CANADIAN-DESIGNED SOFTWARE FOR SPEECH ANALYSIS AND SYNTHESIS

B.C. Dickson, A.G. Wynrib, 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.
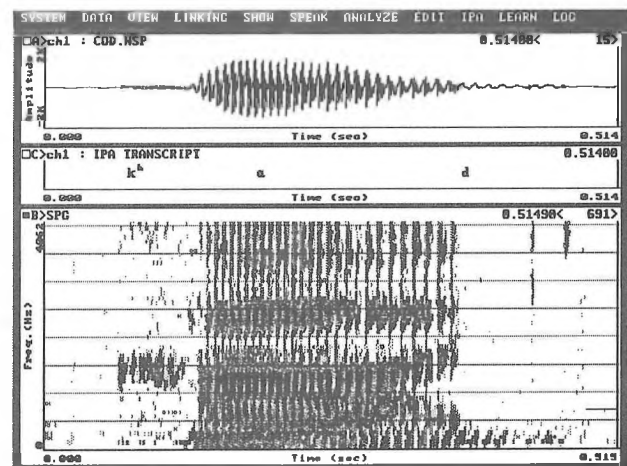

FIGURE 1: *Graphic display from the CSL program for the word "cod", showing the speech waveform (top), IPA transcription (middle) and spectrogram (bottom).*

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.
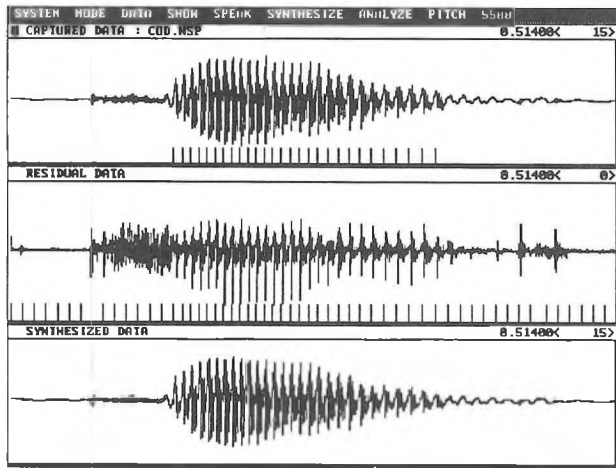
N	M	RES	PK	FM	LEN	F1	B1	F2	B2	F3	B3	F4	B4
1	8	1	46	8	100	1568	306	2759	396	4154	301	8	8
2	8	2	35	8	100	1046	737	3164	652	4943	868	8	8
3	8	3	8	8	100	1318	359	2257	515	3305	288	4673	291
4	8	4	28	8	100	1816	487	2231	562	3311	343	4832	534
5	8	5	21	8	100	1862	404	2326	688	3459	378	8	8
6	8	6	8	8	100	1827	232	2282	295	3381	289	4452	159
7	8	7	192	8	101	1364	48	1400	571	3359	442	4310	93
8	8	8	98	8	100	377	225	1445	55	2895	711	4116	187
9	8	9	148	8	100	434	173	1439	39	2473	573	3940	302
10	8	10	177	8	100	1348	53	1596	922	3199	583	4358	194
11	8	11	117	8	100	1317	35	2397	333	4863	261	8	8
12	8	12	111	8	100	1304	76	2172	327	3449	341	4443	184
13	8	13	258	8	100	131	48	1184	156	2299	59	3568	217
14	8	14	699	133	75	1161	63	2778	585	4287	89	8	8
15	8	15	839	149	67	1862	85	2354	169	3980	198	8	8
16	8	16	1041	136	73	247	161	1188	48	2357	179	3742	129
17	8	17	1213	142	78	346	143	1119	34	2351	131	3784	122
18	8	18	1338	147	68	368	117	1187	38	2368	170	3718	118
19	8	19	1385	142	78	386	114	1189	31	2366	155	3675	112
20	8	20	1498	147	68	396	95	1189	28	2336	143	3651	139
21	8	21	1479	149	67	399	85	1897	28	2318	114	3668	614

FIGURE 2: *Graphic display from the ASL module of the CSL program for the word "cod", showing the original speech waveform (top), the LPC residual waveform (middle) and the resynthesized speech waveform (bottom).*

FIGURE 4: *Graphic display from the ASL module for the word "cod", showing the speech waveform (top), and numerical values resulting from LPC analysis (bottom).*

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.
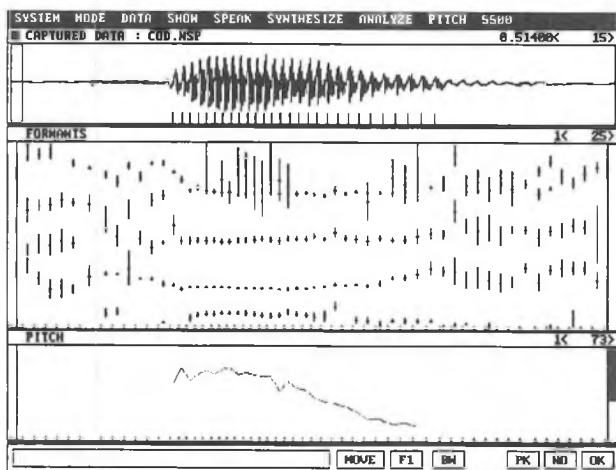
Figure 3 illustrates the **Formants** mode of the ASL module. This mode provides the means for graphic editing of the parameter contours that result from LPC analysis. The figure displays the original speech waveform at the top, the pitch contour at the bottom and a time-history display of formant frequencies 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.

FIGURE 3: *Graphic display from the ASL module for the word "cod", showing the speech waveform (top), formant frequencies derived from LPC analysis (middle) and pitch contour (bottom).*
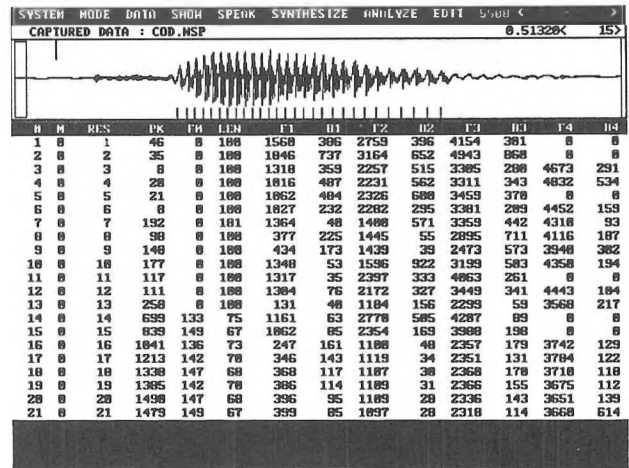
Summary

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