Development and Use of Simulation Modules for Teaching a Distance-Learning Course on Digital Processing of Speech Signals

John N. Gowdy, Eric K. Patterson, Duanpei Wu, and Sami Niska, Clemson University

Abstract

This paper describes the development, utilization, and effectiveness of a set of interactive simulation modules used with a distance-learning version of a graduate class on Digital Processing of Speech Signals at Clemson University. The motivation for developing these modules was to give distance-learning students a special learning tool, to partially compensate for their disadvantages of being separated from campus laboratories, the course instructor, and other students. Although the modules were developed primarily for students taking the course remotely, they have also enhanced the learning experiences of on-campus students. The paper describes the following simulation modules: phoneme segmentation, speech production model, speech coding, speech synthesis using LPC, speech recognition, and speech analysis. The authors have also developed a system for homework distribution, submission, and grading over the Web. This includes JAVA-based tools that permit students to generate and submit diagrams as part of their homework solutions.

A UNITED ENGINEERING FOUNDATION CONFERENCE
Davos, Switzerland 11-16 August 2002
http://www.coe.gatech.edu/eTEE

I. Introduction

Teaching a class using a distance-learning format presents new challenges to the instructor. This is due to the loss or reduction of traditional communication links between students and instructor. Distance-learning students do not have the opportunity of traditional students to visit a professor in his or her office to discuss difficult material and to work out problems. In addition, distance-learning students typically do not have access to special lab equipment or software that on-campus students may have. Furthermore, distance-learning students have reduced opportunities for group interaction with other students.

Instructors can partially compensate for the above restrictions by developing special learning tools that will assist the distance-learning student in understanding the course material. Hands-on, interactive models have been found to be very effective in this regard.

II. Background

Six simulation modules have been developed to provide distance-learning students with an educational resource. These tools were initially developed using C for an X Windows environment on Sun workstations. Because some distance-learning students did not have ready access to Sun computers, and because of system configuration problems for the Suns that were available, the modules were later modified for implementation on standard PC's running under the Linux operating system. This has proven to be an important advancement. The modules have been used with one live TV offering of ECE 846 – Digital Processing of Speech Signals and will be used again in this mode in Fall 2002.

Another difficulty in teaching a course remotely is the problem of submitting, grading, and returning homework. Although fax machines can be used, this mode can present many frustrations due to busy or malfunctioning fax machines. The use of scanners is another possibility, but these may not be available, especially if the student is traveling, as distance students often are. Therefore, we found the need to develop an efficient method for students to submit homework over the Web. A tool developed for this purpose is described in Section V.

III. Speech Processing Modules

The student is provided with a very simple user interface to access the various modules. The initial selection screen is shown in Figure 1.

Brief descriptions of the modules are provided below:

A. Phoneme Segmentation Module

This module permits the student to observe the time domain representation of a speech signal and select a speech segment for analysis. The main use of this module has been to observe the basic features of voiced and unvoiced speech and to assist the students in segmenting a speech segment into phonemes.

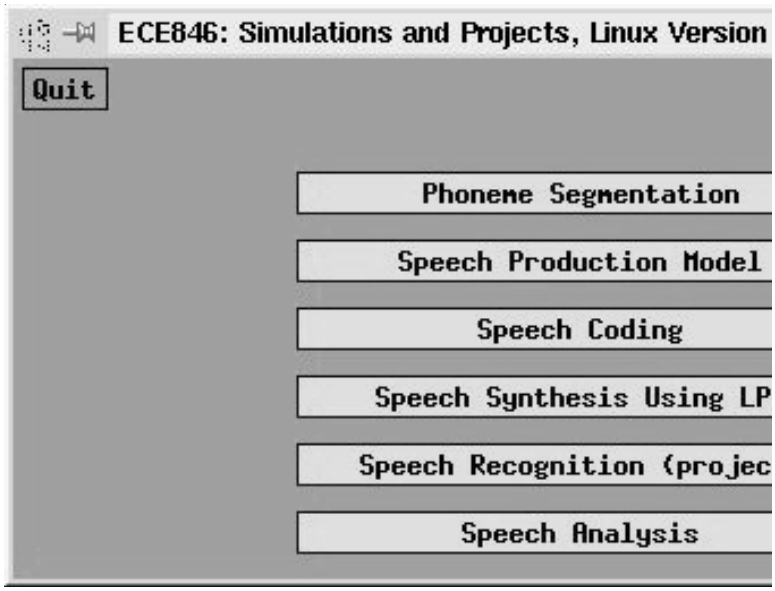

Figure 1. Initial Selection Screen

For example, the student can easily see the quasi-periodic nature of voiced speech and the noise-like characteristics of unvoiced speech. The contributions of pitch to the voiced speech waveform can also be observed.

Using this module, a student can also highlight a portion of the signal to be played back. This provides an interactive method for accurately selecting boundaries between phonemes in a speech utterance.

B. Speech Production Module

This module is based on the speech production model, which represents the vocal track as a concatenation of lossless tubes. The cross-sectional area of these cylindrical tubes can be used as parameters for representing speech. As the cross-sectional areas of the various tubes (up to 8) are changed, the transfer function of the vocal track model also changes. In particular, the resonant frequencies (formants) of the vocal track and their corresponding bandwidths change as the tube sizes are changed. This corresponds to a human speaker moving the throat, jaws, teeth, tongue, and uvula as he or she speaks.

The student can use the computer's mouse to select and adjust any of the tube sizes in this model. For any set of tube sizes, the frequency response of the resulting model can be displayed so that the student can observe interactively the effect of vocal track shape on formant frequencies and bandwidths. A screen display for using this module is shown in Figure 2.

C. Speech Coding Module

Speech coding is a procedure to reduce the number of bits needed to represent speech, within some tolerance of quality reduction, over a fixed time interval. Effective speech coding permits efficient transmission and/or storage of the speech signal. Several methods of speech coding are covered in our graduate course on Digital Processing of Speech Signals. This simulation module permits students to listen to the effects of several coding schemes applied to the same speech utterance. The coded waveform can also be visually compared with the original speech signal. Typical displays of this module are shown in Figure 3.

D. Speech Synthesis Using Linear Prediction Module

This module permits the students to analyze the speech signal by estimating linear prediction coefficients for the underlying speech model. These coefficients represent an estimate of the parameters of an all-pole model of speech production. Students can then use these parameters to resynthesize speech and to investigate how the number of predictor coefficients used affects the quality of the synthesized speech.

E. Speech Recognition Module

This module is still under development. The student will be presented with the basic structure of a recognition system and will be able to choose various options, such as choice of parameters. The user will be able to investigate the recognition performance for several implementation options.

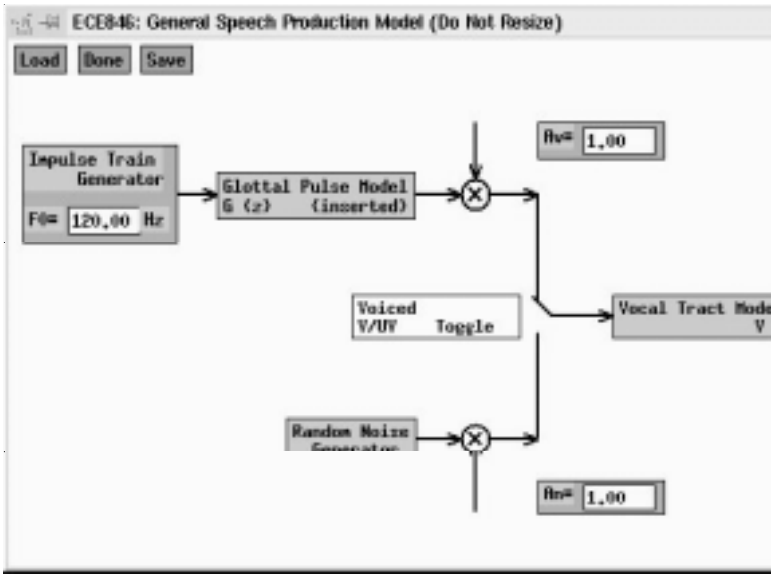

Figure 2. Typical Speech Production Module Display.

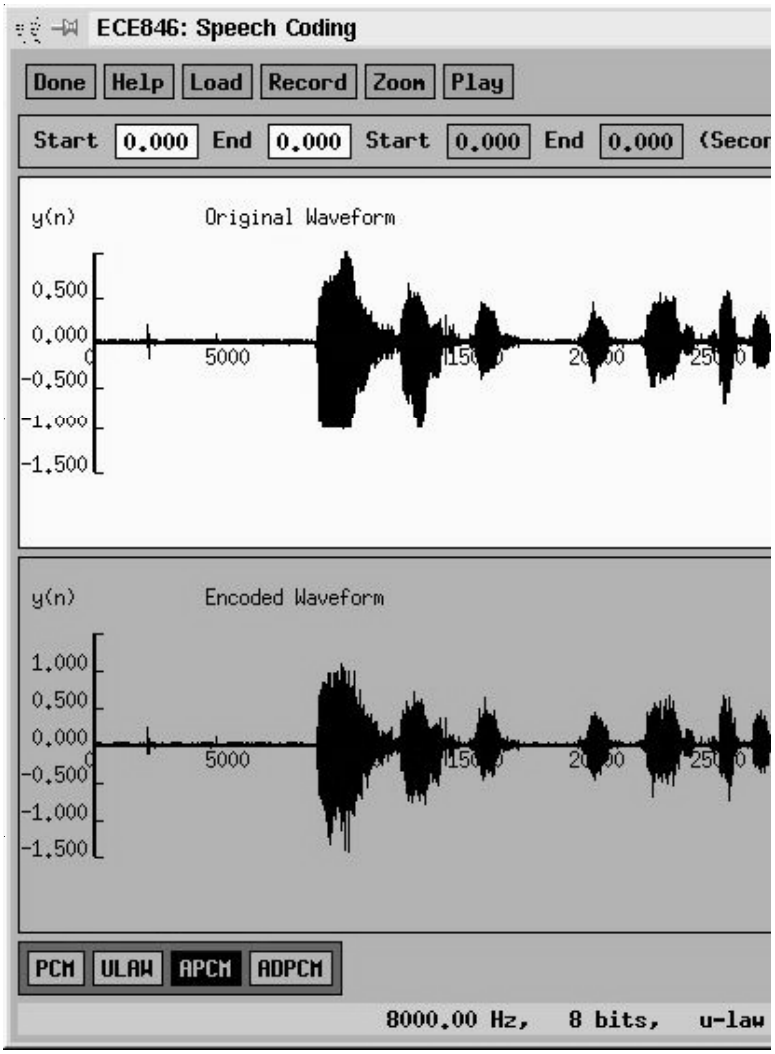

Figure 3. Typical Speech Coding Module Display.

F. Speech Analysis Module

This module was added several years after the original modules were implemented. It permits the student to extract several important features from the speech signal. These include: (1) zero crossing rate, (2) autocorrelation, (3) modified autocorrelation, (4) cepstum, and (5) linear prediction (LP).

1) *Zero Crossing Rate:* The zero-crossing rate is the percentage of speech samples having a change in polarity from the previous sample. It can be used to obtain a crude, but easily determined estimate of the highest frequency content of a signal. It provides a useful tool for determining the start and end of a speech utterance in the presence of background noise.

2) *Autocorrelation:* The autocorrelation function is useful for a number of speech processing tasks. For example, it can be used to determine whether a segment of speech is voiced or unvoiced. In addition, it serves as a preliminary step in more complex speech processing algorithms, such as linear predictive analysis. This module incorporates one of two possible algorithms for implementing the short-term autocorrelation function.

3) *Modified Autocorrelation:* This module implements another algorithm for estimating the short-term autocorrelation function for a speech segment. By comparing the output of this module with the output of the module described above, the student can visualize the different properties of two versions of the short-term autocorrelation function.

4) *Cepstrum:* The cepstrum is one of the most commonly used features for analyzing speech. It is essentially a deconvolution technique that separates the contribution of the excitation signal from system properties of the vocal track. The obtained parameters can be used for a number of speech-related tasks including speech recognition, speaker identification, and speech coding. By observing the cepstrum for a speech segment, the student can visualize the effectiveness of the deconvolution goal of this algorithm. A typical output display of this module is shown in Figure 4.

5) *Linear Predictive (LP) Analysis:* Another valuable set of parameters for speech analysis can be obtained by linear predictive analysis of the speech signal. As already described, this method attempts to find the best set of coefficients for an all-pole model for the speech production system. Like the cepstrum method, it can be used to obtain a characterization of speech which represents the vocal tract properties alone (with the contributions of the excitation signal removed.) This module permits the user to view the LP spectrum of speech, which is the estimated frequency response of the vocal track model. Students can observe formant frequencies and formant bandwidths from the displays of this module. Like the cepstrum, LP parameters are useful for many applications, including speech recognition, speaker identification, and speech coding.

IV. Experience in Using Modules

Students have found the modules very easy to use and very effective in supplementing the course material. Although a small user's handbook has been developed to assist students in utilizing these modules, the user interface is very intuitive and little supplementary information is necessary.

Although the textbook used with our course on Digital Processing of Speech Signals is a good one, it is short on real-world examples. The developed modules therefore provide valuable balance to the treatment of the text. Especially in a course topic like Digital Processing of Speech Signals where it is important in many cases to hear the processed speech, a module with sound capability has proven to be very effective. In addition, the interactive nature of the modules is very useful for helping students develop an intuitive feel for some of the key mathematical models of the speech signal.

A byproduct of developing tools to assist distance-learning students is that these same tools have also proven very popular with on-campus students.

V. JAVAGRAM

JAVAGRAM is a system we developed to give students a means of submitting home problems over the Web. Our first system, implemented in 1996, simply presented students with multiple-choice questions which they could answer with a click of the mouse. The answers were submitted directly over the Web and automatically graded when received. The next step in the homework system permitted students to enter text answers to questions. The following enhancement permitted students to draw boxes, lines, and circles, and to enter text labels. This kind of entry is typical for answering many engineering problems. We also have provided limited equation editing capability, including the use of mathematical operators and Greek symbols. Although this system needs additional development, it has proven to be a very useful tool for students to answer and submit a wide array of homework problems over the Web.

VI. Future Plans and Projections

The next step in the evolution of our simulation modules will be to re-development all the simulation modules as Java applets so they can be implemented over the Web. This should eliminate system compatibility problems that can sometimes exist when using PC's to implement the current software. In addition, using the Web will provide easy access to students with travel obligations while enrolled in a class. Web implementation would also

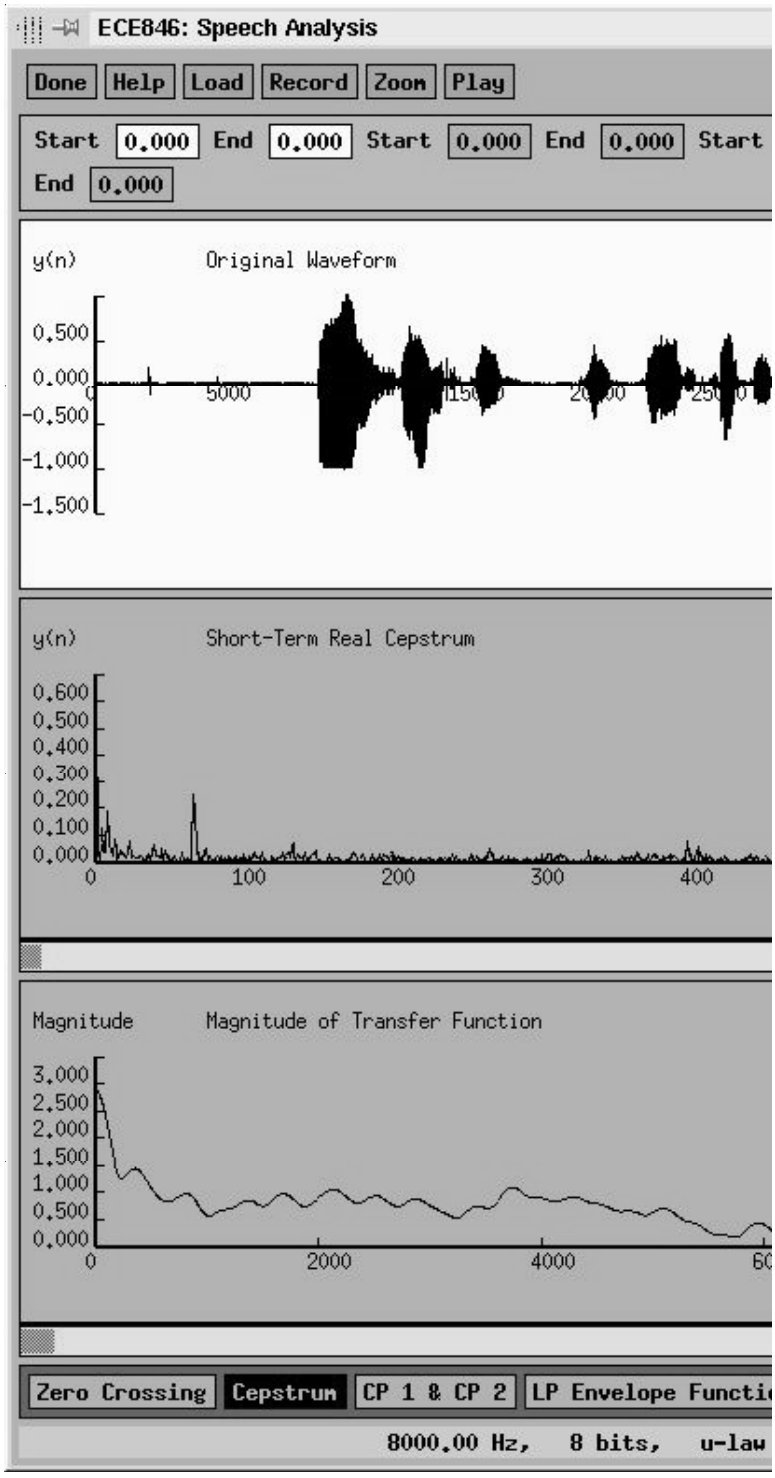

Figure 4. Typical Cepstrum Module Display.

permit students to interact with modules during presentation of a live TV class. Wireless Web access would be valuable for this kind of use.

We also plan to extend the JAVAGRAM system to make it even more powerful as a homework submission system. These plans include adding more options and capabilities for both diagram sketching and equation editing.

References

[1] Patterson E. K., Wu D., and Gowdy J. N., "Multi-Platform CBI Tools Using Linux and Java Based Solutions for Distance Learning," *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Seattle, WA.

[2] Deller J., Proakis J., and Hansen J., *Discrete-time Processing of Speech Signals*, Macmillan, New York, 1993.

[3] McClellan J., Shafer R., Schodorf J., and Yoder M., "Multi-Media and World Wide Web Resources for Teaching DSP," *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, Georgia.

[4] Beck M., Bohme H., Dziadzka M., Kunitz U., Magnus R., and Verworner D., Linux Kernel Internals, Addison-Wesley, 1996.

Authors' Biographies

John N. Gowdy is Professor and Chair of Electrical and Computer Engineering at Clemson University. He has been involved in research in speech signal processing since 1972. He has taught three distance-learning courses for the Clemson Telecampus system.

Eric K. Patterson received his Ph.D. degree from Clemson University in May 2002. His dissertation research was in the area of audio visual speech recognition. His graduate research was supported by a NASA GSRP Fellowship. In August 2002, he will become an Assistant Professor of Computer Science at the University of North Carolina at Wilmington.

Duanpei Wu received his Ph.D. degree from Clemson University in 1996. His dissertation was on a neural network based speech recognition system. He has performed speech research for Sony and now works for Cisco Systems.

Sami Niska attended Clemson University in 2000 as a visiting student from Finland and worked in the Speech and Audio Processing Lab.