Voice Track Computer Based Simulation For Medical Training

2005

Alpesh Purshottam Makwana
University of Central Florida

VOICE TRACK COMPUTER BASED SIMULATION FOR MEDICAL TRAINING

by

ALPESH PURSHOTTAM MAKWANA
B.E. M.S. University of Baroda, 2002

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Modeling and Simulation
in the College of Arts and Science
at the University of Central Florida
Orlando, Florida

Summer Term
2005

ABSTRACT

Varying the delivery rate of audio-based text within web-based training increases the effectiveness of the learning process and improves retention when compared with a fixed audio-based text delivery rate. To answer this question, two groups of 20 participants and one group of 10 participants were tested using the Web-based Anatomy & Physiology course modules developed by Medsn, Inc. The control group received the static speed of 128 words per minute while the experimental group received the initial speed of 128 words per minute with the option to change the speed of the audio-based text. An additional experimental group received the initial speed of 148 words per minute also having the option to vary the speed of the audio-based text. A three way single variable Analysis of Variance (ANOVA) was utilized to examine speed of voice presentation differences. The results were significant, $F(2, 47) = 4.67$, $p=0.014$, $\eta 2 = 0.166$. The mean for the control group was ($M = 7.2$, $SD = 1.69$) with the experimental groups at, ($M = 8.4$, $SD = 1.31$) and with extra groups at ($M = 8.6$, $SD = 1.26$).

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Peter Kincaid, who supported me during my graduate program. I would also like to thank Dr. Bala Jaganathan for his insight and contribution. I would like to acknowledge Medsn. Inc., for their instruction module and Speakonia, Text to Speech converter for use of synthetic sound during the experiment. Also, a special thanks to all the participants who contributed their precious time in doing the test and making it successful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

The purpose of this research is to focus on how and to what degree the ability to change speed of E-Learning module of complicated basic medical science subjects such as Anatomy and Physiology will improve the overall learning process and retention of the subjects. To improve this situation, new methods of training using the latest technology are being developed and tested.

In 1960 Bell laboratories developed a material as first practical mechanism for presenting “speeded speech” for instructional purposes, Breuel and Levens (1969). It presented the accelerated speech in a natural pitch and was at the same time understandable up to a note of 240 words per minute Stitch (1969). Current approach is the use of the software program like study by which students accessing streaming video in distance education courses tend to use speed control of the multimedia presentation when available. The limited information available suggests that the features offered by the software can improve the efficiency and effectiveness of learning.

Does the changing of speed of web-based training increase the effectiveness of learning process and improve the retention of the participants? To answer these questions two groups of twenty participants and one group of ten participants were tested. A 1×2 factorial analysis was utilized to examine all research questions. The outcome of the study clearly suggests that there is a significant difference between control group with population size of twenty compared with two experimental groups with population size of twenty and ten.

CHAPTER TWO: LITERATURE REVIEW

For so many years the publication of books and magazines is increasing rapidly. Conversion of this material into recorded form is keeping pace. There are great interest and effort in reducing the bulk of this material in space and time. For decades the researchers put an effort to increase the time density by speech compression. Speech compression refers to any method of reducing the time required to transmit a spoken message. Time-compressed speech is also referred to as accelerated, compressed, time-scale modified, sped-up, rate converted, or time-altered speech. For time-compression speech various techniques have been developed since decades. The primary motivation for time-compressed speech is to reduce the time needed for user to listen to a message and to increase the communication capacity of the ear, and to do data reduction to save storage space and transmission bandwidth for speech messages.

Time –compressed speech can be used in a various applications such as teaching, aids to the disabled and human-computer interfaces. According to Stitch (1969) listening to teaching materials twice that have been speeded up by a factor of two is more effective than listening to them once at normal speed. Speech can be slowed for learning languages, or for the hearing impaired. Time compression techniques have also been used in speech recognition systems to time normalize input utterances to a standard length (Malah, 1979). Time-compressed speech has been used to speed up message presentation in voice mail systems (Hejna 1990).

2.1 General Time-Compression Techniques

There are various techniques for changing the playback speed of speech these methods are described briefly. These techniques are primarily concerned with reproducing the entire recording, not scanning portions of the signal. Most of these methods also work for slowing speech down. Much of the research summarized here was performed between the mid 1950's and the mid 1970's, often in the context of accelerated teaching techniques, or aids for the blind.

2.1.1 Speaking Rapidly

The normal English speaking rate on an individual is between 130–200 words per minute. When speaking fast, a talker unintentionally changes relative attributes of his speech such as pause durations, consonant-vowel duration, etc. According to Beasley and Maki (1976) talkers can only compress their speech to about 70% because of physiological limitations.

2.1.2 Speed Changing

Speed changing is analogous to playing a tape recorder at a faster or slower speed. This method can be replicated digitally by changing the sampling rate during the playback of a sound. These techniques are undesirable since they produce a frequency shift proportional to the change in playback speed, causing a decrease in intelligibility.

2.1.3 Speech Synthesis

With purely synthetic speech it is possible to generate speech at a variety of word rates. Current text-to-speech synthesizers can produce speech at rates up to 550 wpm. This is typically done by selectively reducing the phoneme and silence durations. This technique is powerful, particularly in aids for the disabled, but is not relevant to recorded speech.

2.1.4 Vocoding

Vocoders that extract pitch and voicing information can be used to time-compress speech. Most vocoding efforts, however, have focused on bandwidth reduction rather than on naturalness and high speech quality.

2.1.5 Silence removal

The speech is made natural by removal of silences or pauses. Though the resulting speech is natural people find it exhausting to listen because the speaker never pause the breath.

2.2 Time domain Techniques

2.2.1 Sampling

Miller and Licklider (1950) demonstrated the temporal redundancy of speech with an experiment. It increased the channel capacity by switching speech on and off at regular intervals so the channel could be used for another transmission. It was established that if interruptions were made at frequent intervals, large portions of a message could be deleted without affecting intelligibility.

Listening time could be saved by abutting the interrupted speech segments, which was first done by Garvey who manually spliced audio tape segments together (Garvey 1953). In the Fairbanks, or sampling, technique, segments of the speech signal are alternatively discarded and retained by the use of a modified tape recorder with four rotating pickup heads. This has traditionally been done isochronously—at constant sampling intervals without regard to the contents of the signal.

2.2.2 Sampling with Dichotic presentation

This sampling method is achieved by playing the standard sampled signal to one ear and the “discarded” material to the other ear (Scott 1967). Under this dichotic condition, intelligibility and comprehension increase. The subjects in this technique prefer to a diotic presentation of a conventionally sampled signal. Listeners initially reported a switching of attention between ears, but they quickly adjusted to this unusual sensation. For compression ratios up to 50%, the two signals to the ears contain common information. For compressions greater than 50% some information is necessarily lost.

2.2.3 Selective sampling

The basic sampling technique periodically removes pieces of the speech waveform without regard to whether it contains any redundant speech information. David and Mc-Donald (1956) demonstrated a bandwidth reduction technique that selectively removed the redundant pitch periods from speech signals. Scott applied the same ideas to time compression, setting the sampling and discard intervals to be synchronous with the pitch periods of the speech. Discontinuities in the time compressed signal were reduced, and intelligibility increased (Scott and Gerber 1972). Neuburg (1978) developed a similar technique in which intervals equal to the pitch period were discarded but not synchronous with the pitch pulses.

2.2.4 Synchronized Overlap Add Method

The synchronized overlap add method (SOLA) was first described by Roucos and Wilgus (1985) and has recently become popular in computer-based systems. It is a fast non-iterative optimization of a Fourier-based algorithm (Griffin and Lim 1984). “Of all time scale modification methods proposed, SOLA appears to be the simplest computationally, and therefore most appropriate for real-time applications” (Wayman, Reinke, and Wilson 1989). Conceptually, the SOLA method consists of shifting the beginning of a new speech segment over the end of the preceding segment to find the point of highest cross-correlation. Once this point is found, the frames are overlapped and averaged together, as in the sampling method. This technique provides a locally optimal match between successive frames; combining the frames in this manner tends to preserve the time-dependent pitch, magnitude, and phase of a signal. The shifts do not accumulate since the target position of a window is independent of any previous shifts (Henja 1990). The SOLA method is simple and effective as it does not require pitch extraction, frequency-domain calculations, phase unwrapping, and is non-iterative (Makhoul and El-Jaroudi 1986). The SOLA technique can be considered a type of selective sampling that effectively removes redundant pitch periods.

2.3 Frequency Domain Techniques

2.3.1 Harmonic Compressor

The American Foundation for the Blind produced devices for increasing the speed of tape and disc reproducers. It was limited solution because the intelligibility of recorded speech is lost to most people at about 75% over-speed due to pitch distortion. Grant Fairbanks (1954) of the University of Illinois, developed a method of speech compression without pitch distortion. This method sampled speech into small segments, discarded some segments and eliminated gaps. Compression capability was provided by this method.

The scientist of Bell Laboratories had thought of the harmonic compressor while studying frequency compression as a possible means for reducing the transmission bandwidth required in voice communication. The device was first simulated on a digital computer. Schroeder, Logan, and Presti of Bell Telephone Laboratories described the basic theory of harmonic speech compressor. In this the harmonic compressor halves the frequencies of the individual harmonic frequency components and from these half frequency components synthesizes a new signal. If the input signal was applied at twice normal rate, the new signal will approximate the speech of a person speaking at twice his normal rate, with his normal pitch.

2.3.2 Phase Vocoding

A vocoder that maintains phase can be used for time-compression (Dolson 1986). A phase vocoder is similar to harmonic compressor which can be interpreted as a filterbank. However, a phase vocoder is significantly more complex because calculations are done in the frequency domain, and the phase of the original signal must be reconstructed. Phase vocoding techniques are more accurate than time domain techniques, but are an order of magnitude more computationally complex because Fourier analysis is required.

The phase vocoder is particularly good at slowing speech down to hear features that cannot be heard at normal speed—such features are typically lost using time domain techniques. According to Dolson (1986) a number of time-domain procedures can be employed at substantially less computational expense. But from a standpoint of fidelity i.e., the relative absence of objectionable artifacts, the phase vocoder is by far the most desirable.

2.4 Combined Compression Techniques

2.4.1 Silence Removal and Sampling

According to Maxemchuk (1980) eliminating every other non-silent block (1/16th second) produced "extremely choppy and virtually unintelligible playback. This technique produced compressions of 33 to 50 percent. The characteristic of this technique is that those words which the speaker considered to be most important and spoke louder were virtually undistorted, whereas those words that were spoken softly are shortened. After a few seconds of listening to this type of speech, listeners appear to be able to infer the distored words and obtain the meaning of the message." This technique can be useful for users of a message system to scan a large number of messages and determine which they wish to listen to more carefully or for users of a dictation system to scan a long document to determine the areas they wish to edit. Silence compression and sampling can be combined in several ways. Silences can first be removed from a signal that is then sampled. Alternatively, the output of a silence detector can be used to set boundaries for sampling, producing a selective sampling technique. Using silences to find discard intervals eliminates the need for a windowing function to smooth (de-glitch) the sound at the boundaries of the sampled intervals.

2.4.2 Silence Removal and SOLA

Removing silences and time-compressing speech using a technique such as the verlap-add method would be linearly independent, and could thus be performed in either order. There are some minor differences, because the SOLA algorithm makes assumptions about the properties of the speech signal. It was found that there is a slight improvement in speech quality by applying the SOLA algorithm before removing silences. The timing parameters must be modified under these conditions. For example with speech compressed 50%, the silence removal timing thresholds must also be cut in half. This combined technique is effective, and can produce a fast and dense speech stream

2.4.3 Dichotic SOLA Presentation

In this a sampled signal is compressed by 50% and presented dichotically so that exactly half the signal is presented to one ear, while the remainder of the signal is presented to the other ear. Generating such a lossless dichotic presentation is difficult with the SOLA method because the segments of speech are shifted relative to one another to find the point of maximum similarity. However, by choosing two starting points in the speech data carefully it is possible to maximize the difference between the signals presented to the two ears. This technique is effective as it combines the high quality sounds produced with the SOLA algorithm with the binaural effect of the dichotic presentation.

CHAPTER THREE: METHODOLOGY

3.1 Materials and Apparatus

Paper materials covered informed consent, demographic survey, post-test multiple-choice questions, and subjective questionnaires. Additional instruments included web-based modules developed by Medsn Inc., Speakonia text to speech converter developed by CFS technologies, and SPSS 11.5 for Windows. Two computer systems, a Desktop and Notebook with a pair of external speakers were required for the experiment.

A crucially important instrument in this study is the web-based model (Figure 1). This interface includes four windows which include a navigation system for the learner. Additional windows include an auto scrolls text window, a presentation window at the center top and an animation window in the center. The animation and the text are followed by a real time voice track (Image courtesy of Medsn, Inc).

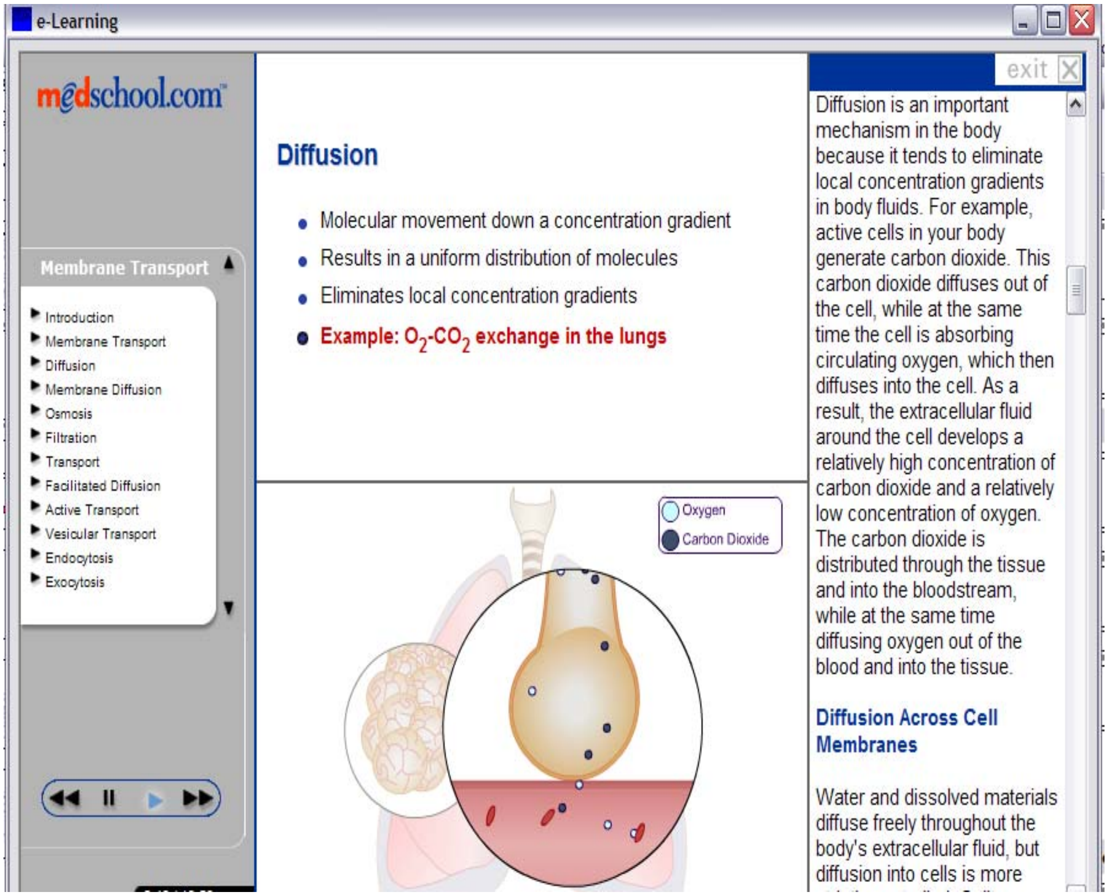

Figure 1: Web based model

An essential apparatus in the study is the CFS-Technologies Speakonia, a Text-To-Speech (TTS) Graphical User-Interface (GUI) Program which uses the Microsoft ® Speech Technology. It is a user friendly GUI which is comparable to Notepad. The reading can be paused, resumed as well as exported to a wave file. It allows Clipboard Reading, i.e., whenever text is copied to the clipboard Speakonia reads it aloud.

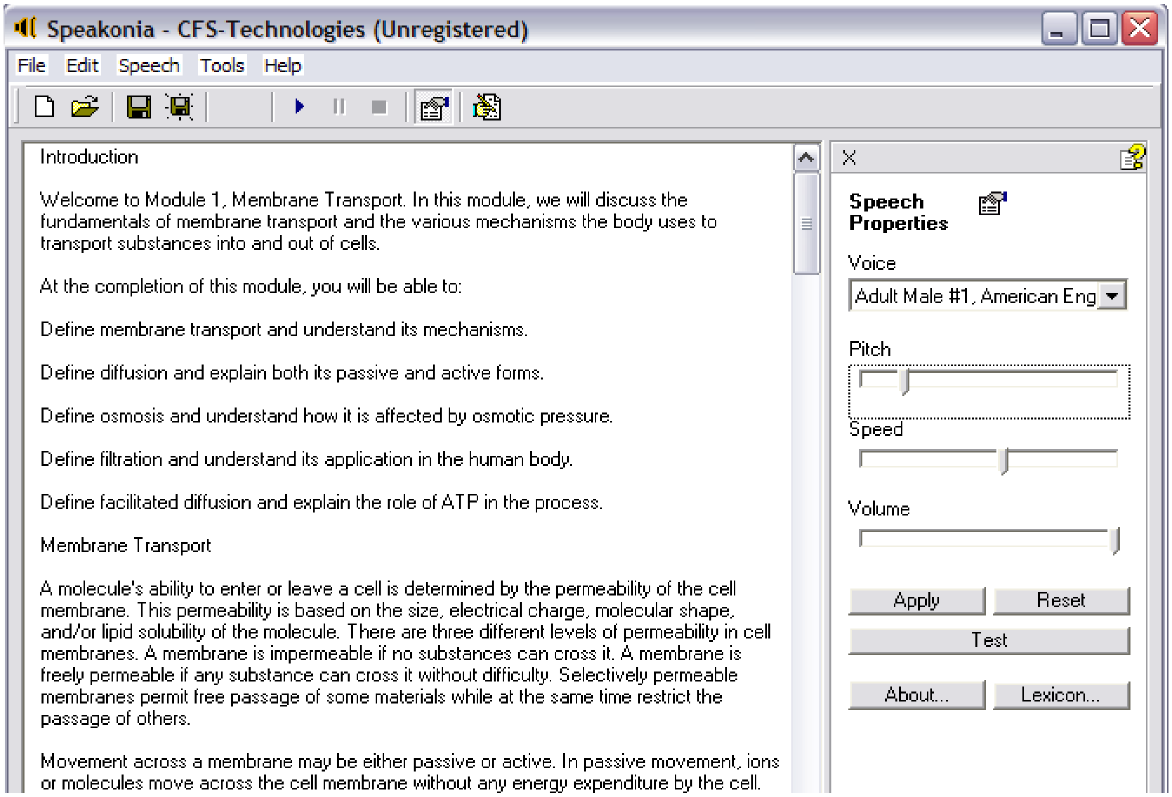

Figure 2: CFS-Technologies Speakonia, Text-To-Speech

3.2 Procedure

First, the participants were brought to the Institute for Simulation and Training (IST) where the experiment was conducted and asked to fill out an informed consent (Appendix A). They were then required to fill out a demographic survey (Appendix B) followed by a web tutorial lasting 3 minutes and 33 second in order to orientate them with the features of the E-learning platform.

The three groups under study were the Control Group as well as Experimental Group one and two. The population size of Experimental Group 1 and the Control Group was 20. The initial delivery speed for both the control and experimental group 1 was 128 words per minute. The Control Group did not have the option to change the speed, while the Experimental Groups could either increase or decrease the speed. Experimental Group 1 had the same properties as Experimental Group 2, but the latter received the initial delivery speed of 148 words per minute with a population size of 10.

Next, as participants appeared for the study, they were assigned to one of the three groups until testing was completed. The participants were required to use the desktop and interact with the web-based instructional modules on Human Anatomy & Physiology developed by Medsn, Inc with no sound/voice coming from the desktop computer. The text from the E-Learning Module of Medsn Inc. was copied to the clipboard of Speakonia in the notebook. The audio-based text is a synthetic voice which is delivered through the speaker system. The Figure below shows the participant using the web-based modules with a desktop computer and the Speakonia text to voice converter.

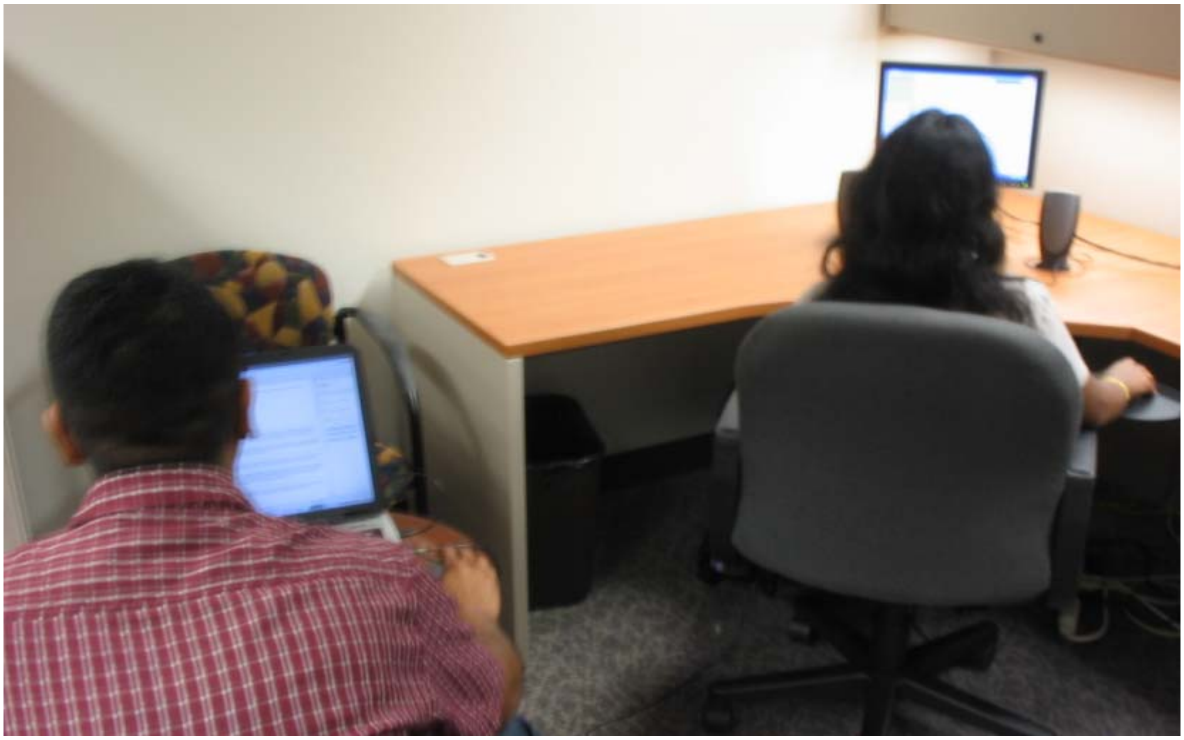

Figure 3: Experimental Setup

After the participants familiarize themselves with the features of E-learning platform and synthetic sound the participants begin the first session which is Membrane Transport. In this session they learn four modules pertaining to Membrane Transport such as Diffusion, Osmosis, Filtration and Facilitated diffusion. There is no time limit in reviewing the mechanisms involved with membrane transport. After the participants completed the Membrane Transport module they were asked to take 10-question multiple choice test (Appendix D).

CHAPTER FOUR: RESULTS

Results for quantitative measures of the multiple-choice test scores for the three groups are shown in the below figures. Figure 4 shows the multiple-choice test responses for Experimental Group-1. In Experimental Group-1 the mode was 9 with a minimum score of 5 and maximum of 10. The mean score for this group was 8.4 with a standard deviation of 1.31.

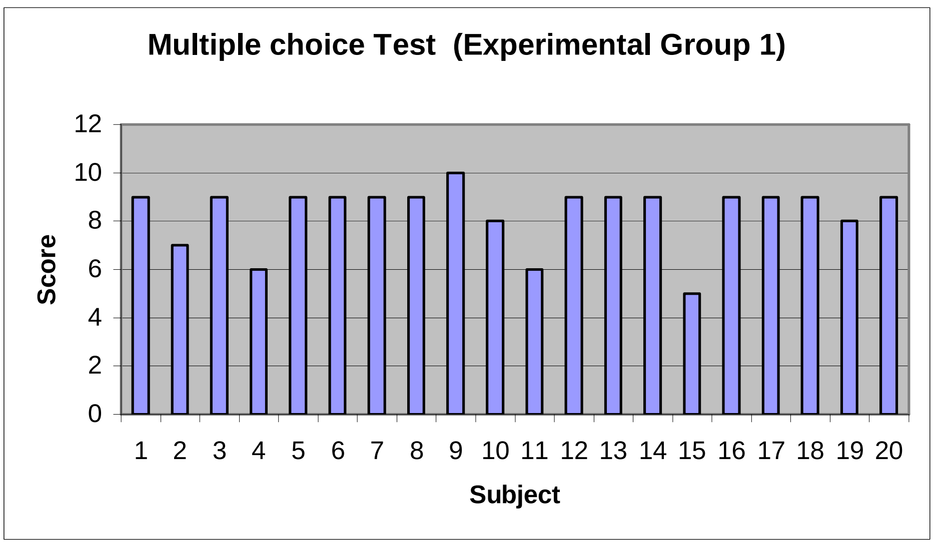

Figure 4: Correct responses for Experimental Group 1

The highest speed used by the participants was 195 words per minute and the lowest was 128 words per minute with a mean of 135.6 words per minute.

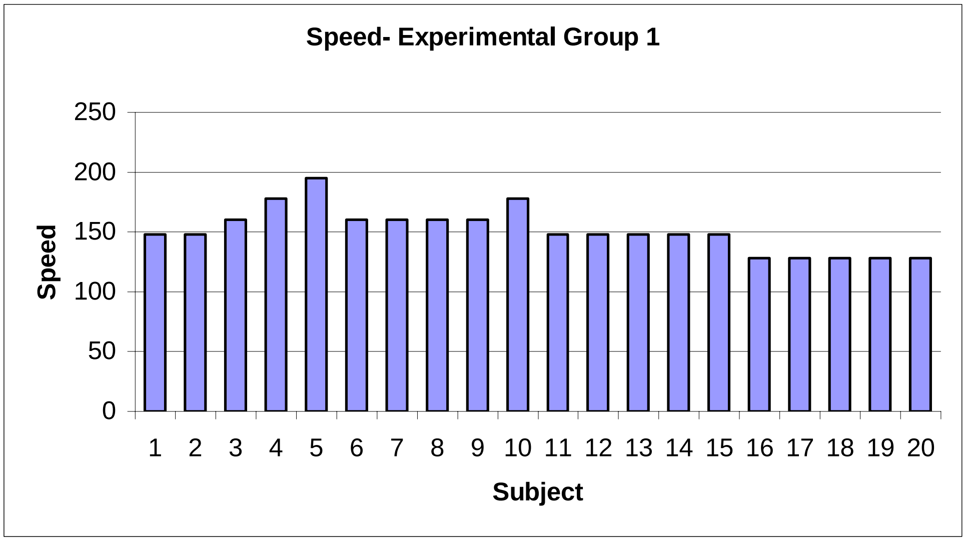

Figure 5: Speed for Experimental Group 1

The maximum and minimum score for Experimental Group-2 was ten and six. The mean score was 8.6 with standard deviation of 1.26.

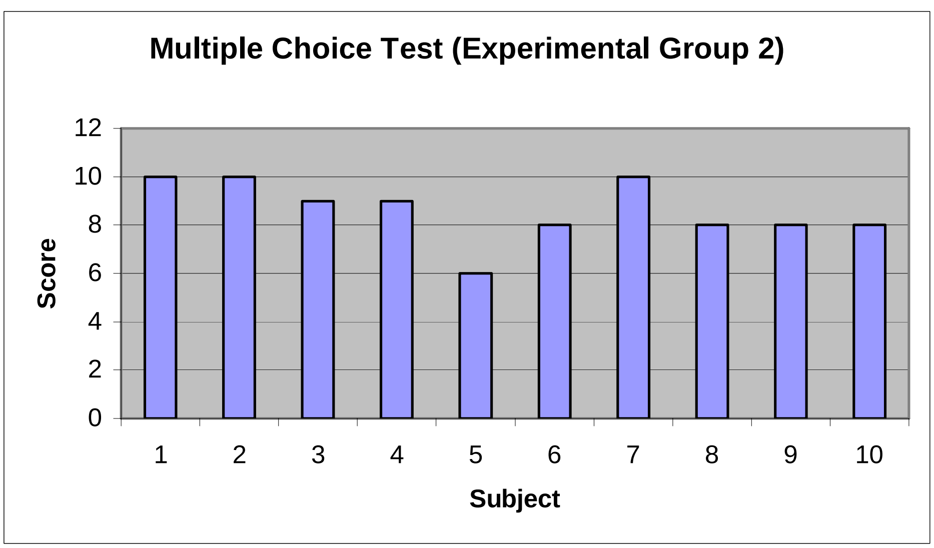

Figure 6: Correct responses for Experimental Group 2

As shown in the below figure five participants of Experimental Group-2 decreased the speed while one participant increased the speed (initial speed of 148 words per minute).

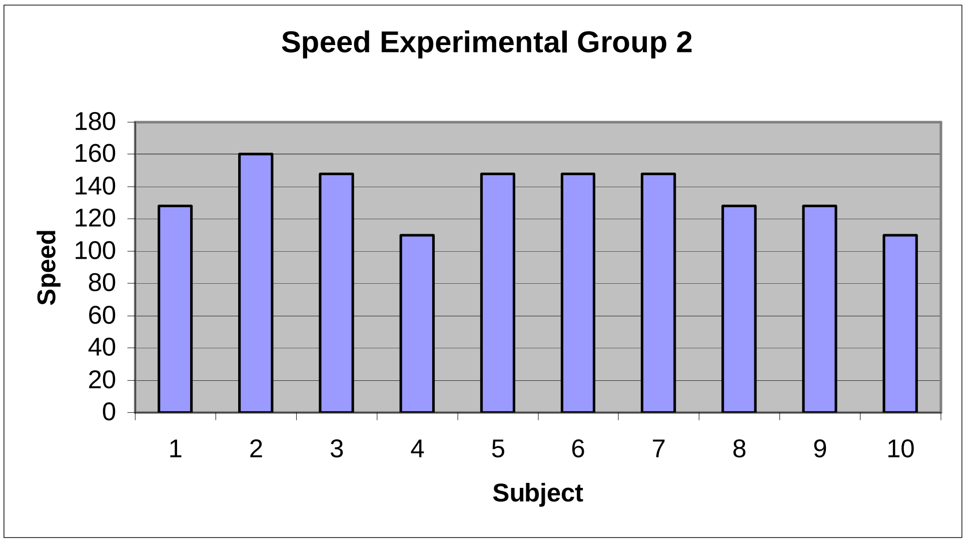

Figure 7: Speed for Experimental Group 2

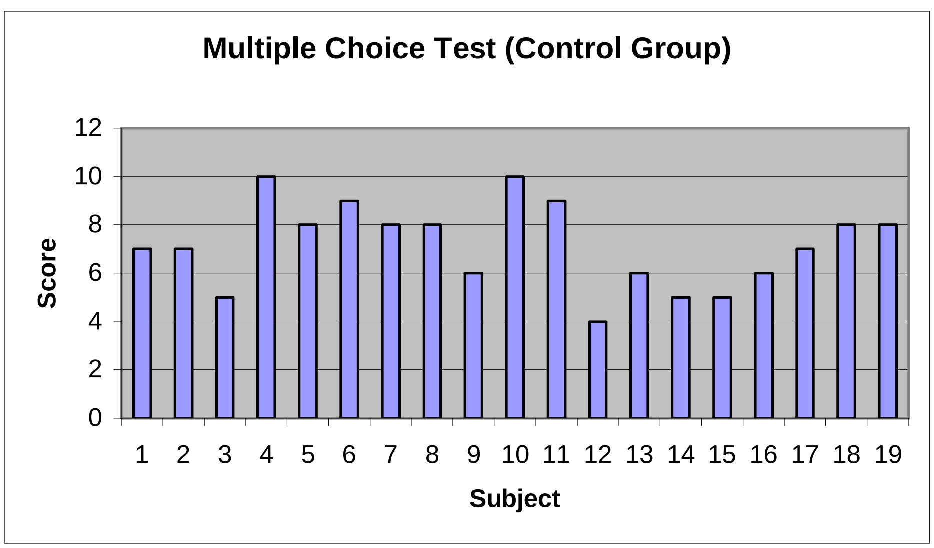

Figure 8: Correct responses for Control Group

The results of the Control Group are considerably lower as compared to Experimental Group-1 and Experimental Group-2. For the Control Group, the mean score was 7.2 with standard deviation of 1.69. The maximum and minimum score was ten and four respectively.

Table 1**:** Mean and Standard deviation analysis of three groups

Group	**Mean**	**Std. Deviation**	**Population Size**
Control Group	7.15	1.694	20
Experimental Group 1	8.35	1.309	20
Experimental Group 2	8.60	1.265	10

4.1 Univariate Analysis of Variance

Table two provides the output of a standard ANOVA summary table. From the below table we have $F(2, 47) = 4.67$, $p=0.014$, $\eta^2 = 0.166$.

Table 2: Standard ANOVA summary table

Source	**Type III Sum of Square**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Square**
GROUP	20.180	2	10.090	4.672	0.014	0.166

4.2 Post Hoc Test

The multiple comparisons table (Table 3) is shown below.

Table 3: Post Hoc Test – Multiple comparisons

Group(I)	**Group(J)**	**Mean Difference (I - J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower Bound	**Upper Bound**
Control Group	Experimental Group 1	-1.20	0.465	0.034	-2.32	-0.08
	Experimental Group 2	-1.45	0.569	0.037	-2.83	-0.07
Experimental Group 1	Control Group	1.20	0.465	0.034	0.08	2.32
	Experimental Group 2	-.25	0.569	0.899	-1.63	1.13
Experimental Group 2	Control Group	1.45	0.569	0.037	0.07	2.83
	Experimental Group 1	.25	0.569	0.899	-1.13	1.63

4.3 Chi-square Test

Table 4: Chi-square Test – Comparison between Experimental Group-1 and Experimental Group-2

	Experimental Group-1	Experimental Group-2	Total
Speed Changed	15	6	21
Speed not Changed	5	4	9
Total	20	10	30

In the above Chi-square test the degree of freedom is 1. The value of Chi-square is 0.7142. For the significance at 0.05 levels Chi-square should be greater then or equal to 3.84. Therefore there is no significant difference between Experimental Group-1 and Experimental Group-2 and the distribution is not significant. Here the p-value is less then or equal to 1.

As shown in Figure 9, the values in overall impression, the amount of listening effort and voice pleasantness shows the participant's difficulty in understand the synthetic voice. Apart from lacking Naturalness it was observed that pertaining to the synthetic voice the concatenation of isolated words degrades intelligibility and naturalness of synthetic speech.

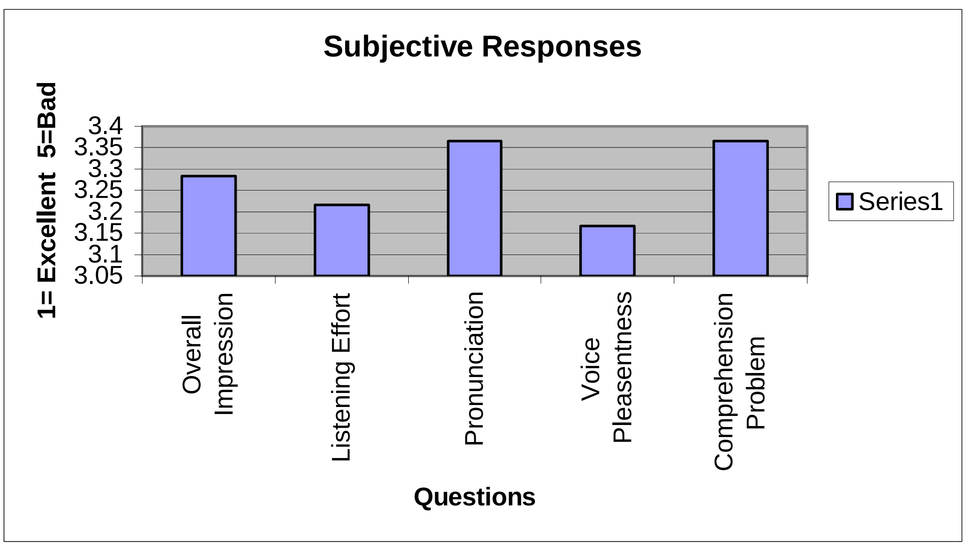

Figure 9: Participant response for synthetic voice

CHAPTER FIVE: DISCUSSION AND CONCLUSION

Although text clearly remains dominant, designers have to cope with the limited space available on a screen. Likewise, large color monitors are available and remain expensive as well as require high-end computers. Most programs that use monitors have a screen area of about two-thirds of an A4 page or less. This generates crowded displays when the text and graphics are displayed together. Procedures such as multiple screens, scrolling fields, windowing, abbreviated point-form 'bullets', and reduced font size are used to cope with this difficulty. If the text relates to a graphic, as is often the case, then work on split attention effects has shown that it is clearly not good design to have to switch or 'page' between graphics and text (Chandler and Sweller, 1992).

According to Fulford (1993) simple speeded speech (synthesized and speeded human voices) found that either normal recorded speech seemed slow after about 30 minutes as compared to speech presented at about 175% of the normal speed. There is also a suggestion that attention or concentration on the task is improved under Time-compressed Speech presentation.

At the time of doing the experiment there were occasions where both the computers were not synchronized with each other. It was resolved by pause and play for few seconds, either desktop or notebook. Controller was responsible for doing all the possible adjustment.

The limitations of this study can be examined using Kirkpatrick’s four level approach to the evaluation of training programs. In this study the experiment only measured levels one and two. Level (1) measures satisfaction and can be demonstrated through the qualitative results of the subjective questionnaires. Five attributes were measured on a 1 to 5 scale with 5 being the highest. The Level (2) assesses the amount of information learned. This is demonstrated by the quantitative results which illustrate impressive advantage of changing the speed as compared to not changing the speed during the test. Levels three and four were not measured. Not many studies relating to the effectiveness of web-based training and specifically the effect of speed of voice speed in multi-media presentation are available. However it is clear from the results of this study that giving the learner control of the speed of voice presentation is a valuable instructional feature.

APPENDIX A: INFORMED CONSENT

Consent

February 22, 2005

Dear Student:

My name is Alpesh Makwana and I am a graduate student working under the supervision of faculty member, Dr. J. Peter Kincaid. You are being asked to participate in an experiment designed to gather information on how performance is affected by changing or not changing the speed of hearing by using text to speech converter. This research project was designed solely for research purposes and no one except the research team will have access to any of your responses. All responses will be kept confidential. Your identity will be kept confidential using a numerical coding system.

Your participation in this project is voluntary. You do not have to answer any question(s) that you do not wish to answer. Please be advised that you may choose not to participate in this research, and you may withdraw from the experiment at any time without consequence. Non-participation will not affect your grade. There are no other direct benefits or compensation for participation. This experiment will take approximately 45 minutes outside of your regularly scheduled class time. There are no anticipated risks associated with participation.

If you have any questions or comments about this research, please contact Alpesh Makwana or his faculty supervisor, Dr. J. Peter Kincaid, Institute for Simulation and training, Orlando, FL; (407) 882- 1330. Questions or concerns about research participants' rights may be directed to the UCFIRB office, University of Central Florida Office of Research, Orlando Tech Center, 12443 Research Parkway, Suite 207, Orlando, FL 32826. The phone number is (407) 823-2901.

Sincerely,

Alpesh Makwana

____________ I have read the procedure described above.

____________ I voluntarily agree to participate in the procedure and I have received a copy of this description.

______ I would like to receive a copy of the procedure described above.

______ I would not like to receive a copy of the procedure described above.

____________________________________ / __________________

Participant Date

APPENDIX A: DEMOGRAPHIC SURVEY

Demographic Information

Occupation: ____

Highest level of education:_________________________________

Department: ____

Major: ____

Gender: Male/Female (Please Circle one)

Age (optional): ____________

Date (MM/DD/YYYY):____/____/________

Do you have any hearing problem?

Yes / No

If yes, please describe it.

____

____

Mother tongue: ____________________

Which are the other languages that you know very well?

1) ________________ 2)________________3)________________4)________________

APPENDIX C: MEMBRANE TRANSPORT QUESTIONNAIRE

Membrane Transport and Osmosis Questions:

Select the best answer choice

1. The permeability of the cell membrane is based on

a. Color, shape, and temperature of the molecule
b. The presence of flagella on the cell
c. Size, shape and electrical charge of the molecule
d. The presence of cilia on the cell membrane

2. The different levels of permeability in cell membranes are

a. Impermeable and selectively impermeable
b. Selectively permeable and selectively impermeable
c. Impermeable, freely permeable and selectively permeable
d. None of the above

3. The process by which ions or molecules move across the cell membrane using the energy from the cell is known as

a. Passive movement
b. Active movement
c. Diffusion
d. All of the above

4. The movement of molecules and ions from an area of relatively high concentration to an area low concentration, due to a concentration gradient is known as

a. Metastasis
b. Transfusion
c. Infusion
d. Diffusion

5. Movement of water across a freely permeable membrane to maintain an equilibrium is known as

a. Carrier mediated transport
b. Osmosis
c. Vesicular transport
d. None of the above

6. The amount of pressure that must be applied to prevent osmosis across a membrane is known as

a. Osmotic pressure
b. Intracranial pressure
c. Vesicular pressure
d. None of the above

7. The process by which water and small molecules are forced through a membrane by a hydrostatic pressure gradient is known as

a. Osmosis
b. Vesicular transport
c. Filtration
d. Metastases

8. With regard to carrier-mediated transport, all of the following are true except

a. Membrane proteins bind specific ions or organic substrates and move them across the cell membrane
b. This movement can be passive or active
c. Facilitated diffusion and active transport are the two major types
d. This is not an important type of membrane transport mechanism

9. ATP and enzymes provide the energy needed to move ions and molecules across the cell membrane in the case of

a. Active transport

b. Passive transport
c. Active and passive transport
d. None of the above

10. Which of the following ions is <u>not</u> carried by active transport across the membrane:

a. Sodium.
b. Magnesium
c. Lithium
d. Potassium

APPENDIX D: SUBJECTIVE QUESTIONNAIRE

SUBJECTIVE QUESTIONNAIRE

Please tick-mark one of the available choices:

Over all impression

How do you rate the quality of the sound of what you just heard?

□Excellent □Good □ Fair □ Poor □ Bad

Listening effort

How would you describe the effort were required to make in order to understand the message?

- □Complete relaxation possible; no effort required
- □Attention necessary; no appreciable effort required
- □Moderate effort required
- □Effort required
- □No meaning understood with any feasible effort.

Pronunciation

Did you notice any anomalies in pronunciation?

- □No
- □Yes, but not annoying
- □Yes, slightly annoying
- □Yes, annoying
- □Yes, very annoying

Speaking rate

The average speed of delivery was:

- □Much faster than preferred
- □Faster then preferred
- □Preferred
- □Slower than preferred
- □Much slower than preferred

Voice pleasantness

How would you describe the voice?

- □Very pleasant
- □Pleasant
- □Fair
- □Unpleasant
- □Very unpleasant

Comprehension problem

Did you find certain words hard to understand?

- □Never
- □Rarely
- □Occasionally
- □Often
- □All of the time

Acceptance

Do you think that this voice could be used for any information service?

- □Yes
- □No

Please rate the instructional material that you have just used

1 = Highest rating

2 = Lowest rating

Quality of illustration

1	2	3	4	5

Defining the instructional objective

1	2	3	4	5

Presentation of main points

1	2	3	4	5

User friendliness

1	2	3	4	5

Retention of Material

1	2	3	4	5

Will you improve if you are given a choice to increase/decrease speed?

1	2	3	4	5

Will you improve if you are given a choice to increase/decrease?

1	2	3	4	5

Will you improve if you are given a choice to change the voice?

1	2	3	4	5

APPENDIX E: IRB

University of Central Florida

Office of Research

April 12, 2005

Alpesh Makwana
11925 Pasteur Drive
Orlando, FL 32826

Dear Mr. Makwana:

With reference to your protocol #05-2431 entitled, "Voice Track Computer Based Simulation" I am enclosing for your records the approved, expedited document of the UCFIRB Form you had submitted to our office. **This study was approved by the Chairman on 4/11/05. The expiration date for this study will be 4/10/06.** Should there be a need to extend this study, a Continuing Review form must be submitted to the IRB Office for review by the Chairman or full IRB at least one month prior to the expiration date. This is the responsibility of the investigator. **Please notify the IRB when you have completed this study.**

Please be advised that this approval is given for one year. Should there be any addendums or administrative changes to the already approved protocol, they must also be submitted to the Board through use of the Addendum/Modification Request form. Changes should not be initiated until written IRB approval is received. Adverse events should be reported to the IRB as they occur.

Should you have any questions, please do not hesitate to call me at 407-823-2901.

Please accept our best wishes for the success of your endeavors.

Cordially,

Barbara Ward

Barbara Ward, CIM
IRB Coordinator

Copy: IRB file

12443 Research Parkway • Suite 302 • Orlando, FL 32826-3252 • 407-823-3778 • Fax 407-823-3299
An Equal Opportunity and Affirmative Action Institution

REFERENCES

[1] Kincaid, J.P., Bala, J., Hamel, C, Sequeira, W.J., and Bellette, A. (2001). "Effectiveness of Traditional vs. Web-based Instruction for Teaching an Instructional Module for Medics". IST-TR-01-06. Orlando: Institute for Simulation and Training, University of Central Florida.

[2] Martini, F. H., and Bartholomew, E. F. (2000). "Essentials of Anatomy and Physiology". (2^{nd} Edition). Upper Saddle River, NJ: Prentice-Hall.

[3] Instruction J. Bala, M.D., J. Peter Kincaid, Ph.D., Cheryl Hamel, Ph.D. Tao Le, M.D., Balinder Sahota, MBA. "Teaching Human Anatomy and Physiology: Effectiveness of Traditional vs. Web-Based "

[4] M. Dolson. The phase vocoder: A tutorial. Computer Music Journal, 10(4): 14-27, 1986

[5] D.W. Griffin and J.S. Lim. Signal estimation from modified short-time fourier transform. IEEE Transactions on Acoustics, Speech, and Signal processing. ASSP-32 (2):236-243, April 1984.

[6] D.J. Hejna Jr. Real-time time-scale modification of speech via the synchronized overlap-add algorithm. M.I.T. Masters Thesis, Department of Electrical Engineering and Computer Science, February 1990.

[7] D. Malah Time domain Algorithms for harmonic bandwidth reduction and time scaling of speech signals. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-27(2): 121-133, April 1979.

[8] J. Makhoul and A.El-Jaroudi. Time-scale modification in medium to low rate coding. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pages 1705-1708. IEEE, 1986.

[9] G.A. Miller and J.C.R. Licklider. The intelligibility of interrupted speech. Journal of the Acoustic Society of America, 22(2): 167-173, 1950.

[10] E.P. Neuburg. Simple pitch-dependent algorithm for high quality speech rate changing. Journal of the Acoustic Society of America. 63(2):624-625, 1978.

[11] S.Roucos and A.M. Wilgus. High quality time-scale modification for speech. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pages 493-496. IEEE, 1985.

[12] R. J. Scott and S.E. Gerber. Pitch-synchronous time-compression of speech. In Conference on Speech Communication and Processing, pages 63-65. IEEE, 1972.

[13] T.G. Stitch. Compression of repeated time-compressed recordings. The Journal of Experimental Education, 37(4), Summer 1969.

[14] J.L. Wayman, R.E. Reinke, and D.L. Wilson. High quality speech expansion, compression, and noise filtering using SOLA method of time scale modification. In 23rd Asilomar Conference on Signals, systems, and Computers, pages 714-717, October 1989 Vol.2.

[15] Web Chi-square Calculator http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html