CANADIAN-DESIGNED SOFTWARE FOR SPEECH ANALYSIS AND SYNTHESIS

B.C. Dickson, A.G. Wyarib, R.C. Snell, S.J. Eady, and J.A.W. Clayards

Speech Technology Research Ltd., Suite D, 1623 McKenzie Avenue, Victoria, B.C. V8N 1A6, Canada

Introduction

This paper describes a microcomputer-based system, called Computerized Speech Lab (CSL), which has been designed for analysis and synthesis of speech signals. The system, which runs on an IBM AT-compatible computer, consists of executable software modules and a hardware module that plugs into the host computer.

Hardware Module

The hardware for the system consists of an external module and a printed circuit board that fits into a slot in the host computer. The external module provides for signal conditioning and volume control on two channels. The external module connects to the printed circuit board, which has a DSP16A chip for 16-bit data acquisition and playback at sampling rates between 2.5 and 51.2 kHz on each channel. The DSP16A also handles digital filtering and downsampling in order to control aliasing. A TMS-320C25 digital signal processing chip is also included on the printed circuit board for high-speed data processing.

Software Modules

The software for this system has been designed to make a wide range of operations easily accessible to the user. The software operates in a windows-type environment using pull-down menus and a mouse. Commands may be entered by choosing from the selections in the pull-down menus, by user-defined function keys or by command strings input at the keyboard. Further flexibility is achieved by allowing the software to run through a series of command files or macros, which combine many commands into one instruction.

System Features

The system provides the user with numerous features for handling speech signals:

1. Signal Acquisition and Playback

   The software provides capabilities for dual-channel speech acquisition, disk storage, retrieval and playback with user-selected sampling frequencies.

2. Speech Editing

   The program also provides capabilities for speech editing, including mixing, subtracting, digital filtering, amplitude scaling, time warping, appending, splicing and downsampling. Editing also includes the ability to apply window weighting to the edges of splices for glitch-free cuts.

3. Speech Analysis

   The software includes routines for analysis of energy, pitch, FFT spectrum, LPC frequency response, LPC formant history and colour or grey-scale spectrograms. Results of analyses can be displayed in graphic or numerical modes. An example of a spectrogram analysis is displayed in Figure 1.

4. Phonetic Transcription

   In addition, there is a Phonetic Transcription feature for producing International Phonetic Alphabet (IPA) standardized character sets which are time-linked to the speech waveform. An example of such a phonetic transcription is shown in Figure 1.

FIGURE 1: Graphic display from the CSL program for the word "cod", showing the speech waveform (top), IPA transcription (middle) and spectrogram (bottom).

5. Parameter Manipulation

   The LPC Parameter Manipulation/Synthesis module (also called ASL) provides a number of different speech processing and data manipulating capabilities which allow the user to investigate the properties of a speech signal. This module uses linear predictive coding (LPC) for analysis and resynthesis of speech data. Synthesis is of high quality using the residual-excited pitch-synchronous LPC approach.

Parameters that can be manipulated in this module include energy, pitch, duration, formant frequencies and formant bandwidths. Parametric values resulting from LPC analysis may be examined and edited in three different modes. In addition, waveforms resulting from analysis and synthesis may be compared in a fourth mode. Three of these four modes are illustrated in Figures 2, 3 and 4.
Figure 2 shows a sample display from the Waveform mode of the ASL module. The top waveform in this figure is the original digitized speech data that has been produced by a male speaker. The vertical striations beneath the waveform illustrate the locations of pitch pulses during voiced portions of the speech signal. These pitch locations are used as frame boundaries for the pitch-synchronous LPC analysis that has been performed on this data. The LPC residual waveform resulting from analysis is displayed in the middle viewscreen of Figure 2. This waveform can be used as the excitation for resynthesis of the speech data. The waveform at the bottom of the figure is the result of resynthesizing the speech data using the residual-excitation approach. Note that, since none of the synthesis parameters has been changed in this case, the two waveforms in the top and bottom viewscreens of Figure 2 should be identical. In fact, listening tests show that the two signals are perceptually indistinguishable.

Figure 3 illustrates the Formants mode of the ASL module. This mode provides the means for graphic editing of the parameter contours that result from LPC analysis. The figure displays the original speech waveform at the top, the pitch contour at the bottom and a time-history display of formant frequencies in the middle. Pitch and formant parameters can be modified by redrawing the pertinent contours using a mouse. Resynthesis of a new speech signal can be done quickly to provide immediate feedback to the user.

Figure 4 displays the Numeric Values mode of the ASL module. As can be seen in the figure, this mode provides a waveform display at the top and a list of numerical values resulting from LPC analysis at the bottom. Each column of data represents a separate LPC parameter. The parameters include residual frame number, energy, fundamental frequency, frame length, formant frequencies and formant bandwidths. Each of these parameters can be modified by entering new numerical values on a frame-by-frame basis or by marking a column and smoothing between endpoints. As in the Formants mode, resynthesis following parameter manipulation provides instant feedback on the effects of parameter changes.

Summary

The Computerized Speech Lab is a microcomputer-based system that provides a wide range of functions for the analysis and synthesis of speech signals. The software for this system has been designed and developed in Canada and is currently being sold to speech researchers around the world.

Acknowledgement

The software described here was developed by Speech Technology Research Ltd., under contract to Kay Elemetrics Corp.