CSpeech(Version 3.1)

연세대학교 의과대학 이비인후과, 음성언어의학연구소
최 홍 식

= Abstract =

CSpeech is a software package that implements an audio waveform/speech analysis workstation on an IBM Personal Computer or hardware compatible computer. Features include digitizing audio waveforms on single or multiple channels, displaying the digitized waveforms, playing back audio waveforms from selected intervals of single channels, saving and retrieving waveforms from binary format disk files, and analysing audio waveforms for their temporal and spectral properties. The distinguishing characteristics of CSpeech are its support for multiple channels, minimal restrictions on sample rate and waveform duration, support for a variety of hardware configurations, fast graphics display, and its user−extensible menu−based command structure.

Introduction

CSpeech, pronounced "See-Speech," is the name for the computer speech waveform acquisition, display, editing and analysis software package. The name is intended as a pun because CSpeech is useful in visualizing temporal and spectrographic properties of the acoustic speech signal. CSpeech supports multi-channel waveform acquisition at equal sample rates on each channel. It is intended for making simultaneous measurements of the acoustic speech signal and signals derived from instruments such as the electroglottograph, neck wall accelerometer, airflow mask, and intraoral pressure sensor. While the analysis functions of CSpeech are tailored to the requirements of speech, CSpeech is useful in other applications requiring acquisition of audio bandwidth signals on multiple channels, such as physiological data recording.

Price

$ 1800 : CSpeech Version 3 Core(waveform display, recording, and playback) and Supplement(gray scale spectrogram, LPC formant tracking, waterfall spectrogram, pitch and voice perturbation analysis, spectrum moments, data conversion, and Pascal source code for data interface)

$ 495 : CSpeech Version 3 Core(waveform display, recording, and playback only)

$ 495 : Update from earlier CSpeech version to Version 3 Core and Supplement

Price is in U.S. dollars as of June, 1990. Price includes software onlyany hardware needs to be purchased from the appropriate vendor. Shipping charges are included-sales tax and import duties are extra, as required. To order, send check or purchase order to :

Paul Milenkovic
118 Shiloh Dr.
Madison, Wisconsin 53705−2433 U.S.A 608)833-7956 (weekdays between 7 and 9 AM Central Time).

Computer requirements

1. **Hardware requirements**

IBM PC-XT, PC−AT, PS/2 or compatible

Fixed disk drive

Math coprocessor chip(one of 8087, 80287 or 80387)

512kB memory(practical minimum)

Graphics card(one of CGA, EGA(all modes), MCGA, VGA, Hercules, or AT&T 640 by 400 monochrome)

Data acquisition(A/D and D/A) card(one of IBM ACPA(now called M-ACPA), DACA ; Scientific Solutions Lab Master, Lab Master DMA, MCDAS ; Data Translation DT2801A, DT2821, DT2823 ; Metrabyte DAS20, DAS16, DAS8/DAC02).

2. Software requirement

DOS Version 2.1 or later (DOS 3.3 recommended).

Features of CSpeech

1. Data acquisition and playback

– Data acquisition on up to 8 equal rate channels with single channel playback

– Waveform duration up to capacity of fixed disk with DMA A/D card.

2. Graphics interface

– Functions selected by single keystroke commands from a menu

– Facility for adding DOS commands and programs to the CSpeech menu

– Time interval cursors controlled either by the cursor keys or optionally from a Microsoft compatible mouse

– Multilevel zoom of time scale on cursor selected time intervals

– Text labelling of cursor positions

– Readout of time, time interval, fundamental frequency, and waveform values in user designated units using the time interval cursors.

3. Data storage and conversion

– Signal waveform storage and retrieval from binary format disk files

– Waveform editing by storing and retrieving waveforms at cursors

– Two way conversion of signal waveforms between binary and ASCII format

– Import waveform data from ILS1 TIMIT database and other formats.

4. Analysis functions

Gray scale spectrogram(requires VGA graphics display, uses IBM ACPA card, if available, for fast display)

LPC formant tracking

Single frame LPC and Fourier narrowband spectrograms Waterfall spectrogram display with linear or Bark frequency scale RMS amplitude envelope of waveforms Pitch analysis by either FPRD or short term autocorrelation Vocal jitter, shimmer and periodicity signal-to-noise ratio LPC inverse filter estimate of glottal volume velocity waveform Computation of power spectrum moments.

5. A/D D/A Converters supported by CSpeech

ACPA : IBM Audio Capture and Playback Adapter

DACA : IBM Data Acquisition and Control Adapter

DT2801A : Data Translation DT2801A

DT2821 : Data Translation DT2821

DT2823 : Data Translation DT2823

LABMASTER : Scientific Solutions Lab Master

LABMASTERDMA : Scientific Solutions Lab Master DMA

MCDAS : Scientific Solutions MCDAS 1612

DAS20 : Metrabyte DAS20

DAS16 : Metrabyte DAS16

DASH16 : Old model Metrabyte DAS16 with 1 MHz clock

DAS8 : Metrabyte DAS8/DAC-02 combination

6. Commands on the CSpeech menu

1) Screen

Bias : Change the vertical position of the selected channel

Visible : Make a hidden channel visible, make a visible channel hidden

All : Make all channels visible

None : Hide all channels

Mark : Mark the selected interval with a text label

Find : Position the waveform cursors by finding the text label among the marked intervals Label : Show the text labels for the marked intervals.

2) Files

New : Create a zero-valued waveform buffer of the specified time duration

Get : Create a waveform buffer containing waveforms from one or more files created with the Put command Put : Output one or more channels into waveform files

Leader : Specify the offset from the initial(left) cursor for waveform input or output DOS : Enter a window to run DOS commands

Exit : Exit CSpeech, saving the displayed waveforms for later redisplay

Filter : Invoke a DOS command, redirecting waveform buffer channels to standard input and output.

3) Edit

Get : Input a single waveform file into the selected channel at the initial(left) cursor position Put : Output the selected interval of the selected channel into a waveform file Read : Input an ASCII(text) file into the selected channel at the initial(left) cursor Units : Change the units name(such as Volts) or the units scale(such as 20 PP) of the selected channel

Copy : Copy the screen interval of the selected channel into one of the other channels of the waveform buffer

Zero : Zero the selected interval of the selected channel

Wlindow : Apply a raised cosine taper to the initial and final 200ms of the selected interval of the selected channel.

4) Analysis

(1) Filter

Hi pass : Apply a zero phase 20 Hz high pass filter to the screen interval of the selected channel

Integrate : Integrate the screen interval of the selected channel

Diff : Differentiate the screen interval of the selected channel

LPC : Apply an LPC inverse filter to the screen interval of the selected channel

APF : Apply an LPC all-pole(forward) filter to the screen interval of the selected channel

5) Spectrum

Fourier : Compute a Hamming window Fourier spectrum of the selected interval of the selected channel

LPC : Compute LPC coefficients and spectrum of the selected interval of the selected channel

Zero spec : Zero the spectrum display buffer

Decibels : Compute the RMS value of the selected interval of the selected channel in dB

Display : Display computed spectra

Ncoef : Change the number of coefficients for LPC analysis and the preemphasis coefficient for LPC and Fourier spectra

6) Voice

RMS : Compute the RMS amplitude envelope of the selected interval of the selected channel

Pitch : Compute the pitch contour of the selected interval of the selected channel using the FPRD algorithm

Jitter : Measure voice perturbation of the selected interval of the selected channel : pitch, jitter, shimmer, and voice aperiodicity SNR

Formants : Compute LPC formant tracks for the selected interval of the selected channel

Spectro : Display the gray scale spectrogram of the selected interval of the selected channel and optionally plot formant tracks

7) Waterfall

Fourier : Compute and display Hamming window Fourier spectra of the selected channel starting at the initial cursor and ending at the final cursor

LPC : Compute and display LPC spectra of the selected channel starting at the initial cursor and ending at the final cursor

1. BarkF : Compute and display Fourier spectra on a Bark transform frequency scale 2. BarkL : Compute and display LPC spectra on a Bark transform frequency scale.

Record : Digitize analog waveforms with the A/D converter card and store the samples in the waveform buffer

Play : Play back the selected interval of the selected channel through the D/A converter card

Quit : Exit CSpeech without saving the waveform buffer and waveform display configuration

hi/lo : Cursor Change between high and low increments of cursor movement.

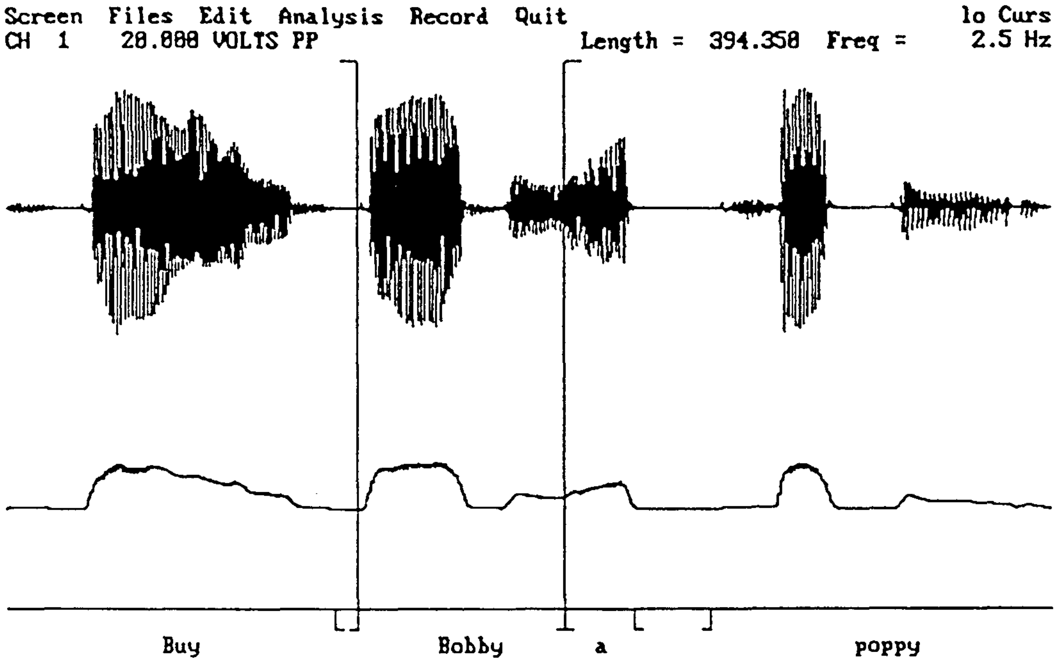

Fig. 1. CSpeech waveform display showing menu line, status line, waveform cursors, two waveform plots, and text labels.

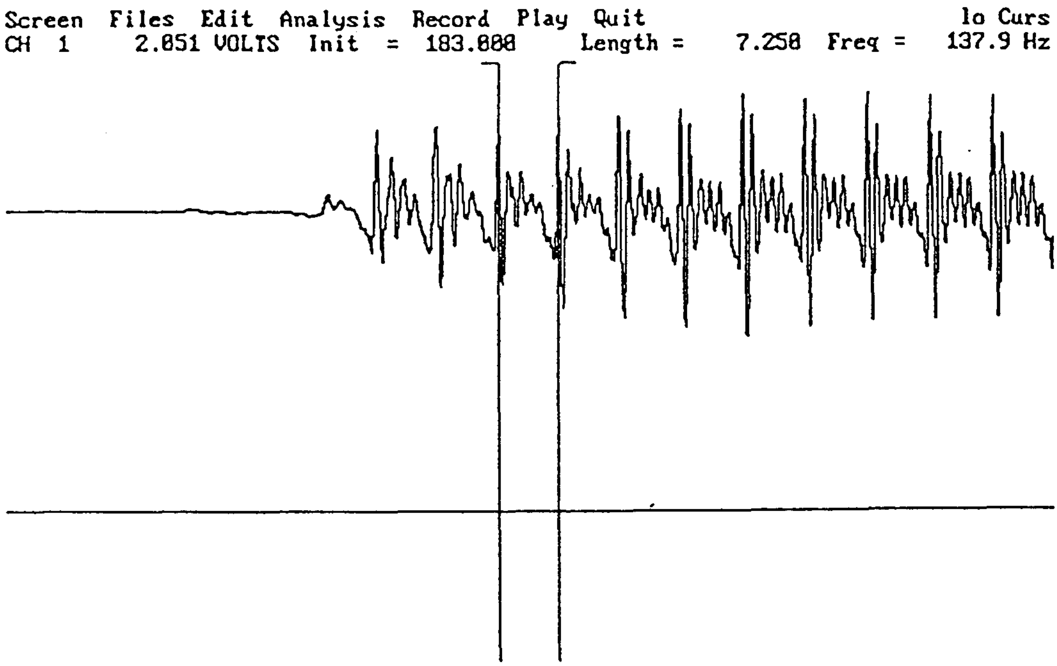

Fig. 2. Manual measurement of the pitch period and fundamental frequency using the cursors.

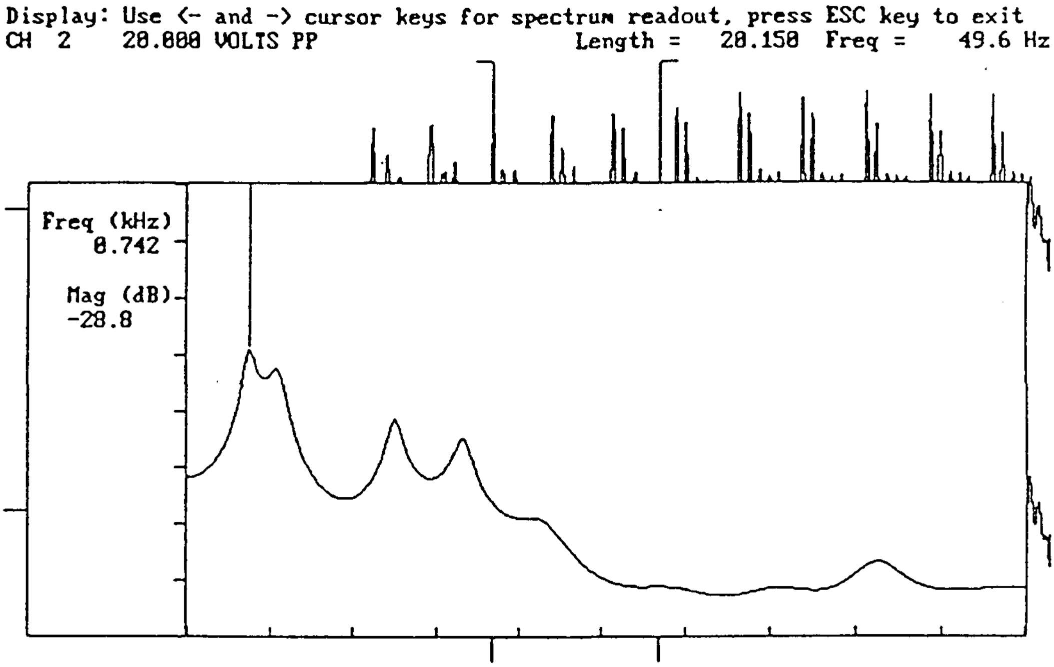

Fig. 3. Plot of the LPC spectrum of the cursor selected waveform interval.

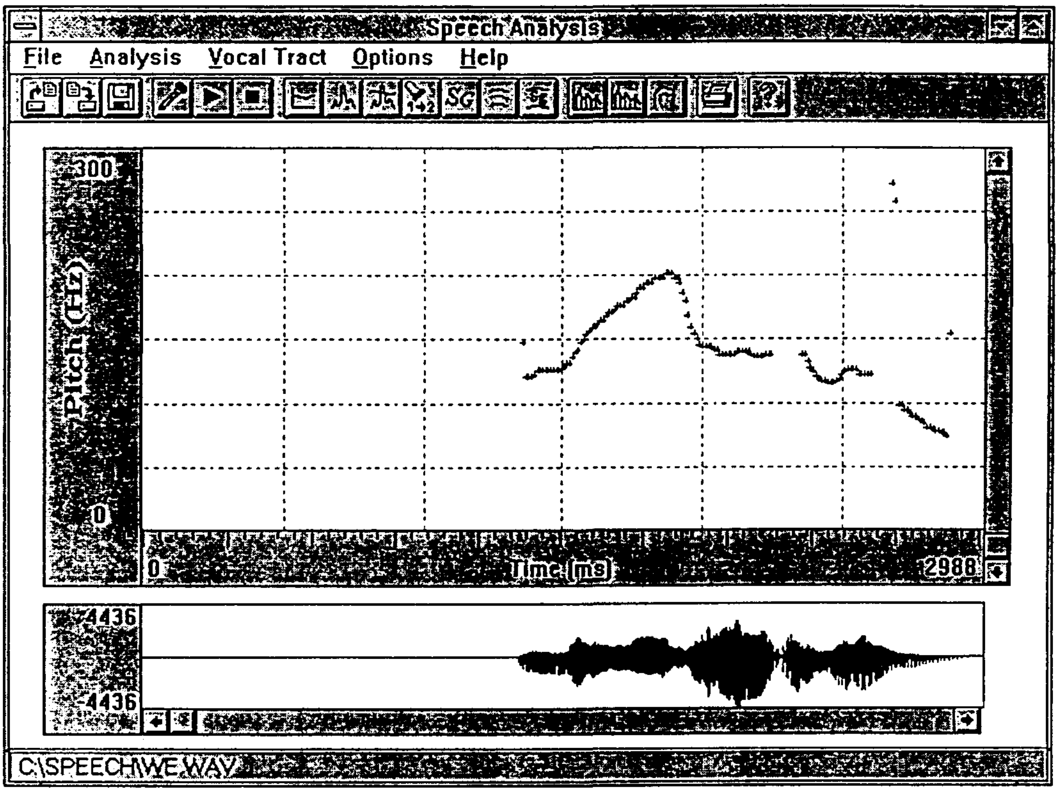

Fig. 4. A display of pitch change versus time.

Dr. Speech Science for Windows(Version 1.01, 2.0)

연세대학교 의과대학 이비인후과, 음성언어의학연구소
최 홍 식

Introduction

Dr. Speech Science for Windows is a PC-based speech/voice assessment and training system designed specifically for Speech-Language Pathologists, Otolaryngologists, speech scientists, and voice scientists. Dr. Speech Science consists of seven programs (ver 1.01). It is possible to use them together or independently. Application areas for this easy to use program include voice disorders, motor speech disorders, speech training, acoustic phonetics, voice identification, accent modification, and English as a second language.

The most important feature of Dr. Speech Science is that it is a fully window-based program without any extra Digital Signal Processing(DSP) hardware required. This system can be used in a PC-based desktop or laptop computer equipped with a 16-bit sound card. The following sections list the programs that come with Dr. Speech Science for Windows, explain their function, and discuss how they complement each other.

1. Speech analysis

Speech Analysis is a sophisticated clinical tool that provides highly versatile information for speech evaluation and therapy features including :

1) Recording, playback, and editing of audio waveforms

2) On-screen cursors for quick, precise measurements

3) Pitch extraction and sound intensity displays

4) Spectral analysis, Formant analysis, and Long Term Average Spectrum(LTAS)

5) Spectrogram display, and Formant tracking programs

6) Dynamic Vocal Tract Display

2. Speech training

Speech Training is a powerful educational and clinical tool with easy to use dual-screen graphic displays to compare an utterance provided by a teacher(clinician) with another provided by a student (client). This is an ideal tool for use in phonetic modeling, speech therapy, or accent modification of a non-native language speaker. This program includes :

1) Dual-screen recording, playback, and editing of audio waveforms

2) Dual-screen pitch and/or intensity displays

3) Dual-screen spectrograms and/or formant tracking

4) Dynamic Vowel Tracking(F1-F2 plot)

5) IPA transcription tutorial

6) Vowel identification

3. Voice assessment

Voice Assessment is a comprehensive clinical tool that provides information for use in quantitative assessment of voice quality. The Voice Assessment program offers a number of exciting features including :

1) Recording, playback, and editing of audio waveforms

2) Spectral analysis, Formant analysis, and Long Term Average Spectrum(LTAS)

3) Perturbation Measurement, such as jitter, shimmer, NNE, and so forth

4) Voice Profile

5) Quantitative Assessment of Voice Quality, based on a built-in database of 2,937 normal and 902 pathological voices

4. Voice synthesis and therapy

Voice Synthesis and Therapy is an integrated clinical and research tool for creating idealized acoustic targets. Speech/voice scientists can use this powerful synthesis program to produce stimuli for use in speech perception experiments.

With this programs, voice clinicians can create target vowels for use in voice therapy(For example, by synthetically generating vowels with acoustic features that lie between the patient's current profile and the normal profile, the clinician can provide intermediate targets that gradually bring the patient in the direction of the normal profile.). This successive approximations approach to treatment of voice disorders consists of the following steps :

1) Voice evaluation

2) Formation of reasonable clinical hypothesis

3) Identification of special parameters, such as F0, F0 contour, jitter, shimmer, glottal noise energy, and so forth

4) Creating a synthetic vowel target

5) Ear training and model matching

6) Repetition

5. EGG assessment

For voice clinicians who use electroglottographic assessment of vocal function, this program offers an easy way to obtain quantitative information from the EGG signal. The electroglottograph(EGG) or laryngograph is an instrument for non-invasive investigation of vocal fold movement. The EGG Assessment program includes the following features :

1) Recording, playback, editing of the EGG signal

2) Spectral analysis, Formant analysis, and Long Term Average Spectrum(LTAS) of EGG

3) Perturbation Measurements, such as EGGjitter, EGG-shimmer, EGG-Noise, Contact Quotient(CQ), CQ Perturbation, Contact Index(CI), CI Perturbation

4) EGG Profile

5) Quantitative Assessment of Vocal Function, based on a built-in database of EGG profiles from normal and pathological voices

6. Clinical progress tracking

The Clinical Progress Tracking program is an interactive tool you can use to monitor the process of voice therapy for your client. This program can provide a graphic display of voice or electroglottography parameters across different therapy sessions. The Clinical Progress Tracking program offers a number of exciting features, including :

1) Monitoring the effect of voice therapy

2) Automatic data collection and reporting

3) Graphic display with 2-D and 3-D charts

4) Print option

7. Wave generator

Wave Generator is an educational and research tool that provides the user with the capability to generate, modify, and display a wide variety of audio waveforms, such as simulations of the glottal sound source or vowel-like harmonic waveforms. Audio waves are digitized, synthesized for output, viewed, analyzed, used to hold the results of analysis, and exported to other application programs. This program can be used to generate the audio waveforrns, including :

1) Harmonic waveform

2) Sawtooth waveform

3) Sine waveform

4) Square waveform

5) Triangular waveform

8. Real speech

Real Speech provides real time displays of continuously varying speech waveforms, or fundamental frequency, intensity, spectrum, or EGG waveform displays. The capabilities of this software provide dynamic interactive displays(instantaneous visual feedback of the acoustical performances of talkers or singers) during sound production.

Speech-Language Pathologists, Otolaryngologists, voice pathologists, teachers of singing, singers and teachers of languages find these displays to be useful in clinical assessrnent, in the teaching of dynamic patterns of intonation, the patterning of linguistic stress contrasts and temporal patterning of speech (VOT, differential vowel duration, noise burst durations, etc.).

A special feature of Real Speech model matching is the ability to store the acoustical features of a dynamic speech sample on the computer screen in one color(for example, the pitch pattern of the instructor), and then compare the performance of an attempt to match the pattern(the student's attempt) by tracing the second pattern in another color.

9. Phonetogram

Phonetogram, displays the dynamic range of the human voice in terms of both fundamental frequency (pitch) and intensity(loudness). Speech-Language Pathologists, Otolaryngologists, voice teachers, and singers find this display to be useful in identifying the limits of vocal function.

Applicatiions

1. Speech

Speech Evaluation
Speech Training
Speech Pathology
Acoustic Phonetics

2. Language

Language Training
Accent Modification
English as a Second Language(ESL)

3. Voice

Voice Assessment
Electroglottography Assessment
Quantitative Assessment of Voice Quality
Quantitative Assessment of Vocal Function
Monitoring the Effects of Voice Therapy
Voice Therapy
Vowel Perception and Synthesis

System Requirements

Before you can install Dr. Speech Science for Windows, make sure your computer has met the minimum hardware and software requirements listed below.

1. Operating system

This software package runs with Microsoft Windows 3.1 or later(or IBM OS/2), which requires DOS 3.1, or a later version of DOS.

2. Computer

This software package is designed to run on an IBM PC or fully-compatible computers using the Intel 386, 486, Pentium, or higher processor. For the full set of version 2.0 including Real speech, PC 486 DX 66 or better and 2 MB RAM is recommended.

3. Monitor

A color monitor(640 * 480 resolution and 16 colors or higher) is required. A 256 color monitor is highly recommended.

4. Memory

The memory requirement for running Dr. Speech Science for Windows(ver. 2.0) is 2 MB RAM. 4 MB RAM is highly recommended.

5. HardDisk

Your computer must have a hard disk with at least 15 MB of free disk space(version 2.0).

6. Mouse

It is highly recommended that you use a mouse when running any of the programs in the Dr. Speech Science for Windows.

7. Sound card

Your computer must have a Windows-compatible sound card that has a 16-bit resolution and a sampling frequency of 11025, 22050, or 44100 Hz.

If you install your program in a desktop computer, it is recommended that you use one of the following sound cards :

*Creative Lab., Inc. or its distributor

Sound Blaster 16(Value Edition), Sound Blaster 16 (Basic Edition), Sound Blaster 16 CD, or Sound Blaster AWE 32

*TRM

IBM Audio Capture and Playback Adapter

Ask : Item : Audio Adapter for non-Micro Channel computers with software and manual

IBM part number : 92F3378 If you install your program in a laptop computer, it is recommended that you use a portable sound card from the following source :

*Digispeech Inc. or its distributor

Digispeech Portable sound plus Adapter

8. Microphone

You must have a good microphone to run Dr. Speech Science for Windows software. A quality microphone is very important to receive the best assessment of your client's voice and speech.

9. Speaker

Your computer should have a speaker to attach to the sound card.

10. Printer

This is optional equipment. You can use any printer that is supported by Microsoft Windows 3.1 or later to print out the information obtained through use of Dr. Speech Science.

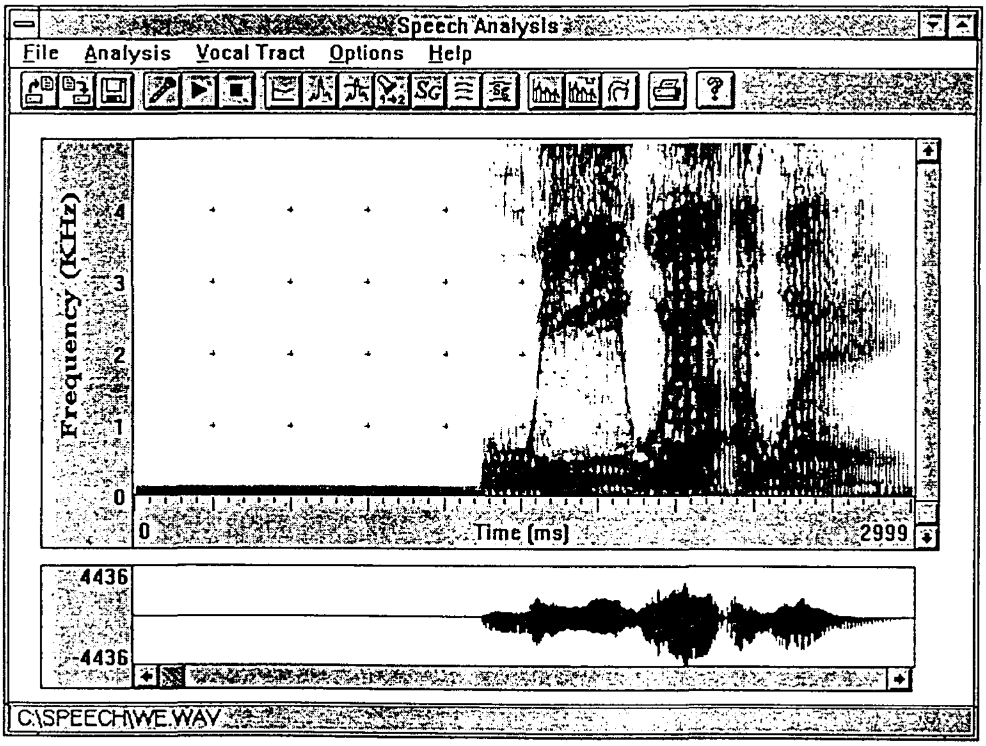

Fig. 5. Spectrogram display.

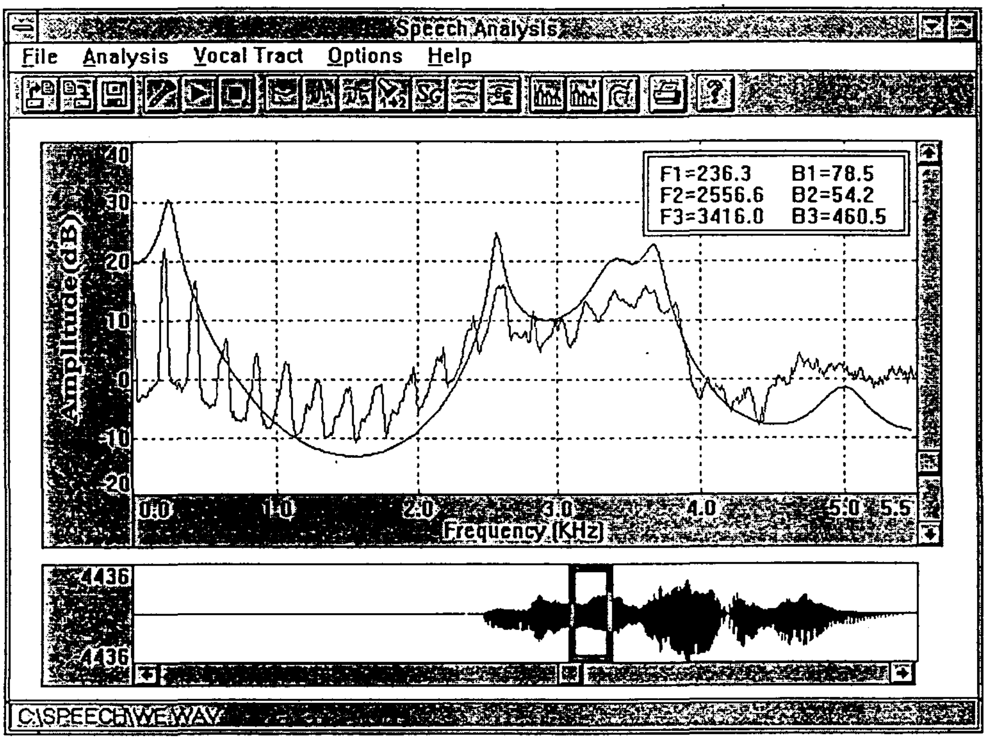

Fig. 6. Power and LPC Spectrum(LTAS) within green frame.

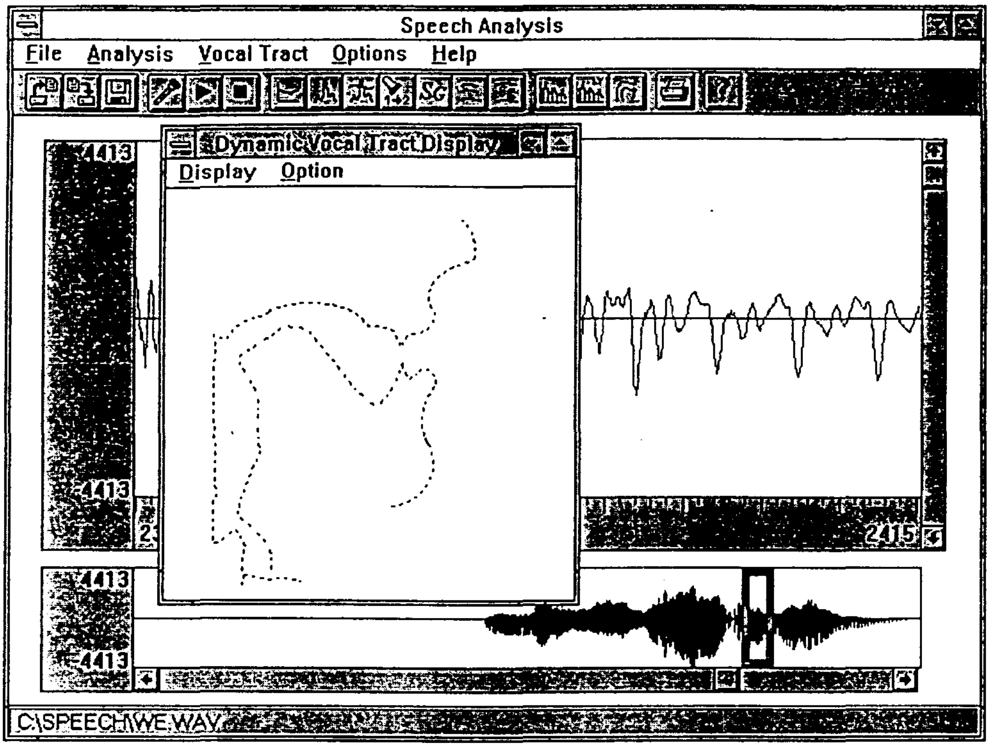

Fig. 7. Dynamic vocal tract display.

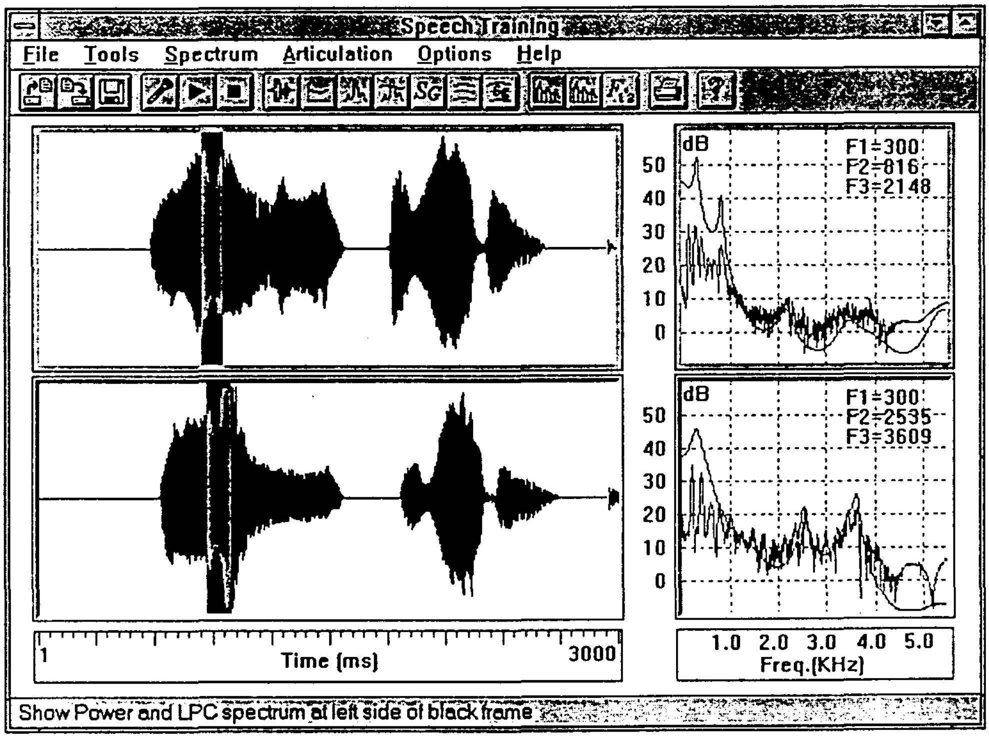

Fig. 8. Comparing phonetic similarity.

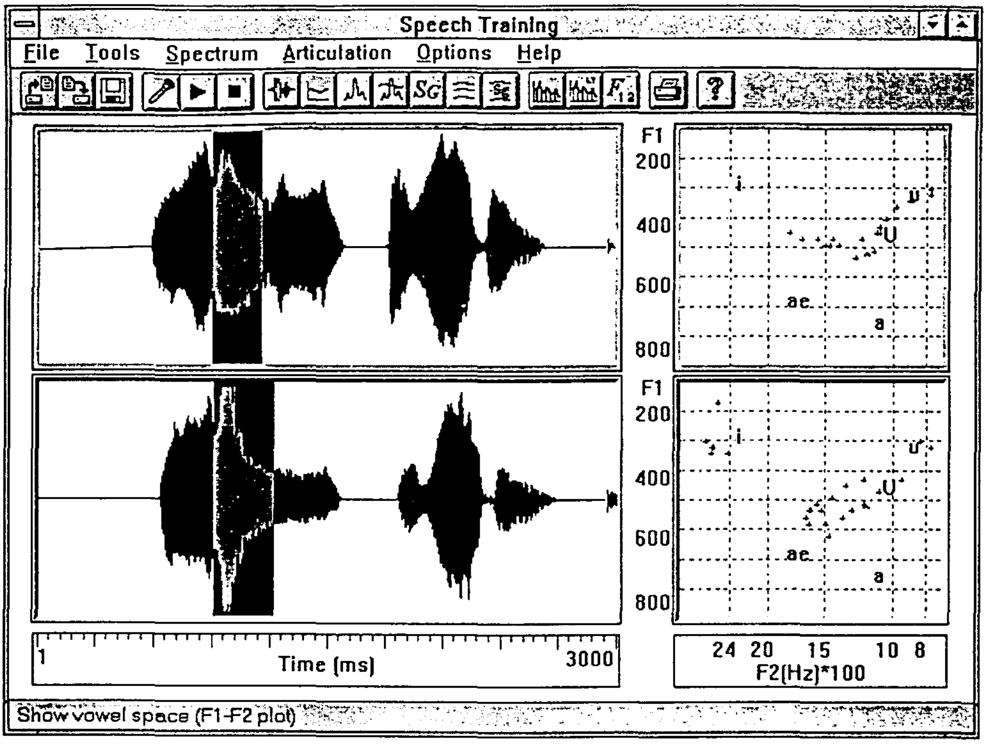

Fig. 9. *Comparing articulation.*

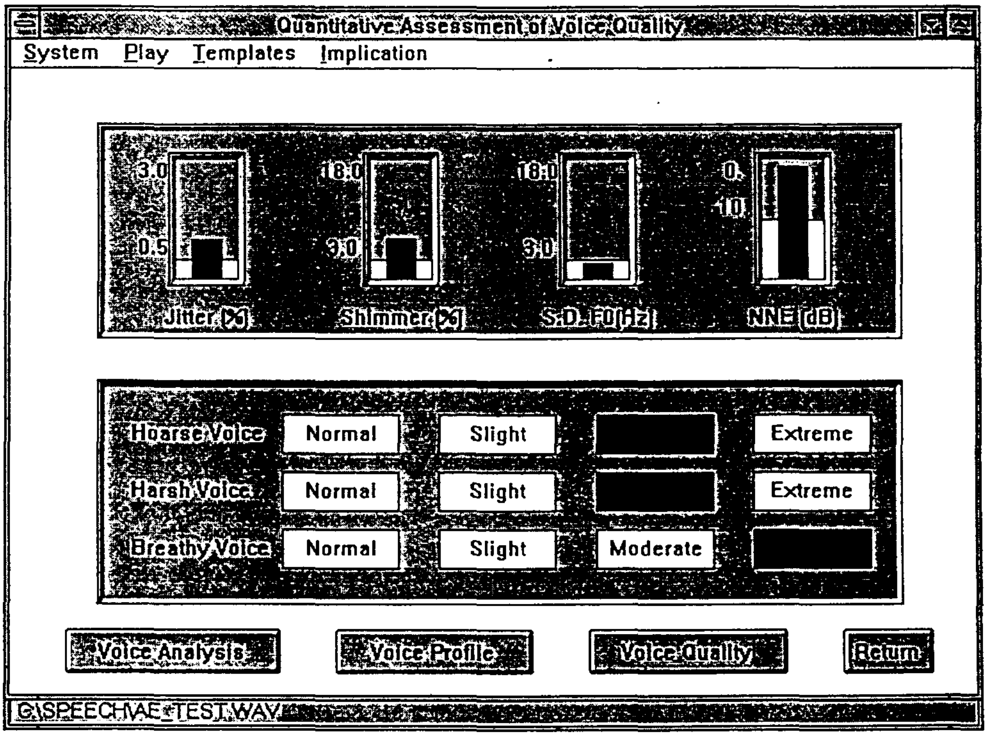

Fig. 10. *Voice quality display.*