in the mean time, testers can easily share their test with colleagues in different parts of the world (Roever, 2001). WBT can be independent from the tester when supported with scoring scripts for dichotomously-scored items, offering flexibility and convenience for test takers (Brusilovsky & Miller, 1999). Even when high stakes are involved, there are still advantages to delivering a test with web technology. Whereas traditional computerized tests require high programming skills and specially-designed delivery platforms, WBTs are comparatively easy to write and implement (Roever, 2001). Most people with a computer and some basic HTML skills are able to write a WBT, and it only requires a computer and a web browser for a student to take the test (Rover, 2001). In addition, HTML allows test designers to insert images, videos and sound and free or low-cost editing programs are available that further ease the burden of test designing (Wang,

2003). WBT is not only inexpensive for test designers and test takers; it is also inexpensive for test administrators since WBTs contain mostly small files that can be easily hosted in a free space on an institute or commercial server (Roever, 2001).

Disadvantages of Computerized Testing

The biggest limitation of computerized testing is the problem of cheating and item confidentiality (Roever, 2001). Since a test taker can take a computerized test, especially web-based test, without supervision, it is difficult to ensure that the test taker answers the test questions independently (Wang, 2003). A possible way of solving such a problem is to include web cam support in the test (either through a plug-in or a built-in module) so that students' activities can be monitored or recorded if they decide to choose their own location or time for the test (Wang, 2003). Item confidentiality is also at risk since test takers could copy items off the screen when they are taking the test in uncontrolled conditions (Roever, 2001). Moreover, WBT items are usually downloaded into the web browser's cache on the test taker's computer and therefore they can be easily accessed by a computer-savvy person (Hamilton et al., 2000). To solve this problem, Hamilton et al. (2000) suggest that the item pool be constantly refreshed so that each item is used as few times as possible. The self-scoring feature of WBT can cause cheating and item exposure problems. JavaScript is often used to make computerized tests self-scoring, but the script may contain all the answers, which a computer-savvy person can easily discover by viewing the test's source code (Rover, 2001).

The second major disadvantage of computerized testing lies in the issue of fairness. Besides the difference in Internet connection quality, one type of computer may

perform better than another type and even computers of the same type may perform differently at the time of testing (Wang, 2003). WBT relies on the use of computers, which might be a problem for those schools that do not have enough qualified computers for test administrations (Sereci, 2003). To gain fairness in implementing WBT, research should be conducted on which schools need improvement in this aspect (Hamilton et al., 2000). Testers may need to standardize equipment and connectivity for WBT by adopting facilities (computers and network structure) dedicated solely to testing (Wang, 2003). As a supplementary measure, additional programs can be integrated in web-based tests that will detect the test takers' connection speeds and hardware abilities and give warning for or reject the machines that do not meet required processing speed or data transfer ability (Wang, 2003).

Design and Development of Web-based Testing

In discussion of the process and good practice of CBT interface development and design, Fulcher (2003) pointed out that interface development and design is extremely important in computer based language testing since they may constitute a threat to construct validity. Fulcher (2003) suggested that CBT interface design should consist of three phases: initial planning and design, concurrent activities, and field testing. The first phase was for developing test specifications and designing interface prototypes; the second phase was for usability testing; and the third phase was for trials and fine tuning of the product (Fulcher, 2003).

Storage of test items can be achieved with different technology methods. Brusilovsky and Miller (1999) distinguished two different ways to store test questions:

presentation format and internal format. According to Brusilovsky and Miller (1999), storing a question in presentation format means storing it as HTML code, while storing a question in an internal format means storing it in database systems such as QuestWriter, Carnegie Mellon Online and LearningSpace. It is also possible to use structured directories to store files for randomizing tests or test items (Wang, 2003).

According to Brusilovsky and Miller (1999), five technologies can be used to deliver test items for web-based assessment: HTML links, HTML CGI forms, scripting language, plug-in, and Java. A commonly used technique in developing web-based tests is to combine HTML forms and CGI-compliant evaluation scripts (Brusilovsky & Miller, 1999). Brusilovsky and Miller (1999) advocated the use of a JavaScript for interactive questions and the use of plug-ins for more advanced interface design.

Comparability of Paper Based Tests and Computer Based Tests

Haas and Hayes (1986) compared paper based assessment and computer based assessment with different presentation format. Haas and Hayes (1986) reported that when a text passage was displayed on multiple pages on the computer screen, computer administered test resulted in lower scores than paper-and-pencil test due to the difficulty of reading the extended text on the screen. In a study with a post-test-only design involving 105 freshman business under graduates, Clariana and Wallace (2002) also examined the comparability of paper based and computer based tests. Through observing such key factors in the two forms of test as content familiarity, computer familiarity, competitiveness and gender, Clariana and Wallace (2002) concluded that gender, competitiveness, and computer familiarity were NOT related to a particular form of test,

but that content familiarity appeared to interact with test modes. Clariana and Wallace (2002) argued that computer based items may increase transition time and memory load and force test takers to focus on individual items.

MacCann, Eastment, and Pickering (2002) conducted a study that compared two modes of responding to essay questions: by computer and by pen and paper. MacCann et al. (2002) found that there were no significant differences in mean scores between the participants tested with pen-and-paper mode and those tested with computer response mode. Some other researchers also reported no differences in students' performance when computer based and paper-and-pencil test scores were compared (Alexander, Bartlett, Truell & Ouwenga, 2001).

Another study on the comparability between computerized tests and paper based tests was conducted by Cole, MacIsaac and Cole (2001). Based on data collected from 1,313 students, the researchers looked at the differences between paper-based and web-based administrations of Force Concept Inventory (FCI), concluding that the web-based administration of the test has the same effect as the paper-based administration (Cole et al., 2001). Similar findings were achieved by Choi, Kim, and Boo (2003) in a study aimed at addressing the issue of the comparability between paper based language tests (PBLT) and computer based language tests (CBLT), which found that PBLT and CBLT were highly comparable both in content and in linguistic features.

Current Technology That Supports Web-based Testing

While the design and development of web-based testing used to rely on hand–coding, current technology allows designers to use HTML or JavaScript generating tools

such as Macromedia DreamWeaver and Macromedia Flash and work in a WYSIWYG (What you see is what you get) environment (Macromedia, 2005). With these tools, test developers only need to type in the questions and answers in specified text boxes and HTML, PHP or JavaScripts code will be generated accordingly (Li, 2006). Multimedia contents such as graphics, sounds and videos can be easily incorporated into HTML (Li, 2006) for test designing. The currently existing database technology is sophisticated, and it is easy to set up a database which is capable of collecting test result from test takers (Wang, 2003).

The delivery of web-based tests is supported by fast-advancing computer hardware technology, drastically improved internet connection speed, and sophisticated server side technology (Wang, 2003). The hardware abilities of modern computers offer more than what is needed to deliver and execute a web-based test, even when the test contains multimedia contents (Roever, 2001). Meanwhile, the connection speed of modern computer networks has reached 100Mb or even 1000Mb, and new network communication protocols have been developed to accommodate the high demand on network data transmission (Pimentel, 1999).

Studies on Audio Technology

The Nature and Characteristics of Sound

Sound is caused by the vibration of materials that is sensed by the human ear (Nisbett, 1995). Sound has such characteristics as frequency and loudness (Alkin, 1998). Frequency is the number of times a sound source vibrates in one second (Luther, 1997). The frequency of a sound is measured in Hertz (Hz) or kilohertz

(kHz) (Liang, 2001). Higher frequencies produce higher pitches for the human ear (Pohlmann, 1989). Loudness is the intensity of the sound judged subjectively by the human ear (Nisbett, 1995).

Analog Audio vs. Digital Audio

Analog audio refers to sound signals that represent electronically the waveforms of a sound (Luther, 1997). Analog audio is a "pictorial representation" of sound voltages (Clifford, 1992, p.148). Analog audio has such weaknesses as noise and interferences, distortion, and instability (Pohlmann, 1989). In digital audio, a sound signal is represented by a series of digits that a computer can recognize as certain values (Luther, 1997). To produce digital audio, analog audio signals are sampled at regular intervals by an analog-to-digital converter and then translated into binary digits (Clifford, 1992). Digital audio has such advantages as high fidelity, easy to store, easy to process, error-free transmission and low cost (Luther, 1997). Digital audio has two main weaknesses: digital systems can be slow in operation and digital devices can be expensive (Rosen & Howell, 1998).

Sample Rate and Sample Depth of Digital

The sample rate is the number of times the voltage of the analog signal is measured in each second (Pohlmann, 1989). The commonly used sample rates are 44 kHz for CD quality audio, 22 kHz for radio quality audio, and 11 kHz for phone quality audio (Luther, 1997). In order to faithfully digitize analog audio, the sample rate must be at least twice as high as the highest frequency in the analog audio (Pohlmann, 1989). Sample depth of digital audio is the number of bits used to

record each sample (Luther, 1997). Two of the commonly used sample depths are 16 bit and 8 bit (Huber & Runstein, 2005).

Human Perception of Sound

According to Watkinson (1994), the human ear is most sensitive between 2 kHz and 5 kHz, and most listeners cannot tell if the upper most frequency of a sound is 20 kHz or 16 kHz. Since the sampling theorem (also known as the Nyquist's Theorem) dictates that the sampling rate must be at least twice the input frequency, sampling rate of at least 40 kHz is needed to reproduce digital audio that covers the full human hearing range (Watkinson, 1994) . On the other hand, 11 kHz of sample rate is suitable only for 5 kHz sound, and is considered adequate for recording human speech for ordinary communication purpose (Liang, 2001).

Digital Audio Compression

Today's advanced audio compression technology is another factor that drives the development of web-based tests, especially web-based language tests. MPEG-1 Audio Layer 3 (MP3), for example, is a popular digital audio encoding algorithm (Brandenburg & Popp, 2000). MP3 is a lossy compression format that uses Huffman codes to reduce the size of an audio file while maintaining the original audio quality (Pan, 1993). RealAudio is another audio format that may be used for the purpose of web-based language test. This audio uses a variety of codecs ensuring that users receive appropriate encoded content regardless of what bandwidth is available to him (Beggs & Thede, 2001). It is often used as a streaming audio format that can be played while being downloaded (Beggs & Thede, 2001).

One of the audio formats that could be used for web-based language listening comprehension test is Windows Media Audio (WMA). Developed by Microsoft, WMA maintains good audio quality at great compression ratios (Stockdale, 2002). At a typical CD sampling rate (44 kHz), WMA can achieve CD-like sound quality at a bit rate as low as 48 kbps (Ribas-Corbera, 2003). This indicates that even at such a low bit rate a frequency of up to 20 kHz is still maintained. Windows Media Audio files can be downloaded as discrete files, or they can be streamed from a media server (Ribas-Corbera, 2003).

Computer-Assisted Language Learning and Language Testing

Computer-assisted language learning (CALL) began in the 1960s (Chapelle, 2001). Computer applications have been used to teach vocabulary, grammar, and pronunciation, and have functioned as spell checkers and electronic workbooks (Reover, 2001). Computers have been used as a source of motivation for language learners and, given the interactive and multimedia capabilities, as an attractive teaching tool for language teachers (Blake, 2000; Chun & Brandl, 1992). Computer mediated communication (CMC) allows students from different countries to communicate in and practice a common target language (Beatty, 2003), or to collaborate in joint exploration of culture (Cummins & Sayers, 1995). Recently, Web-based desktop videoconferencing has been used for language teaching, especially the teaching of EFL (Xiao, 2007).

Computerized language tests have been used for a wide range of purposes such as determining to what degree a student has met the objectives of a language course and to what extent an individual has mastered the listening comprehension or the grammar of

the target language (Dunkel, 1999). Compared to traditional paper-and-pencil based language tests, computer aided language tests allow immediate result reporting and relatively frequent evaluation (Larson & Madsen, 1985). Computerized language tests are considered user-friendlier than paper based tests and, therefore, they may allow testing practices to be directly incorporated into the teaching and learning process (Alderson, 2000). Roever (2001) suggests that it is fairly easy to implement grammar and vocabulary tests using a variety of item types such as multiple choice, filling blanks and discourse completion, and it is easy to test writing using large text areas and reading comprehension using frames. Alderson (2000), however, argues that computer based language tests are limited to only a few item types and that item types other than multiple-choice may not be appropriate for computer based delivery. Some researchers (Choi et al., 2003) suggest that computer based language tests are comparable to traditional paper-and-pencil based language tests in terms of content and linguistic features. Some other researchers, however, are concerned that an computer-adaptive test may be less valid than a paper and pencil based test or a test conducted through face to face interview (Chaplle, 2001).

Technology in Computerized Language Listening Assessment

Computer technology has been used in many ways to test students' ability to listen and understand a target language. Some computer assisted listening tests present only aural input such as mini-dialogs or short monologs (Dunkel, 1999) while other tests may contain such test items as sentence completion or dictation (Coniam, 1996). Some computer based language listening comprehension tests are designed with non-linguistic

elements such as diagrams or pictures displayed on the monitor with question items (Chapalle, 2001; Choi et al., 2003). However, Choi et al. (2003) suggest that a listening comprehension test that is designed without the use of visual elements may result in better test validity. Brusilovisky and Miller (1999) also suggest that misleading pictures may serve as distracters rather than creating contextualization, and different test takers may understand a visual element differently.

To the knowledge of the author of this paper, no previous study has focused on the issue of audio in English listening comprehension tests. However, some researcher did express the concern over the quality of the aural input, suggesting that the testing environment for a listening test should be as free from ambient noise as possible and that test takers may need to wear headphones to keep environmental noise away (Dunkel, 1999).

Summery of Literature Review

Web-based testing is a simple concept built on complex theories and technologies. To truly understand the characteristics of web-based testing, one must look into not only the field of language testing but also the tools and methods that enable tests to be conducted online and through a distance. This chapter has reviewed the literature on computerized testing in general and web-based testing in particular. The review also covered the studies on the characteristics of sound, the characteristics of analog and digital audio, as well as the technology for recoding, compressing, storing and delivering audio in digital format. In addition, this chapter has reviewed the studies on computer assisted language learning and language testing. To the knowledge of the author of this

paper, there has been no previous study that focused on the issue of digital audio quality in the context of web-based English listening comprehension tests. The review has provided a foundation upon which the current study has been built.

CHAPTER THREE: METHODOLOGY

Digital audio quality is a crucial factor for test validity and is a complicated issue in the context of a web-based English listening test. Previous research in the field of digital audio has found that high sample rate and sample depth digital audio, such as the audio recorded at 44 kHz and 16 bits, can achieve a very close representation of its analog source (Liang, 2001) while low sample rate audio, such as the audio recorded at 11 kHz and 8 bits, although not as good as high sample rate audio, can still sound good for human ears (Currier, 1996). For the purpose of language testing, these findings may not apply, since language students, who are therefore language test takers, may not process the target language in the same way as a native speaker does (Conrad, 1985; Bacon, 1992). In this study, the following research questions have been addressed:

1. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample rates (44 kHz, 22 kHz and 11 kHz)?
2. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample depths (16 bit and 8 bit)?
3. Do sample rate and sample depth interact in affecting students' performance in web-based English listening comprehension tests?

These research questions were answered by comparing test results of six groups of participants taking a LAN delivered web-based English listening comprehension test with digital audio of different sample rate and sample depth combinations. This chapter

describes the approaches for finding answers to these questions. The following components are provided in this chapter: 1) the sample used for the study, 2) a description of the instrumentation, 3) identification of independent and dependent variables, and 4) a statement of the statistical hypothesis.

Population and Sample

The population of this study consists of English learners who are non-English majors in an EFL setting in eastern China. In this setting, the curricula of English teaching are strongly based on the traditionalist view of education and, as a result, the language classes are mostly teacher centered and lecture oriented, and grammar translation and audiolingualism play a dominant role (Liu & Shi, 2007). Students spend most of the time learning grammatical rules and memorizing words and phrases, or practicing translation between English and Chinese in written forms. The teaching of speaking and listening is based on the teaching of phonetic sounds. For speaking, students are first taught the individual sounds of the English language, and then they are taught to 'build' words using these sounds, and then 'build' sentences using words. For listening, students are required to distinguish every sound in a word, and every word in a sentence. According to Conrad (1985) and Bacon (1992), the process of listening comprehension may be different between language learners and native language speakers. This may well be true in the EFL environment such in China. For English learners in such an environment, every phoneme of the language may be important for the listening comprehension of the language, and any imperfect rendition of an individual sound can affect the students' performance in a listening comprehension test.

The participants of this study are first year students from a university in eastern China. This university currently has about 4,800 students who, although not majoring in English, are required to learn English as a foreign language and pass the College English Test (CET) Band 4 to be awarded a bachelor degree or Band 6 to be awarded a master's degree. In a social background where knowing a Foreign language, especially English, is highly popular and beneficial, these students are eager to take any available opportunity to improve their English skills and are enthusiastic at participating in research activities for foreign language education. The curricula of the university are strongly based on the traditionalist view of education and the English classes are mostly teacher centered and lecture oriented, and grammar translation and audiolingualism are the main teaching methods. The university has 14 modern computer labs, three of which were chosen to complete this study.

The ages of the participants range from 19 to 22. About 52% of the participants are male and 48% are female. The participants, although non-English majors, have taken English classes ever since they entered the university and all of them are making preparation to take the CET Band 4. Educated in the cultural background of China, these students are well known for being cooperative with their professors and school authorities and are willing to participate in research projects related to their studies.

According to Lipsey (1990), the sample size is determined by the desired statistical power, significance level, and the effect size. While most social studies adopt the power of .80 (Murphy & Myors, 1998), this study aimed at a power of .90 considering that web-based language tests are typically aimed at large scale and often

high stakes assessment, and that any flaw in design and administration will have a wide and deep negative impact. Regarding the effect size, there is a lack of research on the difference of effects between different audio qualities in web-based English listening comprehension tests. Based on the findings of research in similar fields such as the human hearing (Xu, Thompson & Pfingst, 2005), this study aimed at a medium effect size between the three sample rates (44 kHz, 22 kHz and 11 kHz) and two sample depths (16 bit and 8 bit). Based on the need to generate adequate statistical power for an anticipated medium effect size as well as a consultation with the software SamplePower, a total sample size of 576 for the 2 x 3 ANOVA was proposed by the researcher. However, due to the fact that each of the computer labs suitable for data collection for this study could only seat a maximum of 64 students, only 360 students were chosen to participate in the study. Those 360 students were randomly chosen from around 2,400 students in the first grade and were then randomly assigned to six different groups, with each group having 60. Some of the students who had been chosen were absent from the test, which made the actual number of participants to be 335. This issue is further discussed in the limitations and weaknesses section of this dissertation. All the hypotheses were tested at the .05 level of significance.

In order to achieve true random selection, a computer program on the GraphPad Software website (http://www.graphpad.com/quickcalcs/randomN1.cfm) was used to generate 360 random numbers between 0001 and 2400. The students with ID cards that have the random numbers as the last four digits were selected. These 360 students were then randomly divided into 6 groups. The chosen students were notified that they would

be given an English listening comprehension test for the purpose of a research. Information about the difference in the audio that they would hear was deliberately kept secret from these students.

Instrument Used for the Study

The main instrument used for this study is a web-based English listening comprehension test derived from the listening comprehension section of disclosed College English Test of 1995. The test items were not changed in any way except that they were converted to a computerized format. As mentioned in Chapter One, the test as the instrument of this study was aimed for delivery on the Internet using web-based technology. However, due to the restrictions of the Internet environment at the university in eastern China, this web-based English listening comprehension test was delivered in a LAN environment where common Internet services were installed. This test is referred to as a web-based English listening comprehension test throughout this paper in order to reflect its essence of web-based technology.

The College English Test, better known as CET, is a national English test for college and university students in the People's Republic of China. This test is administered twice a year nationally, in June and December respectively. The CET is primarily for college students whose majors are not English. It has two levels, Band 4 and Band 6. CET is administered by the National Educational Bureau of China, and is known to have high item reliability (Yang, 2006). Many colleges and universities in China require the CET certificate as a necessity for a bachelor diploma, and employers in China prefer applicants with a CET certificate. A free sample of the CET Band 4 test has been

provided by the National College English Testing Committee (NCETC) and is available on the official CET website (www.en.cet.edu.cn). The listening section of this sample test has been attached to this paper as Appendix A. The items used in the instrument for this study were taken from a disclosed CET Band 4 test from 1995. Disclosed test items are properties of the university and are only available to teachers who specialize in training the students to pass the CET test. There has been some year to year change in the format of CET and most teachers prefer to use the latest versions of disclosed CET in their training. Therefore, it is quite unlikely the participants had seen these out-of-date test items prior to the study.

The English listening comprehension test as the instrument of this study consists of two sections, which are described as follows:

Section One. In Section One, test takers hear short conversations between two speakers. At the end of each conversation, a third person asks a question about what was said. Test takers hear each conversation and question about the conversation only once, therefore, they must listen carefully to understand what each speaker says. After they hear a conversation and a question, they must read the four possible answers and decide which one is the best answer to the question they have heard.

Section Two. In Section Two of the test, test takers hear some talks or passages. After each short talk or passage, test takers are asked some questions. Both the talks and passages and the questions about them are spoken just one time. After they hear a question, test takers must read the four possible answers and decide which one is the best answer to the question they have heard.

The scores for each of these two sections were collected along with the total scores. The section scores were not analyzed for this study. Instead, the section scores would be used in future studies.

Once the items for both the two sections were selected, the corresponding audio recording on the original CET cassette tape was digitized using six different sample rate and sample depth combinations, namely 44 kHz/16 bit, 44 kHz/8 bit, 22 kHz/16 bit, 22 kHz/8 bit, 11 kHz/16 bit and 11 kHz/8 bit. All the audio files were converted to Windows Media Audio using variable bit rate (VBR) at highest quality setting.

Six different versions of a web-based listening comprehension test were composed, each using a different one of the digitized audio files (audio with a different sample rate/sample depth combination). These versions of test were designed in a non-adaptive style, with each question contained within a single web page that was automatically displayed when each recording is heard. The time interval between the recordings for any two test items was about 20 seconds, same as the interval between two recordings on the original CET tape. The test was created in a proprietary language testing system bundled with the three computer labs used for this study. The result of the test was collected instantaneously and saved as a Microsoft Excel spreadsheet for statistical analysis. A copy of the test items used for the instrument of this study is provided as Appendix B.

As discussed in the previous chapter, web-based test can be administered either through the Internet or through a LAN. Ideally, data collection should be done in the Internet environment to emulate a large scale test over a distance. However, there existed

a possibility that the bandwidth or settings of the Internet environment at the site of data collection would not allow smooth streaming of the audio files, especially the high quality audio files, from the test server to the hundreds of client computers at the same time. In order to determine whether the Internet or a LAN should be the environment in which data would be collected, a pilot study was conducted three weeks prior to the actual data collection process to find out whether the test as the instrument would function as intended. The pilot study used the listening comprehension test for the actual data collection as well as a short interview. The result of the pilot study indicated that the Internet connection at the university in eastern China did not allow smooth transfer of the highest quality audio (44 kHz / 16it) for this study. The researcher decided that it was more appropriate to collect data in a LAN environment than the Internet environment. Details concerning the pilot study are discussed in Chapter Four.

Equipment Used for the Study

The equipment used for this study includes a high fidelity audio cassette tape player, an audio digitizing computer, a test server, and three computer labs in a LAN environment. The equipment is described in detail as follows.

The Audio Cassette Tape Player

The original CET recordings selected for the purpose of this study are carried analogously on wide frequency range cassette tapes. These tapes have been well preserved in a dry and ventilated storage house and have been fast rewound at least twice each year to prevent the deterioration of signal strength. In order to exploit the full frequency range and the full dynamic range of the original CET recordings, a JVCTD-

W718 cassette deck was chosen as the player for the digitizing process. This player approximately matches reel-to-reel decks with frequency response of 20 Hz to 17 kHz range and 58 dB signal-to-noise ratio. A professional audio cable was used for signal transmission from the audio tape player to the digitizer (the computer sound card in this case). This cable has been designed to provide maximum isolation from electromagnetic and radio frequency interference with shielded construction and full-coverage aluminum foil shield.

The Audio Digitizing Computer

The audio digitizing computer for this study was a Lenovo M5100G computer with an AMD SEMPRON 3000+ processor. It had 2 x 512 megabytes of memory and an 80-gigabyte hard drive. It was equipped with a sound card which is capable of recording sound at a sample frequency of as high as 96 kHz and a sample depth as deep as 24 bits. The operating system running on this computer was Windows XP. Cool Edit Pro 2.0 was installed on this computer, which has been used to capture audio and save it on the hard drive. A USB 2.0 external hard drive was used to transfer finalized audio files to the test server.

The Test Server

For this study, the six versions of web-based English listening comprehension test were hosted on a customized HP computer as the LAN server that has a 3. 2 GHz Pantium D processor and 4 gigabytes random access memory. It had Windows 2003 Server as its operating system and IIS 6.0 as its web server. PHP, MySQL and Windows Media Services were installed to run data base driven web pages. The hard drives were

set to a Redundant Array of Independent Disks (RAID) configuration to minimize the chance of failure during the test.

The Computer Labs

The English listening comprehension test for data collection was conducted in two sessions in three computer labs owned by a university in eastern China. These labs have been designed for multiple purposes and can function both as ordinary computer labs and as language teaching and learning labs. Except for the difference in location, these labs are identical in design and configuration. Each lab contains 64 HP XW4300 desktop computers. Each of these computers is equipped with a 3.2 GHz Pentium D processor, 1 GB of DDR randomly access memory, 3 x 80GB SATA hard drive, and an integrated ALC260 sound card. Plugged into each computer is a pair of high definition headphones designed especially for the rendition of high quality digital sound. These headphones are capable of responding to frequencies up to 22 kHz, enough to cover the frequency range of the audio used in this research. Besides, these headphones also feature a sensitivity of 90 dB with a wide dynamic range suitable not only for playing back the human voice but also for CD quality music. All the computers are connected through high speed Ethernet cables and network cards. In order to minimize the effect of screen size, screen resolution and display rates on test performance, the 19 inch LCD monitors in these labs were all set to 1280 x 1024 resolution, 32 bit color depth and 75 Hz of refresh rate. The operating system on all computers was fully updated Windows XP Service Pack 2, with Internet Explorer 7 as the web browser. Windows Media Player 10

was installed along with all current codecs. All computers were checked and maintained shortly before the test in order to achieve maximum performance.

Data Collection Methods

Data collection was completed in several separate steps. An English listening comprehension test was first composed on paper using items from disclosed CET test of China. The items used in the instrument for this study were taken from a disclosed CET test of 1995. Permission from the university authority was obtained in advance to avoid any copyright issues and a copy of the permission has been attached to this paper as Appendix C. There were a total of 20 questions in this test. Questions 1 to 10 were for Section One, and questions 11 to 20 were for Section Two. Once the test items were selected, the original recording of these items were digitized using the following combination of sample rate and sample depth: 44 kHz/16 bit, 44 kHz/8 bit, 22 kHz/16 bit, 22 kHz/8 bit, 11 kHz/16 bit and 11 kHz/8 bit. The audio was recorded in mono mode to reduce file size. The audio files were compressed to WMA using variable bit rate (VBR) at highest quality setting without changing the original sample rate and sample depth.

Six different versions of the web-based English listening comprehension test were developed using the bundled testing system in the computer labs. These six versions of the test differed only in the quality of the digital audio files embedded in the HTML codes, each version containing only one of the six audio files. Audios of different quality were not mixed in any single version of the test. Participants had been randomly divided into six groups and those in the same group took the same version of the test. To be more specific, group one took the test with digital audio of 44 kHz/16 bit, group two took the

test with digital audio of 44 kHz/8 bit, group three took the test with digital audio of 22 kHz/16 bit, group four took the test with digital audio of 22 kHz/8 bit, group five took the test with digital audio of 11 kHz/16 bit, group six took the test with digital audio of 11 kHz/8 bit. A research assistant was hired to organize and supervise the test in the three computer labs (See Appendix D for contact information of the researcher and the research assistant). More details on the data collection procedures are discussed in Chapter Four.

Operational Definitions of the Variables

The data for this study was analyzed using 2 x 3 ANOVA. Based on the research hypotheses, the following independent and dependent variables were defined:

Independent Variables

1. Sample rate. This independent variable has three levels—44 kHz, 22 kHz and 11 kHz. The contents of the audio in the web-based English listening comprehension tests are identical, but they were delivered in three different levels of sample rate.
2. Sample depth. This independent variable has two levels—16 bit and 8 bit. The contents of the audio in the web-based English listening comprehension tests are identical, but they were delivered in two different levels of sample depth.

Dependent Variables

This study focused on one dependent variable—the test score. In the 2 x 3 analysis of variance the test scores of the six groups were compared in order to find any significant difference that might exist.

Methods for Data Analysis

This study focused on investigating the effects of digital audio quality in a web-based English listening comprehension test delivered in a LAN environment. Based on collected data, descriptive statistics for independent and dependent variables were provided and illustrated. All the data analysis procedures were implemented by using the Statistical Package for Social Science (SPSS) Version 14 program. The main effects of both independent variables (sample rate and sample depth) were examined and analyzed as well as the interaction between them. The significance level of $\alpha = .05$ was adopted to test all the hypotheses in this study. More details of the data analysis procedures are revealed in Chapter Four.

Methods for Answering Research Questions

In order to answer the research questions of whether or not differences among three different sample rates and between two different sample depths, as well as the interaction effects, exist on web-based English listening comprehension test scores, the following null hypotheses were proposed and tested:

HO_1: There is no significant difference of the test scores among test with 44 kHz digital audio, test with 22 kHz audio and test with 11 kHz digital audio.

HO_2: There is no significant difference of the test scores between test with 16 bit digital audio and test with 8 bit audio.

HO_3: There is no significant relationship between sample rate and sample depth in terms of the test scores.

The submitted test results from the six groups were computed and described. Besides descriptive statistics, this study used two by three Factorial Analysis of Variance (2 × 3 ANOVA) to investigate the effect of the two factors, sample rate and sample depth, on test results as the dependent variable.

Post hoc analysis was implemented to further compare the three sample rates in pairs. A test for interaction between the independent variables was conducted to provide complete information of their effects on the web-based English listening comprehension test.

CHAPTER FOUR: FINDINGS AND DISCUSSIONS

The purpose of this study was to examine the effects of different sample rates and sample depths in digital audio on students' performance in web-based English listening comprehension tests. Specifically, it explored how differently students perform on an English listening comprehension test using audio recordings with sample rates of 44 kHz, 22 kHz and 11 kHz and sample depths of 16 bit and 8 bit. Data were collected from the students' scores in an English listening comprehension test given with digital audio of different combinations of sample rate and sample depth. This chapter describes in detail the data collection procedure as well as the analysis of the data. This chapter also describes the assumption testing of the data and the pilot study as part of the data collection procedure. Most importantly, this chapter presents the findings based on the data analysis and discusses the implications of these findings. The Statistical Package for Social Sciences (SPSS) for Windows version 14 was used to analyze the data. Two-way ANOVA was utilized to answer the three previously defined research questions.

The Data Collection Procedure

Data collection was accomplished through an English listening comprehension test given to six groups of students using six different combinations of sample rate and sample depth. As part of the data collection procedure, a pilot test was conducted prior to formally giving the English listening test to the six groups of participants. Based on the findings of the pilot test, the researcher adopted a local testing method, by which the test was hosted on a server for a local area network.

The Pilot Study

Since the purpose of the pilot study was to find out whether there exist such issues as Internet congestion, IP address filtering and unfavorable firewall settings, it was conducted with only the audio being streamed from a server while the test items were printed on paper. A single page website was used as the container for a link to each audio file. The instrument for the pilot study was the same test as the one used in the actual data collection. As already described in the previous chapter, the test consisted of two sections, with each section having 10 questions. The first section tested students' ability to understand short conversations and the second section tested students' ability to understand longer talks or passages. Also as described in the previous chapter, the audio recording for the test was digitized using six different sample rate and sample depth combinations, namely 44 kHz/16 bit, 44 kHz/8 bit, 22 kHz/16 bit, 22 kHz/8 bit, 11 kHz/16 bit and 11 kHz/8 bit. All the audio files were converted to Windows Media Audio format.

The pilot study was conducted with 12 second-year English majors from the same university at which the actual English listening comprehension test was conducted for data collection. The reason that English majors have been used instead of non-English majors for the pilot study was mainly an effort to maintain independence of observation for the real data collection test. In almost all universities in China, including the one at which data collection has been conducted, non-English majors from different departments or different schools are often grouped together to take English courses in the same class. In other words, any non-English major can be classmate with any other non

English major even though they are from totally different departments or schools. Conducting the pilot study with any of the non-English majors would risk exposing the test items to the actual participants, which may cause a serious flaw in the design of this research since one of the assumptions for 2 x 3 ANOVA is independence of observation (Green & Salkind, 2005). Another advantage of using English majors for the pilot test is that these students are far more advanced English learners than the non-English majors. Given an ideal testing environment, e. g., no bad audio quality or transmission delay of the audio, these students are capable of achieving high or even full scores on such a relatively easy test. This helps the researcher to concentrate on the most important point of interest, which is the quality of the Internet connection.

The 12 students were divided into 6 groups, each group having two. Each group was asked to take the test with an audio recording of different sample rate and sample depth combination. The six groups were tested sequentially in one of the three computer labs that were later used for the real test. The total scores as well as the section scores of each student were examined. As shown in Table 4.1, the two participants in the 44 kHz /16 bit group scored the lowest despite the fact that the audio they heard during the test was of the highest quality. A closer look at the scores of each section indicates that the students who took the test with the highest sample rate and sample depth actually scored quite low on the second section of the test, which consists of longer passages and questions.

Table 4.1

Results of Pilot Test Participated in by Twelve English Majors

Group	*Student ID*	*Section One*	*Section Two*	*Total*
Group 1 (44k/16b)	1	9	5	14
	2	9	6	15
Group 2 (44k/8b)	1	10	9	19
	2	10	9	19
Group 3 (22k/16b)	1	10	9	19
	2	10	8	18
Group 4 (22k/8b)	1	10	10	20
	2	10	8	18
Group 5 (11k/16b)	1	9	10	19
	2	9	7	16
Group 6 (11k/8b)	1	10	8	18
	2	9	9	18

Note: The test has a total of 20 points, with 10 points for each of the two sections.

An interview with the participants of this pilot test was conducted through the telephone or through meeting with the research assistant. The key questions asked during the interview were 1) How would you describe the quality of the audio that you heard during the test, clear, unclear or noisy? 2) In what way do you think the quality of the audio has affected your performance in the test? All the participants who were interviewed claimed that the test was not difficult at all but some said that the recording started slow and had break-ups, which had made it difficult for them to answer some of the questions. The participants who complained about the audio were those who were tested with 44 kHz and 16 bit, while the participants from all other groups stated that the recordings sounded clear. The researcher concluded that the audio recording with high sample rate and sample depth had problems being transmitted through the Internet, most probably due to a relatively low bandwidth available at the time and location that the pilot study was being conducted. If the actual data collection were to be conducted through the Internet at the Chinese university, it was likely that the same problem would occur and that the scores of the high quality audio groups would not truly reflect the effects of the corresponding sample rates and sample depths. It would only reflect the effect of the Internet connection quality, which was not what this study was aimed to address. Therefore, the researcher decided to conduct the data collection procedure in a LAN environment rather than over the Internet.

The Actual Test for Data Collection

As discussed in the previous chapter, this study used a web-based English listening comprehension test as the instrument for data collection. The items in the test

were taken from a collection of disclosed CET test items of 1995. There were a total of 20 questions in this test. Questions 1 to 10 were for the first section of the test, which tested students' ability to understand short conversations. Questions 11 to 20 were for the second section, which tested students' ability to understand longer talks or passages. The original recording of the items was digitized using the following combinations of sample rate and sample depth: 44 kHz/16 bit, 44 kHz/8 bit, 22 kHz/16 bit, 22 kHz/8 bit, 11 kHz/16 bit and 11 kHz/8 bit. All the audio files were compressed to WMA. The test was given to six groups of participants, with each group hearing the digital audio made with one of the six sample rate and sample depth combinations. For the purpose of effectively recording test scores, the six groups were labeled Group 1, Group 2, Group 3, Group 4, Group 5 and Group 6, and the computer labs were labeled Lab 1, Lab 2 and Lab 3. Group 1 was tested with 44 kHz/16 bit audio; Group 2 was tested with 44 kHz/8 bit audio; Group 3 was tested with 22 kHz/16 bit audio; Group 4 was tested with 22 kHz/8 bit audio; Group 5 was tested with 11 kHz/16 bit audio; Group 6 was tested with 11 kHz/ 8 bit audio.

The test took place in three computer labs that had identical hardware and software. Two test sessions were arranged for the test with one session immediately following the other. Group 1, Group 2 and Group 3 were tested during the first session in Lab 1, Lab 2 and Lab 3 respectively. Group 4, Group 5 and Group 6 were tested in the second session also in Lab 1, Lab 2 and Lab 3 respectively. Participants did not have any time to exchange information between the two sessions. The possible answers for each question were displayed on the monitor within one single page along with the four

buttons marked A, B, C and D. Each time after a choice was made a confirmation dialog box appeared to ask the test taker either to confirm his/her choice or cancel it to make a different choice. Auto feedback function of the test was disabled to help the participants to concentrate on listening and choosing answers. This did not change the test since the auto feedback feature was typically for self-paced exercises only. The test was automatically graded and the scores of each group were stored in Microsoft Excel spreadsheet for analysis in SPSS 14.

Data Analysis and Result

Description of Data

The data collected from the web-based English listening comprehension test was first examined using SPSS 14. There were a total number of 335 values for the dependent variable, the test score. The total mean score across sample rate and sample depth is 14.69, with a standard deviation of 2.69. The means and standard deviations on the dependent variable, the test score, were different among the groups tested with the three different sample rates and the two different sample depths. The groups tested with 22 kHz of sample rate had a total mean score of 15.19 with a standard deviation of 2.58, which was higher than the means of both the groups tested with 44 kHz and the groups tested with 11 kHz. The groups tested with 8 bit of sample depth achieved overall mean score of 14.70 with a standard deviation of 2.55, which was higher than the means of the groups tested with 16 bit of sample depth. The group tested with a combination of 22 kHz and 8 bit achieved the highest mean score, which was 15.54 with a standard deviation of 2.22. Participants in the group that was tested with 11 kHz and 8 bit achieved the lowest

scores with a mean of 13.54 and a standard deviation of 2.60. The highest score was 20, which appeared mostly in the 44 kHz/16 bit group and the 22 kHz/8 bit group. The lowest score was 6, which appeared in the 22 kHz/16 bit group. The numbers of participants in the six groups are different but are roughly equal. No missing values were found (Table 4.2).

Table 4.2

Descriptive Statistics of Participants' Test Scores

Sample Rate	*Sample Depth*	*Mean*	*Std. Deviation*	*N*
44 KHz	16 bit	15.1250	2.85442	56
	8 bit	15.0339	2.41380	59
	Total	15.0783	2.62628	115
22 KHz	16 bit	14.8571	2.85675	56
	8 bit	15.5385	2.21812	52
	Total	15.1852	2.58011	108
11 KHz	16 bit	14.0862	2.70337	58
	8 bit	13.5370	2.59690	54
	Total	13.8214	2.65498	112
Total	16 bit	14.6824	2.82303	170
	8 bit	14.7030	2.54526	165
	Total	14.6925	2.68581	335

Assumption Testing

According to Green and Salkind (2005), three major assumptions must be fulfilled when using two-way ANOVA: 1) independence of observations, 2) normal distribution of observations, and 3) homogeneity of variance. Before conducting 2 x 3 ANOVA with the data collected from the 335 participants, the three assumptions were examined to make sure that they were all satisfied.

The assumption of independence of observations states that each observation should be independent from other observations. This assumption is usually satisfied if random sampling and random assignment strategies are used (Stevens, 2001). For this study, the participants were both randomly selected and randomly assigned. To select the participants, a computer program was used to generate 360 random numbers between 0001 and 2400. The students with ID cards that have these random numbers as the last four digits were selected. Therefore, the process of selecting participants did not involve any human bias. After the 360 participants were selected, names were randomly picked from them to be assigned to six groups.

The participants took the test in a proctored environment and have answered the test questions independently. Even though the six groups took the test in two different sessions, the two sessions of testing had been arranged one immediately following the other and therefore participants who took the test in the first session did not have time to exchange any information about the test with participants who took the test in the second session. Moreover, the participants were highly disciplined students and there is no reason to believe that any of them had cheated on the test, especially after they had been

informed that the purpose of the test was for research only. Based on the above analysis, it is reasonable to say that the assumption of independence of observation was valid.

The assumption of normality was tested with both graphical methods and a theory-driven test. The histogram of the data distribution was first examined to visually check the normality of the distributions. As shown in Figure 4.1, the distribution of the overall test scores appeared to slightly deviate from normal.

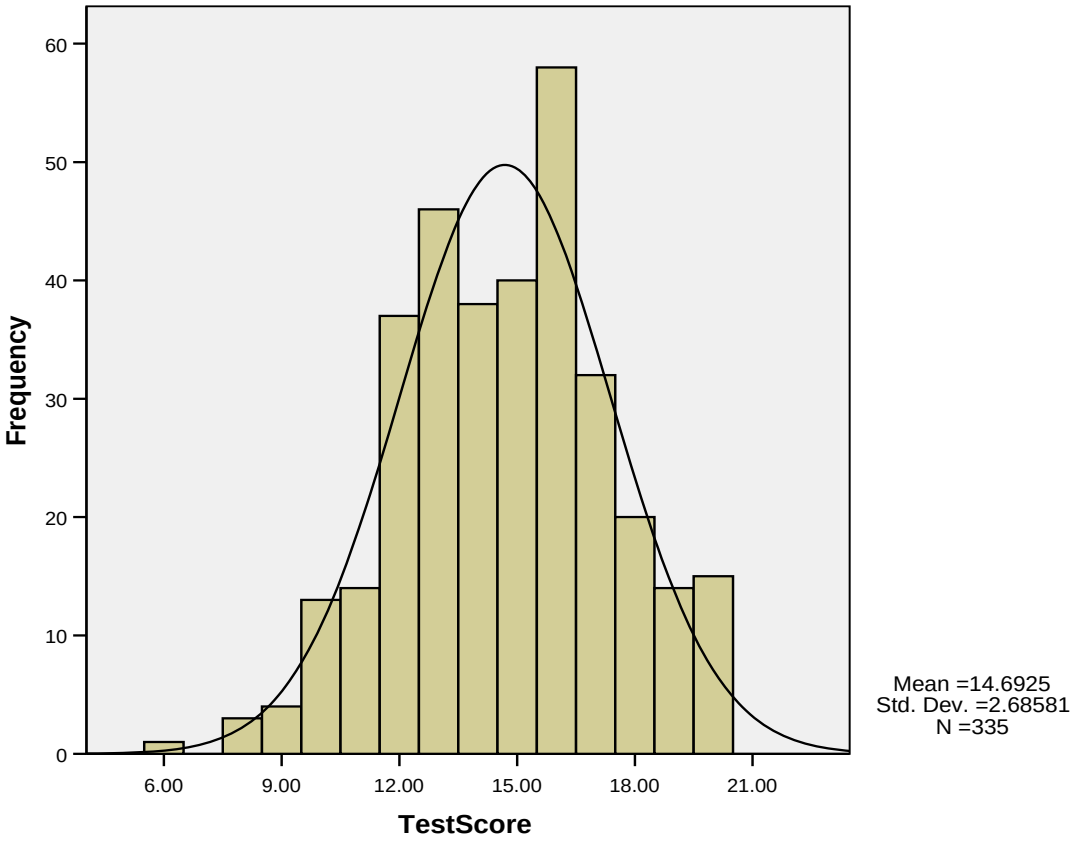


Figure 4.1. Histogram of the Distribution of Participants' Test Scores. Superimposed Curves Shows Normal Distribution.

The boxplots for the six cells also showed that the distributions slightly deviated from normal, with the rectangles in each plot slightly off the middle of the range and the median line noticeably off the center. Two outliers were detected, one in the 22 kHz /16 bit group and the other in the 11 kHz / 8 bit group (See Figure 4. 2).

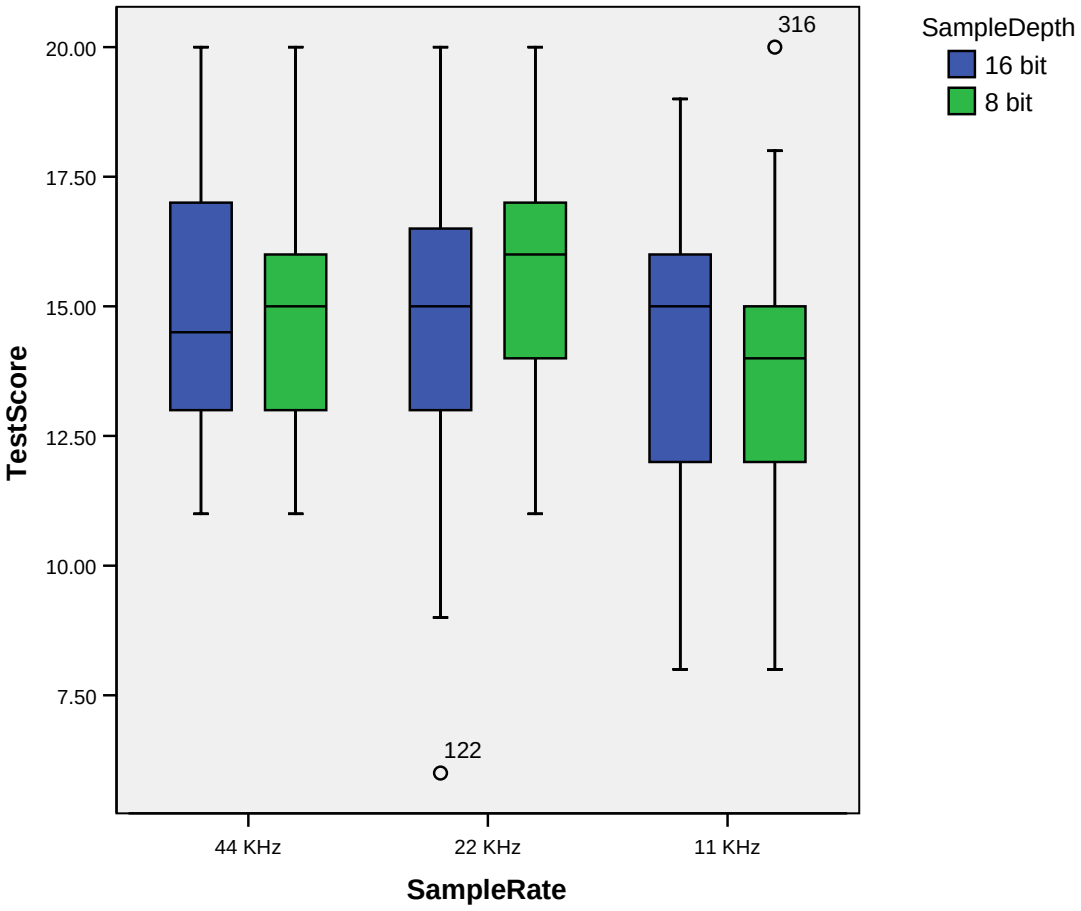


Figure 4.2. Boxplots of the Distribution of Participants' Scores in Each Cell for the Two-way ANOVA Design.

In addition, this study examined the normality assumption using Kolmogorov-Smirnov test. When tested at alpha = .05 level, the Kolmogorov-Smirnov test was significant, which indicated that the distribution of the tests scores was not normal (See Table 4. 3).

Table 4.3

Kolmogorov-Smirnov Test for Normality Assumption of Test Scores

N		*335*
Normal Parameters(a,b)	Mean	14.6925
	Std. Deviation	2.68581
Most Extreme Differences	Absolute	.102
	Positive	.088
	Negative	-.102
Kolmogorov-Smirnov Z		1.862
Asymp. Sig. (2-tailed)		.002*

Note: The adopted significant level is .05.

Stevens (2001) argues that two-way ANOVA is robust to the of normality assumption if the distribution in all cells skew to the same direction. In the this study, the histograms of test scores for each level of sample rate and each level of sample depth skewed to the same direction—to the right (See Figure 4.3).

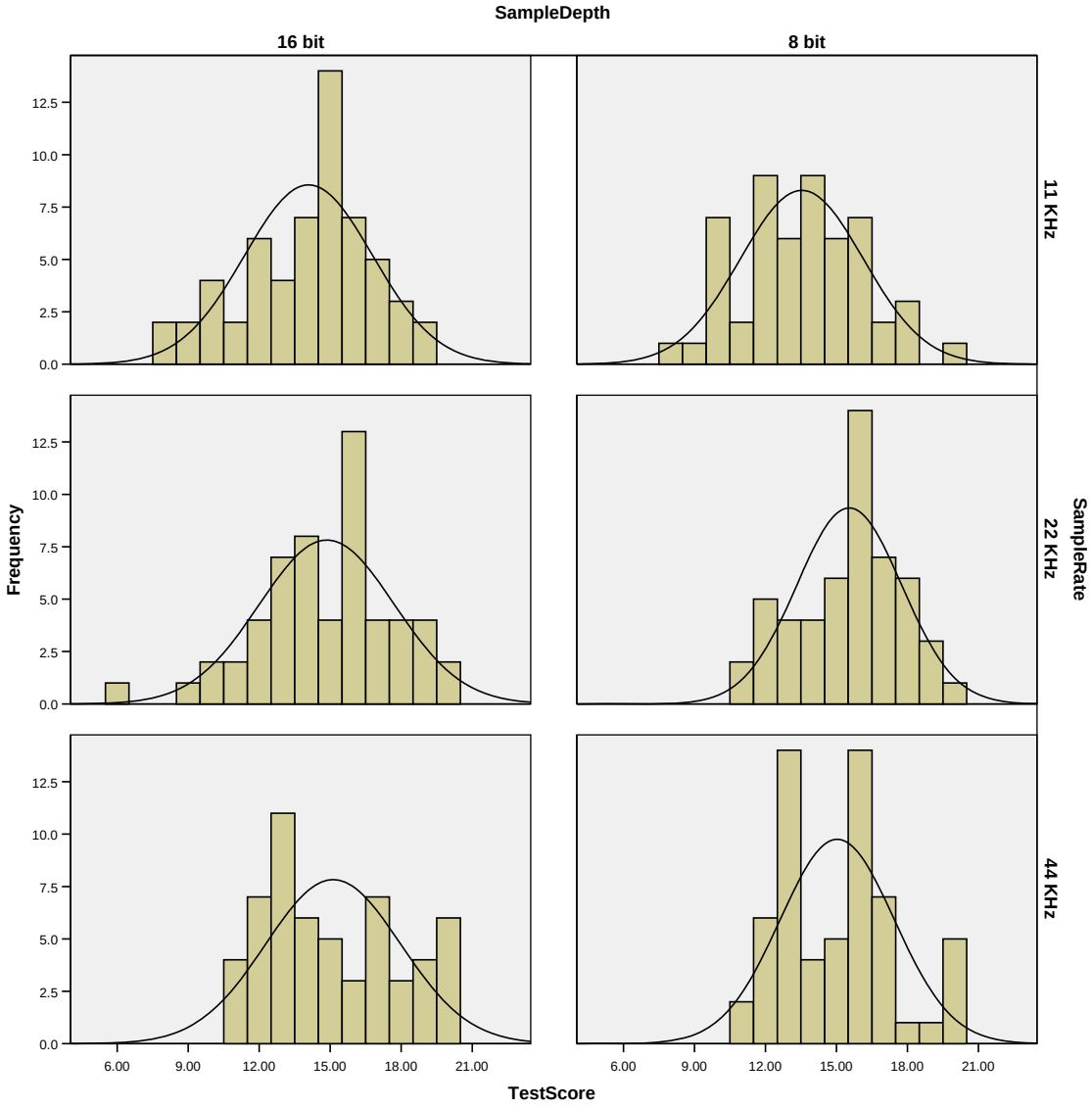


Figure 4.3. Histograms of the Distribution of Participants' Scores in Each Cell for the Two-way ANOVA Design. Superimposed Curves Show Normal Distributions.

Green and Salkind (2005) suggest that 2 x 3 ANOVA is robust to deviations from normality if the sample size is moderate or large, e.g., with more than 15 in each group. This study has a sample size of 335, with cell sizes ranging from 52 to 59. The researcher concluded that the normality assumption was tenable.

For the assumption of homogeneity of variance, this study used Levene's test, which tests whether the error variance of the dependent variable is equal among all groups. The Levene's test resulted in a significant value greater than .05 (p=.243). This indicated that the variances of scores in the six cells formed by the three levels of sample rate and the two levels of sample depth were homogenous. Therefore, the assumption of homogeneity of variance was found to be satisfactory. Table 4.4 displayed the results of Levene's test for test score as the dependent variable.

Table 4.4

Results of Levene's Test of Equality of Error Variances

F	*df1*	*df2*	*Sig.*
1.349	5	329	.243

Note: The significant level adopted is .05.

Hypothesis Testing Results

This study examined the effects of digital audio quality on students' performance in a LAN delivered web-based English listening comprehension test. By comparing

digital audio recorded with different sample rates and sample depths in the context of web-based English listening comprehension tests, the researcher sought to discover how sample rate and sample depth in digital audio affect test takers' performance, and how these two factors interact in shaping the quality of the tests. Three research questions were raised as the focus of this study:

1. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample rates (44 kHz, 22 kHz and 11 kHz)?
2. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample depths (16 bit and 8 bit)?
3. Do sample rate and sample depth interact in affecting students' performance in web-based English listening comprehension tests?

To answer the three research questions of this study, three null hypotheses had to be tested based on the analysis of the data:

HO_1: There is no significant difference of the test scores among test with 44 kHz digital audio, test with 22 kHz audio and test with 11 kHz digital audio.

HO_2: There is no significant difference of the test scores between test with 16 bit digital audio and test with 8 bit audio.

HO_3: There is no significant relationship between sample rate and sample depth in terms of the test scores.

Hypothesis one ($HO_{1)}$ was used to determine if there was a difference of test score between the groups that were tested using different sample rates, 44 kHz, 22 kHz and 11 kHz, and thus providing an answer to the first research question. Hypothesis two ($HO_{2)}$ was used to determine if there was a difference of test score between the groups that were tested using different sample depth, 16 bit and 8 bit, thus providing an answer to the second research question. Hypothesis three (HO_3) was used to determine if an interaction effect existed between sample rates and sample depth, thus answering the third research question.

The main statistical method employed for testing the above hypotheses was 2 x 3 ANOVA, with test score as the dependent variable and sample rate and sample depth as the independent variables. Based on the results the 2 x 3 ANOVA analysis, the decision to reject or fail to reject the null hypothesis was made and post hoc analysis was conducted where necessary. The software SPSS 14 was used as the tool for the analysis. Adopting a significant level of $\alpha = .05$, the 2 x 3 ANOVA analysis found a significant main effect for sample rate (F (2, 329) = 9.58, $p < .05$, $\eta_p^2 = .055$), with a medium effect size. This indicates that higher sample rate generally leads to higher scores than lower sample rate. No significant main effect was found using a critical α of .05 for sample depth (F (1, 329) = .002, $p > .05$), which indicates that the overall scores achieved by participants tested with 16 bit of sample depth (M = 14.68, SD = 2.82) are similar to the overall scores achieved by those tested with 8 bit of sample depth (M = 14.70, SD = 2.55). Tests of between-subject effects indicated that there was no significant interaction

effect between sample rate and sample depth (F (2, 329) = 1.54, p> .05) (See Table 4.5 and Figure 4.3).

Table 4.5

Two-way ANOVA: Sample Rate by Sample Depth on Test Score

Source	*df*	*Mean Square*	*F*	*Sig.*	*Partial Eta Squared*	*Observed Power(a)*
Corrected Model	5	29.900	4.353	.001*	.062	.965
Intercept	1	72224.974	10514.947	.000*	.970	1.000
Sample Rate	2	65.770	9.575	.000*	.055	.980
Sample Depth	1	.016	.002	.962	.000	.050
SampleRate * SampleDepth	2	10.593	1.542	.215	.009	.327
Error	329	6.869				
Total	335					
Corrected Total	334					

Note: The mean difference is significant at the .05 level.

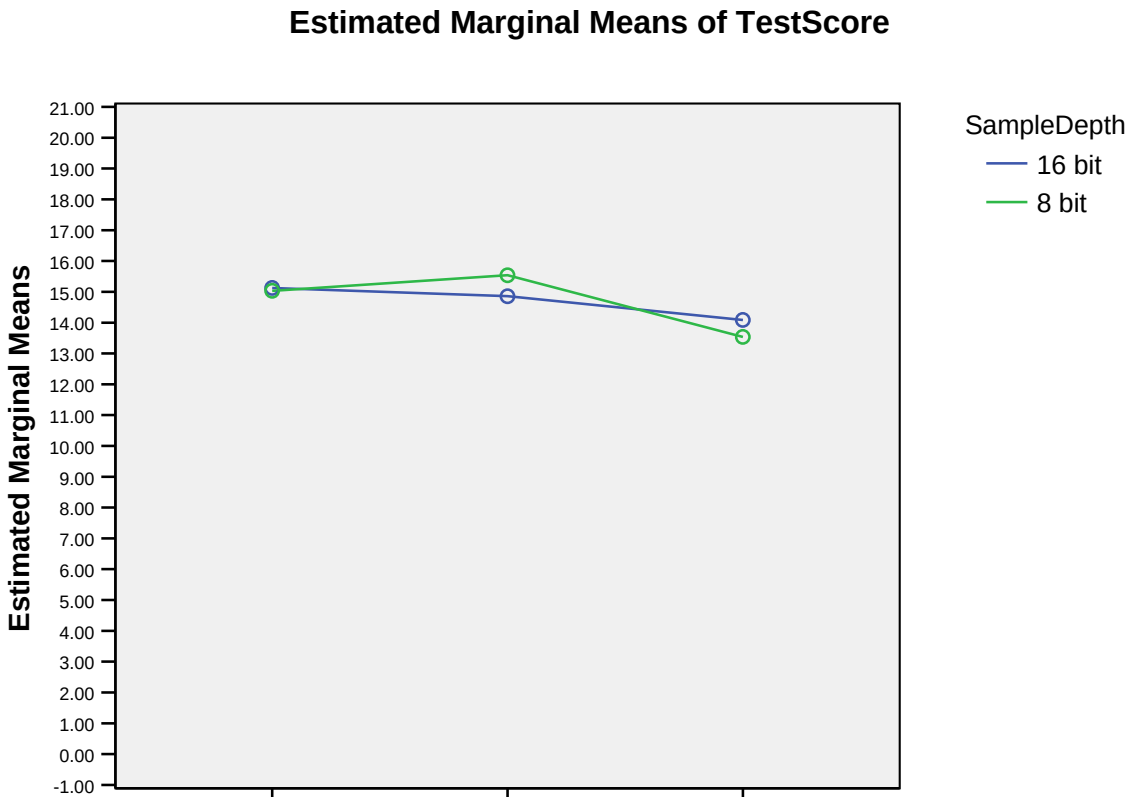


Figure 4.4. Line Graph of Sample Rate and Sample Depth for the Test Score.

Since the main effect of the independent variable, sample rate, had been found significant, a post hoc test, Tukey's test, was conducted to further compare the three levels of sample rate. It was found that the 44 kHz sample rate resulted in significantly higher scores than the 11 kHz sample rate with a mean difference of 1.26 ($p < .05$). The 22 kHz sample rate also resulted in significantly higher scores than 11 kHz sample rate, with a mean difference of 1.36 ($p < .05$). However, there was no significant difference between the 44 kHz sample rate and the 22 kHz sample rate on the participants' scores, with a mean difference of 0.11 (See Table 4. 6 for the post hoc test results).

Table 4.6

Tukey's Test for Comparison among Three Levels of Sample Rates

(I) SampleRate	*(J) SampleRate*	*Mean Difference (I-J)*	*Std. Error*	*Sig.*
44 KHz	22 KHz	-.1069	.35123	.950
	11 KHz	1.2568(*)	.34798	.001*
22 KHz	44 KHz	.1069	.35123	.950
	11 KHz	1.3638(*)	.35350	.000*
11 KHz	44 KHz	-1.2568(*)	.34798	.001*
	22 KHz	-1.3638(*)	.35350	.000*

Note: The mean difference is significant at the .05 level.

Based on the above analyses, the following is a summary of the hypotheses testing results:

HO_1: There is no significant difference of the test scores among test with 44 kHz digital audio, test with 22 kHz audio and test with 11 kHz digital audio.

This hypothesis was rejected since the data analysis through 2 x 3 ANOVA revealed a significant main effect for sample rates (F (2, 329) = 9.58, $p < .05$, $\eta_p^2 = .055$). High sample rates results in higher scores. Post hoc analysis showed both 44 kHz and 22 kHz result in significantly higher test scores than 11 kHz, but there was no significant difference between 44 kHz and 22 kHz.

HO_2: There is no significant difference of the test scores between test with 16 bit audio and test with 8 bit audio.

This hypothesis was retained since the 2 x 3 ANOVA revealed no significant main effect for sample depth (F (1, 329) = .002, $p > .05$). High sample depth did not result in significantly higher score than the low sample depth.

HO_3: There is no significant relationship between sample rate and sample depth in terms of the test scores.

This hypothesis was retained since tests of between-subject effects indicated that there was no significant interaction between sample rate and sample depth (F (2, 329) = 1.54, p> .05).

Answers to Research Questions

Based on the results of the above hypotheses testing, the answers to the three research questions are summarized as follows:

1. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample rates (44 kHz, 22 kHz and 11 kHz)?

Yes, students perform differently in web-based English listening comprehension tests using digital audio of different sample rates. Results of the data analysis show that students tested with 44 kHz audio or 22 kHz audio perform significantly better than students tested with 11 kHz audio, although students tested with 44 kHz audio do not perform significantly better than those tested with 22 kHz audio.

2. Do students perform differently in web-based English listening comprehension tests using digital audio of different sample depths (16 bit and 8 bit)?

No, students do not perform differently in web-based English listening comprehension tests using digital audio of different sample depths. Results of the data analysis do not indicate that students tested with 16 bit audio perform significantly better than students tested with 8 bit audio.

3. Do sample rate and sample depth interact in affecting students' performance in web-based English listening comprehension tests?

No, analysis of the data found no significant interaction effect between sample rates and sample depth.

Additional Findings

The additional findings of this study came from the pilot study with the English majors as the participants. Both the participants' scores and their answers to the follow-up interview questions suggested that the Internet connection in some areas of the world, as in the area where the data for this study was collected, is still not reliable enough to flawlessly transmit very high quality audio, such as 44 kHz and 16 bit audio, for the purpose of web-based English listening comprehension tests. This finding is meaningful especially for those who are considering administering high stakes English listening comprehension test on the Internet. Since high stakes tests usually require high quality audio recording (Wang, 2003), web-based testing through the Internet may not be a good option. Such a test may be more reliable and valid if administered on a local area network where high speed data transmission is possible. It may even need to be restricted to local testing methods by which the test resides on the local testing machines.

Another additional finding also came from the pilot study. A look at the data from the pilot test found that low sample rate and low sample depth did not seem to make a meaningful difference in the test scores of the participants. The participants of the pilot test were English majors and the test material was much easier to them than to the non-English majors. This seems to suggest that the effects of digital audio quality may interact with the degree of difficulty of the test. Therefore, in areas where high quality audio is restricted by the speed of the Internet or by any other factors, lower quality audio may be used if the level of difficulty of the test is adjusted accordingly.

The researcher wishes to point out, however, that the pilot test was conducted with a very small sample size and the data was not formally analyzed using such a statistical method as ANOVA. Therefore, the two additional findings from the pilot test should only be viewed as suggestions for future research. They should not be generalized until proven valid by future studies.

CHAPTER FIVE: CONCLUSIONS

Based on the data analysis results and findings, Chapter Five discusses the implications of this study, the limitations and weaknesses of this study, as well as some recommendations for future research. As the last chapter of this dissertation, this chapter provides the reader with a summary of the study and draws a final conclusion.

Implications of the Research

This study has examined the effects of digital audio quality on students' performance in a LAN delivered web-based English listening comprehension test. It has focused on sample rate and sample depth, which are the two most critical factors that determine the quality of digital audio (Scanlan, 2002). The findings of this study provide several implications on the use of digital audio for web-based English listening comprehension tests.

By comparing digital audios recorded with different sample rate and sample depth in the context of a web-based English listening comprehension test, the researcher confirmed the common belief that digital audio with higher sample rate produces higher quality sound and therefore enables test takers to achieve better performance. This finding is important since web-based language tests are often large scale and high stakes tests. Unlike the difference between 22 kHz and 11 kHz, the difference in students' performance between 44 kHz audio and 22 kHz audio is not significant and should not cause any concern in real world testing practices. Considering that a 44 kHz audio actually takes twice the storage space and twice the data transmission rate of the same audio at 22 kHz, the researcher argues that for the purpose of web-based English listening

comprehension test, the audio files which are played for students to hear do not need to exceed 22 kHz. On the other hand, it is also inappropriate to use a sample rate of 11 kHz since the results of the data analysis clearly show that students' performance is significantly poorer at 11 kHz than at 22 kHz. As for sample depth, the researcher argues that 8 bit is adequate since no significant difference has been found between 16 bit and 8 bit and 8 bit audio takes half the storage space and half the data transmission rate of the same audio at 16 bit. In other words, 22 kHz and 8 bit produce the best compromise between audio quality and audio data size for the purpose of web-based English listening comprehension test in the current computer and network technology environment. The researcher cautions that 11 kHz/8 bit audio or 11 kHz/16 bit audio should not be adopted for the purpose of web-based English listening test even though it may cost less in terms of money and computer and network resource, and even though some people may insist that it sounds just as good to the human ear.

Given the uneven distribution of computer capability or network speed in different geographical locations, designers and administrators of web-based English listening comprehension tests from different places may intend to use different criterion for audio files to suit their own technological conditions or financial situations. In such an EFL environment as the one in China, a national language test like CET is usually simultaneously but separately administered to hundreds of thousands of students in all parts of the country. If such a test were carried out using a web-based administration, as it may be in the near future, a reasonable standard will have to be made for the quality of the digital audio used in the test. Otherwise, serious unfairness will occur among test

takers. To make sure that no test taker is disadvantaged by a particular configuration of the audio, the researcher suggests that 22 kHz and 8 bit be set as the standards for sample rate and sample depth for the purpose of administering CET online.

Findings of this study provide implications for digitizing audio recordings for English listening comprehension tests for the purposes of archiving. As of the date that this study was conducted, the national English test of China, the CET, has had a history of over 20 years and has left more than a hundred volumes of recordings on audio tapes. Besides CET testing materials, each university or college in China may also have hundreds or even thousands of audio tapes containing material potentially useful for future web-based English listening comprehension tests. All these tapes need to be digitized both for prevention of signal deterioration and for quick availability for future web-based tests. It is possible that some universities or colleges have digitized these recording materials using low sample rate and low sample depth for archiving purposes. Some may have compressed the files to MP3 format. On the other hand, some other institutes may want to digitize their material at extremely high sample rate and extremely high sample depth, such as 96 kHz and 24 bit. Considering that this study has found 22 kHz and 8 bit to be adequate for the testing purpose, the researcher suggests that 44 kHz and 16 bit be used for digitally archiving audio material for the purpose of web-based English listening comprehension test. This combination of sample rate and sample depth faithfully reproduces the traditional analog audio signals and has best results when down sampled to 22 kHz and 8 bit (Liang, 2001).

Findings of this study have provided some implications for selecting computer hardware and accessories for the purpose of web-based English listening comprehension test. Since some computer hardware components (such as a sound card) and some computer accessories (such as head phones or audio cables) have certain frequency response properties and dynamic range properties, it is very important to know what specifications these hardware or accessories must have in order to be qualified for use in web-based English listening comprehension tests. Based on this study, the researcher suggests that any equipment used in a web-based English listening comprehension test must have the ability to record, transmit or faithfully play back digital audio of 22 kHz and 8 bit. This is very important since any piece of equipment that does not meet this standard will create a week point in the chain through which the sound signal flows. As for the CET tests in China, it is a true large scaled and high stakes English test that is simultaneously administered to hundreds of thousands of students in hundreds of universities or colleges in the country. In this case, standardized equipment such as headphones or even sound cards may have to be mandated should the CET test be administered online.

Limitations and Weaknesses

As mentioned in the beginning chapter of this dissertation, the major limitation of this study is that it has not examined the whole picture of digital audio. The quality of digital audio is determined by many factors such as sample rate, sample depth, compression rate and noise (Scanlan, 2002; Pan, 1993; Liang, 2001). This study only looked into two of those factors, sample rate and sample depth, since an experimental

design incorporating all the features at deferent levels will be difficult to carry out. The result of this study will only provide ideas on how sample rate and sample depth in digital audio may affect web-based language listening comprehension test performance. Other studies similar to this one are still needed to provide a comprehensive understanding of digital audio for English listening comprehension tests. Another limitation of this study is that only one signal source was looked at, i.e., prerecorded tape material. Although the recordings had been carried professionally on high performance metal cassette tapes, the quality of the signals did not represent what could have been achieved in ideal recording environments such as a digital sound studio. Any generalization of these findings to direct studio digital recording should be carried out with caution. Since this study was conducted using participants and instruments from an EFL environment, its findings are only suitable for reference in such an environment. Findings may not apply to web-based language assessment in ESL settings.

This study has some technical weakness in its design and implementation. One of the weaknesses is that the actual number of participants for this study was much smaller than the number proposed to achieve the power of .90, which this study was aimed at. Due to the limitations in the capacity of the computer labs at the Chinese university where data was collected, the actual number of participants that was chosen for this study was 360, much smaller than the proposed number, 576. There were a few absentees during the data collection process, with the actual sample size ended up being 335.

Another technical weakness that the researcher would like to address is that the data collection process for this study was not as well controlled and organized as the

researcher had intended to. Due to uncontrollable reasons, the researcher could not travel to the university in China to organize, monitor and control the data collection in person. All data collection was done under the coordination and supervision of a research assistant in the university. Despite the effort from both the researcher and the research assistant to adhere to the predefined data collection scheme, miscommunications often occurred due to language barriers or misunderstanding of technical terms. The final phase of this study fell into the longest Chinese holiday season, the Chinese Spring Festival, which interrupted the researcher's effort to review and verify all the communications that occurred during the research period. This weakness may have subsequently compromised the accuracy of some research information as well as the data on which all the findings of this study were based. Considering that a study such as this can have a strong influence on the policies or regulations governing the design and administration of web based English tests, which may include CET in the future, the researcher strongly suggests that this study be replicated using a larger sample size and a more strictly controlled and monitored data collection procedure.

The researcher would like to mention that the instrument of this study, the web-based English listening comprehension test, was not setup in a way that responses to individual items could be recorded. As a result, the internal reliability of the test could not be conducted. Although some research indicates that CET, from which the instrument of this study was derived, has high reliability in general (Yang, 2006), a check of item reliability for this individual test would certainly have given more credit to this study.

Recommendations for Future Research

This study did not examine all the factors that work together to form the whole picture of digital audio quality. Instead, it only focused on sample rate and sample depth with respect to their effects on web-based language listening comprehension tests. Many other questions still need to be answered to provide a comprehensive understanding of digital audio in listening comprehension tests. This study, for example, has found that high sample depth, 16 bit in this case, and low sample depth, 8 bit in this case, have no significant difference in terms of students' performance in web-based English listening comprehension test. However, we also know from the study of literature that low sample depth mainly affect the quality of digital audio in terms of dynamic range and noise level (Bartlett & Bartlett, 2002). Does this suggest that a certain amount of noise does not significantly interfere with students' performance in web-based English listening comprehension tests? If so, what is the maximum noise to signal ratio that can be tolerated? The researcher strongly recommends a future study that addresses this issue.

The following is a list of other topics and questions that interested researchers may want to look into for their future studies:

1. Effects of digital audio quality on the performance of students of different levels of English proficiency. Do sample rate and sample depth affect beginning English learners the same way as they affect advanced learners or native speakers?
2. A comparison of different compression methods or different file formats for the purpose of web-based English listening comprehension tests. Do different

compression methods or different file formats result in different performance among test takers?

3. Effects of digital audio quality on the test performance of students from EFL environments having different native languages. Do sample rate and sample depth affect the performance of EFL test takers the same way even though they speak different native languages?
4. Effects of digital audio quality on students' performance in web-based English listening tests in an ESL environment. Do such factors as sample rate and sample depth affect ESL English learners the same way as they affect EFL learners in web-based English listening comprehension tests?

Summary and Conclusion

This study has looked into the effects of digital audio quality on students' performance in a LAN delivered English listening comprehension test created with web technology and administered in s EFL setting. While the quality of digital audio depends on many different factors such as sample rate, sample depth, noise and compression (Scanlan, 2002; Pan, 1993; Liang, 2001), this study has focused on sample rate and sample depth. By comparing digital audio recorded with different sample rates (44 kHz, 22 kHz, 11 kHz) and different sample depths (16 bit and 8 bit) in the context of web-based English listening comprehension tests, the researcher has sought to discover how sample rate and sample depth in digital audio affect test takers' performance, and how these two factors interact in shaping the quality of a test. Three research questions have been addressed in this study: 1) Do students perform differently in web-based English

listening comprehension tests using digital audio of different sample rates? 2) Do students perform differently in web-based English listening comprehension tests using digital audio of different sample depths? 3) Do sample rate and sample depth interact in affecting students' performance in web-based English listening comprehension tests?

The data for this study was collected through a web-based English listening comprehension test given on a LAN to six groups of English learners as participants at a Chinese university. The participants of this study were both randomly selected and randomly assigned. The instrument of this study was a web-based English listening comprehension test derived from a disclosed CET test of 1995. The test was administered in three computer labs that had identical hardware and software configurations. Data was analyzed through 2 x 3 ANOVA and a post-hoc test. The results of the data analysis showed that there was a significant difference in the test scores between students tested with 22 kHz audio and those tested with 11 kHz audio, but there is not a significant difference between students tested with 44 kHz audio and those tested with 22 kHz audio. The results showed that students tested with 16 bit audio performed approximately the same as those tested with 8 bit audio in the web-based English listening comprehension test. No significant interaction effect was found between sample rate and sample depth. The researcher suggests that 22 kHz and 8 bit should be set as the standard for the quality of digital audio for web-based English listening comprehension tests. The researcher also suggests that equipment used in the process of a web-based English listening comprehension test must be capable of receiving, carrying or responding to this sample rate and this sample depth.

Aimed at a set of serious purposes and ambitious tasks, this study has achieved most of its initial objectives. While limitations and weaknesses exist, this study has been a successful attempt to understand the behavior of digital audio in web-based English listening tests. The methodology was generally effective and the findings may prove valuable for web-based English test designers and administrators. Perhaps an even more important contribution of this study is that it may serve as a pathfinder for future researchers who are interested in the same research subject. The researcher hopes that both the strengths and weaknesses of this study will provide valuable reference for other researchers of the same field.

REFERENCES

Adobe (2003). A digital audio primer. Retrieved June, 20, 2006, from http://www.adobe.com/designcenter/audition/articles/Digital_Audio_Primer/Digital_Audio_Primer.pdf

Alderson, J. C. (2000). Technology in testing: The present and the future. *System, 28,* 593-603.

Alderson, J. C., & Bachman, L. F. (2006). Series editor's preface. In Chapelle, C., & Douglas, D. (Eds.), *Assessing language through computer technology.* Cambridge, UK: Cambridge University Press.

Alexander, M., Bartlett, J. E., Truell, A. D., & Ouwenga, K. (2001). Testing in a computer technology course: An investigation of equivalency in performance between online and paper and pencil methods. *Journal of Career and Technical Education, 18*(1), 69-80.

Alkin, G. (1998). *Sound recording and reproduction*. Boston, Reed Education and Professional Publishing Ltd.

Anderson, N., Bachman, L. F., Cohen, A. D., & Perkins, K. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing, 8,* 41-66.

Arons, A. B. (1984). Computer-based instructional dialogs in science courses. *Science, 224*, 1051.

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17,* 1-42.

Bacon, S. M. (1992). Authentic listening in Spanish: How learners adjust their strategies to the difficulty of input. *Hispania. 75*, 29-43

Barker, F. B. (1989). Computer technology in test construction and processing. In Linn, R. L. (Ed.), *Educatinal Management* (3rd edn).New York: Macmillan Publishing.

Bartlett, B., & Bartlett, J. (2002). *Practical recording techniques: The step by step approach to professional audio recording* (3rd ed). Boston, MA: Focal Press.

Beatty, K. (2003). *Teaching and researching computer-assisted language learning.* London, UK: Pearson Education Limited.

Beggs, J., & Thede, D. (2001). *Designing web audio.* Sebastopol: O'Reilly & Associates, Inc.

Bicanich, E., Silvinski, T., Hardwicke, S., & Kapes, J. (1997). Internet based testing: A vision or reality. *THE journal, 25*(2), 61-65.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 120-136.

Bonham, S. W., Beichner R. J., Titus, A., & Martin, L. (2000). Education research using Web-based assessment systems. *Journal of Research on Computing in Education 33*, 28–45.

Brandenburg, K., & Popp, H. (2000). An introduction to MPEG layer-3. Retrieved March 12, 2007, from http://www.mp3-tech.org/programmer/docs/trev_283-popp.pdf

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education, 16,* 191-205.

Brusilovsky, P., & Miller, P. (1999). Web-based testing for distance Education. Retrieved September 12, 2005, from http;//www.contrib.andrew.cmu.edu/~plb/papers/Wednet99.html.

Chapelle, C. (2001). *Computer applications in second language acquisition.* UK, Cambridge University Press.

Chapelle, C., & Douglas, D. (2006). *Assessing language through computer technology.* UK, Cambridge University Press.

Choi, I. C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing, 20*(3), 295-320.

Chun, D. M., & Brandl, K. K. (1992). *Beyond form based-drill and practice: Meaning enhanced CALL on the Macintosh.* Foreign Language Annals, *25*(3), 255-267.

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Education Technology, 33*(5), 593–602.

Clifford, M. (1992). *Modern audio tedchnology*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences.* Hills Dale, NJ: Lawrence Erlbaum Associates.

Cole, R. P., MacIsaac, D., & Cole, D. M. (2001). A comparison of paper-based and web-based testing. Retrieved January 25, 2009, from http://physicsed.buffalostate.edu/pubs/AERA_Apr01/AERAAcolemac2.doc

Coniam, D. (1996). Computerized dictation for assessing listening proficiency. *CALICO*

Journal, 13(2-3), 73-85

Conrad, L. (1985). Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Education. 7,* 59-72

Cummins, J., & Sayers, D. 1995. *Brave new schools: Challenging cultural illiteracy through global learning networks.* New York: St. Martin's Press.

Currier, B. (1996). Digital audio for multimedia. Retrieved March 12, 2007, from http://www.synthetic-ap.com/digaudio/digaud1.html

Davies, A. (1990). *Principles of language testing.* UK: Basil Blackwell Ltd.

Douglas, D., & Chapelle, C. (1993). *A new decade of language testing research.* Alexandria: Teachers of English to Speakers of Other Languages, Inc.

Dunkel, P. (1999). Considerations in developing or using second /foreign language proficiency computer-adaptive tests. *Language Learning & Technology, 2*(2). 22-93.

ETS (2006). *The TOEFL test details: TOEFL Internet based testing overview.* Retrieved March 12, 2007, from vhttp://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=f138af5e44df4010VgnVCM10000022f95190RCRD&vgnextchannel=b5f5197a484f4010VgnVCM10000022f95190RCRD

Fries, B., & Fries, M. (2005). *Digital audio essentials*. UK: O'Reilly Media.

Fulcher, G. (2003). Interface design in computer-based language testing. *Language Testing, 20*(4), 384-408.

Green, S. B., & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (4th ed). Upper Saddle River, NJ: Prentice Hall.

Greenberg, R. (1998). Online testing. *Techniques: Making Education & Career Connections, 73*(3), 26-28.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff Publishing.

Hamilton, L. S., Klein, S. P., & Lorie, W. (2000). Using web-based testing for large-scale assessment. Retrieved January 25, 2009, from http://eric.ed.gov:80/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED447155

Haas, C., & Hayes, J. R. (1986). *Pen and paper versus the machine: Writers composing in hard-copy and computer conditions.* Pittsburgh: Carnegie-Mellon University, Communications Design Center.

Huber, D. M., & Runstein, R. E. (2005). *Modern recording techniques* (6th ed). Boston, MA: Focal Press.

Larson, J. W., & Madsen, H. S. (1985). Computer adaptive language testing: Moving beyond computer assisted testing. *CALICO Journal, 2*(3), 32-43

Li, Z. (2006). Dreamweaver 8 & PHP 网页资料库[*Dreamweaver 8 and PHP Material Libraries*]. Taipei, Taiwan: Flag Publishing Co., Ltd.

Liang, L. (2001). 数字音频实用手册 [*A Practical Handbook of Digital Audio*]. Qingdao, PRC: Science & Technology Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). *By design: Planning research on higher education.* Cambridge, MA: Harvard University Press.

Lipsey, M. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, London, UK: SAGE Publications.

Liu, Q., & Shi, J. (2007). An analysis of language teaching approaches and methods—effectiveness and weakness. *US-China Education Review*, *4*(1), 69-71.

Luther, A. C. (1997). *Principles of digital audio and video*. London, UK: Artec House, Inc.

Macromedia (2005). The Solution For Creating Rich-Media E-Learning Content. Retrieved Sep. 5, 2005, from http://www.macromedia.com/software/authorware/productinfo/overview/

Mason, B. J., Patry, M., & Bernstein D. J. (2001). An examination of the equivalence between non-adaptive computer-based test and traditional testing. *Journal of Educational Computing Research, 24*, 29–39.

MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: Computer versus pen and paper. *British Journal of Education Technology, 33*(2), 173–188.

Murphy, K., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests.* Mahwah, NJ: Lawrence Erlbaum Associates.

Nisbett, A. (1995). *The sound studio*. London, UK: Focal Press.

NCETC (2006). The testing syllabus for College English Test-Band 4. Retrieved May 28, 2007, from http://www.en.cet.edu.cn/dgdisplaynews.asp?id=298

Pan, D. Y. (1993). Digital audio compression. *Digital Technical Journal, 5*(2), 1-14

Pimentel, R. (1999). Design of net-learning systems based on experiential learning. *Journal of Asynchronous Learning Networks, (3)*2.

Plichta, B., & Kornbluh, M. (2005). Digitizing speech recordings for archive purposes. Retrieved October 8, 2005, form http://www.historicalvoices.org/papers/audio_digitization.pdf

Pohlmann, K. C. (1989). *Principles of digital audio.* Indianapolis: Howard W. Sams & Company.

Pohlmann, K. C. (1991). *Advanced digital audio.* Indianapolis: Howard W. Sams & Company.

Rabinowitz, S., & Brandt, T. (2001). Computer-based assessment: Can it deliver on its promise? Retrieved March 9, 2008, from http://www.eric.ed.gov:80/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/db/80.pdf

Ribas-Corbera, J. (2003). Windows media 9–a platform to deliver audio and video for Internet and broadcast applications. Retrieved May 6, 2007, from http://www.ebu.ch/en/technical/trev/trev_293-ribas.pdf

Richards, J. C., & Rodgers, T. S. (2001). Approaches and methods in language teaching: A description and analysis. UK: Cambridge University Press.

Roever, C. (2001). Web-based language testing. *Language learning & Technology, 5*(2), 84-94.

Rosen, S., & Howell, P. (1998). *Signals and systems for speech and hearing*. London, UK: Academic Press.

Scanlan, C. L. (2002). Digital audio / sound. Retrieved Jan. 3, 2009, from http://www.umdnj.edu/idsweb/idst3400/audio.htm

Sereci, S. (2003). Computerized adaptive testing: An introduction. Retrieved June 25, 2005, from http://rave.ohiolink.edu/databases/record/eric/ED480083

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stockdale, A. (2002). Tools for digital audio recording in qualitative research. *Social Research Update,* 1-4.

van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Netherlands: Kluwer Academic Publishers.

Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement, 30*(1), 1021.

Wang, N. (2003). 电脑及因特网在未来语言测试中的应用 [*Computers, Internet and the Future of Language Assessment*]. Changchun, PRC: Educational Publishing.

Wang, T. H., Wang, K. H., Wang, W. L. et. al. (2004). Web-based assessment and test analysis (WATA) system: Development and evaluation. *Journal of computer assisted Learning, 20*, 57-71.

Warschauer, M., & Kern, R. (2000). Network-based language teaching: Concepts and Practice. Cambridge, UK: Cambridge University Press.

Watkinson, J. (1994). *An introduction to digital audio*. London, UK: Butterworth Group.

Watkinson, J. (1998). *The art of digital audio*. London, UK: Butterworth Group.

Waugh, M. L. (1985). Effects of microcomputer-administered diagnostic testing on immediate and continuing science achievement and attitudes. *Journal of Research in Science Teaching, 23*, 793.

Xiao, M. (2007). An empirical study of using Internet-based desktop videoconferencing in an EFL setting. Retrieved October 8, 2008, from http://www.ohiolink.edu/etd/send-pdf.cgi/Xiao%20Mingli.pdf?acc_num=ohiou1194703859

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *Acoustical Society of America, 117*(5), 3255-3267.

Yang, M. (2006). Validating a simulated test of CET 4. Asian EFL Journal, *12*. Retrieved January 20, 2009, from http://www.asian-efl-journal.com/pta_May_ym.doc

APPENDIX A: SAMPLE PAGES OF LISTENING COMPREHENSION SECTION OF CET BAND 4

4. Below the canopy level of a tropical rainforest grows an overabundance of plants.

5. New tree seedlings will not survive to reach the canopy level unless _____.

6. Epiphytes, which form much of the understory of the rainforest, get all their water and nutrients from _____.

7. Stranglers are so called because they _____ by blocking the sunlight and competing for the nutrients.

8. Since rainforest bacteria and trees depend on each other for life, the relationship they form is termed _____.

9. Plant species are dispersed over a large area with the help of _____.

10. As we are still ignorant of millions of unique life forms in the rainforest, deforestation can be compared to the destruction of _____.

Part III Listening Comprehension (35 minutes)

Section A

Directions: *In this section, you will hear 8 short conversations and 2 long conversations. At the end of each conversation, one or more questions will be asked about what was said. Both the conversation and the questions will be spoken only once. After each question there will be a pause. During the pause, you must read the four choices marked A), B), C) and D), and decide which is the best*

answer. Then mark the corresponding letter on ***Answer Sheet 2*** *with a single line through the centre.*

11. A) She isn't going to change her major.
    B) She plans to major in tax law.
    C) She studies in the same school as her brother.
    D) She isn't going to work in her brother's firm.

12. A) She will do her best if the job is worth doing.
    B) She prefers a life of continued exploration.
    C) She will stick to the job if the pay is good.
    D) She doesn't think much of job-hopping.

13. A) Stop thinking about the matter.
    B) Talk the drug user out of the habit.
    C) Be more friendly to his schoolmate.
    D) Keep his distance from drug addicts.

14. A) The son.
    B) The father.
    C) The mother.
    D) Aunt Louise.

15. A) Stay away for a couple of weeks.
    B) Check the locks every two weeks.
    C) Look after the Johnsons' house.
    D) Move to another place.

16. A) He didn't want to warm up for the game.
    B) He didn't want to be held up in traffic.
    C) He wanted to make sure they got tickets.
    D) He wanted to catch as many game birds as possible.

17. A) It will reduce government revenues.
    B) It will stimulate business activities.
    C) It will mainly benefit the wealthy.
    D) It will cut the stockholders' dividends.

18. A) The man should phone the hotel for directions.
    B) The man can ask the department store for help.
    C) She doesn't have the hotel's phone number.
    D) The hotel is just around the corner.

Questions 19 to 21 are based on the conversation you have just heard.

19. A) To interview a few job applicants.
    B) To fill a vacancy in the company.
    C) To advertise for a junior sales manager.
    D) To apply for a job in a major newspaper.

20. A) A hardworking ambitious young man.
    B) A young man good at managing his time.
    C) A college graduate with practical working experience.
    D) A young man with his own idea of what is important.

21. A) Not clearly specified.
    B) Not likely to be met.
    C) Reasonable enough.
    D) Apparently sexist.

Questions 22 to 25 are based on the conversation you have just heard.

22. A) The latest developments of an armed rebellion in Karnak.
    B) The fall of Karnak's capital city into the hands of the rebel forces.
    C) The epidemic that has just broken out in the country of Karnak.
    D) The peace talks between the rebels and the government in Karnak.

23. A) The epidemic has been brought under control.
    B) There are signs of progress in the peace process.
    C) Great improvements are being made in its capital.
    D) There's little hope of bringing the conflict to an end.

24. A) Late in the morning.
    B) Early in the afternoon.
    C) Sometime before dawn.
    D) Shortly after sunrise.

25. A) Inadequate medical care.
    B) Continuing social unrest.
    C) Lack of food, water and shelter.
    D) Rapid spreading of the epidemic.

Section B

Directions: *In this section, you will hear 3 short passages. At the end of each passage, you will hear some questions. Both the passage and the questions will be spoken only once. After you hear a question, you must choose the best answer from the four choices marked A), B), C) and D). Then mark the corresponding letter on Answer Sheet 2 with a single line through the centre.*

Passage One

Questions 26 to 28 are based on the passage you have just heard.

26. A) One of the bridges between North and South London collapsed.
    B) The heart of London was flooded.
    C) An emergency exercise was conducted.
    D) A hundred people in the suburbs were drowned.

27. A) Fifty underground stations were made waterproof.
    B) A flood wall was built.
    C) An alarm system was set up.
    D) Rescue teams were formed.

28. A) Most Londoners were frightened.
    B) Most Londoners became rather confused.
    C) Most Londoners took Exercise Floodcall calmly.
    D) Most Londoners complained about the trouble caused by Exercise Floodcall.

Passage Two

Questions 29 to 31 are based on the passage you have just heard.

29. A) It limited their supply of food.
    B) It made their eggshells too fragile.
    C) It destroyed many of their nests.
    D) It killed many baby bald eagles.

30. A) They found ways to speed up the reproduction of bald eagles.
    B) They developed new types of feed for baby bald eagles.
    C) They explored new ways to hatch baby bald eagles.
    D) They brought in bald eagles from Canada.

31. A) Pollution of the environment.
    B) A new generation of pest killers.
    C) Over-killing by hunters.
    D) Destruction of their natural homes.

Passage Three

Questions 32 to 35 are based on the passage you have just heard.

32. A) Why people hold back their tears.
    B) Why people cry.
    C) How to restrain one's tears.
    D) How tears are produced.

33. A) What chemicals tears are composed of.
    B) Whether crying really helps us feel better.
    C) Why some people tend to cry more often than others.
    D) How tears help people cope with emotional problems.

34. A) Only one out of four girls cries less often than boys.
    B) Of four boys, only one cries very often.
    C) Girls cry four times as often as boys.
    D) Only one out of four babies doesn't cry often.

Passage Two

Questions 29 to 31 are based on the passage you have just heard.

29. A) It limited their supply of food.
    B) It made their eggshells too fragile.
    C) It destroyed many of their nests.
    D) It killed many baby bald eagles.

30. A) They found ways to speed up the reproduction of bald eagles.
    B) They developed new types of feed for baby bald eagles.
    C) They explored new ways to hatch baby bald eagles.
    D) They brought in bald eagles from Canada.

31. A) Pollution of the environment.
    B) A new generation of pest killers.
    C) Over-killing by hunters.
    D) Destruction of their natural homes.

Passage Three

Questions 32 to 35 are based on the passage you have just heard.

32. A) Why people hold back their tears.
    B) Why people cry.
    C) How to restrain one's tears.
    D) How tears are produced.

33. A) What chemicals tears are composed of.
    B) Whether crying really helps us feel better.
    C) Why some people tend to cry more often than others.
    D) How tears help people cope with emotional problems.

34. A) Only one out of four girls cries less often than boys.
    B) Of four boys, only one cries very often.
    C) Girls cry four times as often as boys.
    D) Only one out of four babies doesn't cry often.

(43) _____ influence on all the great Greek and Roman schools of philosophy. Yet for all his fame and influence, Socrates himself never wrote a word.

Socrates (44) _____ _____ in Athens. They wanted him silenced. Yet many were probably surprised that he accepted death so readily.

Socrates (45) _____ _____. But Socrates, as a firm believer in law, reasoned that it was proper to submit to the death sentence. (46) _____ _____.

Part IV Reading Comprehension (Reading in Depth) (25 minutes)

Section A

Directions: *In this section, there is a short passage with 5 questions or incomplete statements. Read the passage carefully. Then answer the questions or complete the statements in the fewest possible words on Answer Sheet 2.*

Questions 47 to 51 are based on the following passage.

America is a country that now sits atop the cherished myth that work provides rewards, that working people can support their families. It's a myth that has become so divorced from reality that it might as well begin with the words "Once upon a time." Today 1.6 million New Yorkers suffer from "food insecurity," which is a fancy way of saying they don't have enough to eat. Some are the people who come in at night and clean the skyscrapers that glitter along the river. Some pour coffee and take care of the aged parents of the people who live in those buildings. The American Dream for the

APPENDIX B: ENGLISH LISTENING COMPREHENSION TEST AS THE INSTRUMENT FOR THIS STUDY

(Audio available upon request)

Section A

1. A) Librarian and student.
   B) Operator and caller.
   C) Boss and secretary.
   D) Customer and repairman.

2. A) Look for the key.
   B) Repair the car.
   C) Fix a shelf.
   D) Paint a shelf.

3. A) To make the woman angry.
   B) To please the man's mother.
   C) David is the man's good friend.
   D) David is good at carrying on conversations.

4. A) He must meet his teacher.
   B) He must attend a class.
   C) He must go out with his girlfriend.
   D) He must stay at school to finish his homework.

5. A) He wants to pay.
   B) he doesn't want to eat out
   C) He wants to eat somewhere else.
   D) He doesn't like Japanese food.

6. A) He didn't work as hard as he was supposed to.

B) He didn't pass the physics exam.
C) He did better in an earlier exam.
D) He found something wrong with the exam.

7. A) He is attending his sick mother at home.
B) He is on a European tour with his mother.
C) He is at home on sick leave.
D) He is in Europe to see his mother.

6. A) They don't know how to get to Mike's home.
B) They are discussing when to meet again.
C) They went to the same party some time ago.
D) They will go to Mike's birthday party.

9. A) Five lessons.
B) Three lessons.
C) Twelve lessons.
D) Fifteen lessons.

10. A) Find a larger room.
B) Sell the old table.
C) Buy two bookshelves.
D) Rearrange some furniture.

Section B

Passage One

Questions 11 to 13 are based on the passage you have just heard.

11. A) Courses in British history.
B) Language courses.
C) Courses in sports.

D) Teacher training courses.

12. A) To attract more students.
B) To make the courses suitable for students of all levels.
C) To let the students have a good rest.
D) To make the summer school more like a holiday.

13. A) Because they all work very hard.
B) Because their teachers are all native speakers of English.
C) Because they learn not only in but also out of class.
D) Because they are all advanced students.

Passage Two

Questions 14 to 16 are based on the passage you have just heard.

14. A) Because it takes too long to process all the applications.
B) Because its resources are limited.
C) Because it is a library for special purposes.
D) Because there is a shortage of staff.

15. A) Discard his application from.
B) Forbid him to borrow any items.
C) Cancel his video card.
D) Ask him to apply again.

16. A) One month.
B) One week.
C) Two weeks.
D) Two months.

Passage Three

Questions 17 to 20 are based on the passage you have just heard.

17. A) Chemicals.
    B) Vapor.
    C) Water.
    D) Gas.

18. A) By passing steam over dry ice.
    B) By turning ordinary ice into steam.
    C) By heating dry ice.
    D) By mixing dry ice with ordinary ice.

19. A) It takes a longer time to melt.
    B) It is lighter to carry.
    C) It is cleaner to use than ordinary ice.
    D) It is not so cold as ordinary ice.

20. A) In the 1920's.
    B) In the 1930's.
    C) In the 1940's.
    D) In the 1950's.

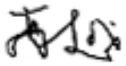
APPENDIX C: LETTER OF PERMISSION FOR USE OF CET TEST ITEMS AS THE INSTRUMENT OF THIS STUDY

October 20, 2008

To whom it may concern:

This letter is to certify that Mr. Xiangui Yang has permission to use CET test items for his study entitled *Effects of Digital Audio Quality on Students' Performance in Web-based English Listening Comprehension Tests*. Your cooperation and assistance are much appreciated. Thank you!

Yue Gao



Director of College English Section
North China University of Technology
Phone: 10-8880-2901 (office)

APPENDIX D: CONTACT INFORMATION OF RESEARCHER AND RESEARCH ASSISTANT

Researcher:

Xiangui Yang
College of Education
Ohio University
Phone: 757-348-2764
Email: xy878301@ohio.edu

Research Assistant:

Yihong Zhang
Department of English Language and Literature
North China University of Technology
Phone: 10-8880-2011
Email: yihongzh@ncut.edu.cn

APPENDIX E: IRB APPROVAL LETTER

OHIO UNIVERSITY

Office of the Vice President for Research

07E114

Office of Research Compliance
Research and Technology
Center 117
Athens OH 45701-2979

T: 740.593.0664
F: 740.593.9838
www.ohiou.edu/research

A determination has been made that the following research study is exempt from IRB review because it involves:

Category 1 - research conducted in established or commonly accepted educational settings, involving normal educational practices

Project Title: Effects of Digital Audio Quality on Student's Performance in Web-Based English Listening Comprehension Tests

Project Director: Xiangui Yang

Department: Educational Studies

Advisor: Teresa Franklin

Rebecca Cale 5/24/07

Rebecca Cale, Associate Director, Research Compliance Institutional Review Board — Date

The approval remains in effect provided the study is conducted exactly as described in your application for review. Any additions or modifications to the project must be approved by the IRB (as an amendment) prior to implementation.

APPENDIX F: IRB AMENDMENT APPROVAL LETTER

OHIO UNIVERSITY
Office of the Vice President for Research

A 07E114

Office of Research Compliance
Research and Technology
Center 117
Athens OH 45701-2979

T: 740.593.0664
F: 740.593.9838
www.research.ohiou.edu

The amendment, detailed below, and submitted for the following research study has been approved by the Institutional Review Board at Ohio University. Approval date of this amendment does not affect the expiration date of the original approval.

Amendment: Change Location to North China University of Technology

Project: Effects of Digital Audio Quality on Student's Performance in Web-Based English Listening Comprehension Tests

Project Director: Xiangui Yang

Advisor: (if applicable) Teresa Franklin

Department: Educational Studies

Rebecca G. Cale
Institutional Review Board

12/22/08
Date