CSpeechSP Demo/Microbeam Viewer

Paul H. Milenkovic [1]
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
1415 Johnson Drive
Madison, Wisconsin 53706

Charles Read
Department of Linguistics
University of Wisconsin-Madison 53706

[1]Please refer inquiries to Paul Milenkovic, 118 Shiloh Dr., Madison, Wisconsin 53705-2433, U.S.A., 608/833-7956

Abstract

CSpeech, pronounced "See-Speech," is a 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. Distinguishing characteristics include acquisition and display of signals in multiple channels, minimal restrictions on sample rate and waveform duration, support for a variety of hardware configurations, fast graphics display, and a user-extensible menu-based command structure. While the analysis functions of CSpeech are tailored to the requirements of speech, CSpeech is useful in other applications such as digital filter design and physiological data recording.

CSpeechSP is the most recent version, named in honor of S. P. Korolev, the noted Soviet aerospace engineer.

1   Introduction

Think of CSpeech as a digital storage oscilloscope with an enormous capacity for waveform samples together with analysis and measurement functions tailored to speech. CSpeech runs on PC compatible computers under the DOS operating system, providing these capabilities on the most widely available type of computer.

CSpeechSP Demo employs a waveform buffer containing from 2 channels. Each channel is sampled at the same rate and contains the same number of samples. CSpeech can display a little or a lot of the time extent of the waveforms, it can change the amplitude scaling of the waveform traces, and it can move the traces up or down on the screen. Cursors placed across these traces take measurements, and key presses activate analysis functions. Because the samples in all the channels line up, it is easy to filter all or part of any channel and place the filtered waveform into any other channel.

While the comparison to an oscilloscope is helpful, you must consider that CSpeech is a set of computer programs that run under the DOS operating system of an IBM PC-compatible computer. With DOS, you are always "in" a particular drive and directory. The drive is indicated in the DOS prompt (such as `C:>`). The directory may also be indicated in the DOS prompt; otherwise, it can be displayed with the command `cd`. We strongly recommend that CSpeech users learn how to select a drive and directory: please review the DOS commands

> `chdir` Change directory (or `cd`)
>
> `mkdir` Make directory (or `md`)
>
> `rmdir` Remove directory (or `rd`)
>
> `dir` List Directory.

You need to know some key terms. CSpeech maintains a *waveform buffer* for recording, displaying, and performing operations on signal waveforms. The waveform buffer may contain one or more *channels* of signal waveforms sampled at equal rate. CSpeech maintains *work files* in the *current directory*, the disk drive and directory selected under DOS when CSpeech is started. Information about the layout of the waveform display is saved in the work file `cspeech.sav` while the waveform sample values for all channels are saved in the work file `cspeech.adc`.

Starting CSpeechSP

The CSpeechSP demo program may be started by entering the command

```
sp
```

at the DOS prompt. You enter a DOS command, such as the one that starts CSpeech, at the DOS prompt, typically `C:>`, by typing in the command and pressing the `Enter` key. CSpeech may, however, be started a different way: with a different command name, from a menu program, or by clicking on an icon in Microsoft Windows. Please check with the person who installed CSpeech on your computer to find out how to start CSpeech.

Exiting CSpeechSP

You may exit with either the `Quit` command, invoked by pressing `q`, or with the `Exit` command found in the `Files` submenu. Select the `Files` submenu by first pressing `f` and then invoke the `Exit` command by pressing `e`. Both commands exit the program. Using `Files Exit` will prompt you to save analysis results. The demo/microbeam viewer version of CSpeech is meant of analysis of existing waveforms and does not permit you to save any waveforms.

Cursor and Menu Commands

Upon starting CSpeech, a menu line, status line, and cursor line appears at the top of a waveform display. Cursor commands control the selection of channel and time extent of the waveform. Commands selected from the menu operate on these waveforms. Both the cursor and menu commands can be selected with the mouse or alternately with key presses.

The menu lists submenus and commands: view the contents of a submenu by pressing the first upper-case letter of the submenu name; invoke a command by pressing the first upper-case letter of the command name. It does not matter whether you press a lower or upper case letter on the computer keyboard. To invoke a command within a submenu, view each submenu leading to that command and then invoke the command. Press `Esc` or the spacebar key at any submenu to return to the main menu.

Menu and option selections may also be made with the mouse. Position the mouse cursor and press the left mouse button to select. Release the left mouse button to confirm the selection; move the mouse cursor before releasing to cancel the selection. Press the right mouse button to exit any submenu or display of options.

The status line and cursor line reflect the state of the waveform display as a result of invoking cursor functions and menu commands. This is important because many commands operate on the selected channel and selected interval indicated by this display.

2   Cursor Keypad Functions in CSpeech

The cursor keypad, and optionally the mouse, can alter the settings of the waveform display: channel selection and scale of individual channels, cursor placement, and screen interval. Position the mouse cursor over one of the arrows located under the status line entries for channel, scale, and interval, and press the left mouse button to alter that setting. Position the mouse cursor over the waveform display; press the left mouse button to place the initial cursor, right mouse button to place the final cursor. If you prefer to employ the cursor keys, please refer to Table 1 where `Ctrl`-*key* means hold down the `Ctrl` key and press *key* once. The functions CSpeech assigns to these keys are shown in Table 2.

Channel selection and scale

      `Home` Select lower-numbered channel (*e. g.* change from 2 to 1)

      `End` Select higher-numbered channel (*e. g.* change from 1 to 2)

      ↑ Scale up display of selected channel

      ↓ Scale down display of selected channel.

Cursor placement

      `Ctrl` ↓ Select initial waveform cursor for movement

      `Ctrl` → Select final waveform cursor for movement

      ↓ Move the selected waveform cursor to the left

      → Move the selected waveform cursor to the right

      `LeftMouse` Position and drag the initial cursor

      `RightMouse` Position and drag the final cursor.

Holding down `Ctrl` and pressing ↓ once (`Ctrl` ↓) selects the initial waveform cursor. Pressing → moves this cursor to the right starting from the left edge of the screen. Each press of → advances the initial cursor an additional increment. Holding down `Ctrl` and pressing → once (`Ctrl` →) selects the final waveform cursor. Pressing → moves this cursor to the right starting from the position of the initial waveform cursor. Each press of → now advances the final cursor an additional increment.

Screen interval

      `PgDn` Zoom in to the interval marked by waveform cursors

      `PgUp` Zoom out to a wider screen interval

      `Ctrl-PgUp` Scroll right by one screen interval

Ctrl-Home Scroll left by one screen interval

Ctrl-PgDn Scroll right by one quarter screen interval

Ctrl-End Scroll left by one quarter screen interval

Ins Restore waveform cursors to initial positions

Press Del to remove any message box from the bottom of the screen.

Home Ctrl-Home	↑	PgUp Ctrl-PgUp
↓ Ctrl ↓		→ Ctrl →
End Ctrl-End	↓	PgDn Ctrl-PgDn
Ins		Del

Table 1: Keys of the keyboard cursor pad.

LowerChan ScrollLeftFull	ScaleUp	ZoomOut ScrollRightFull
MoveCurLeft SelectInitCur		MoveCurRight SelectFinalCur
UpperChan ScrollLeftPart	ScaleDn	ZoomIn ScrollRightPart
RestoreCurs		DelText

Table 2: CSpeech function assignments to keys of the keyboard cursor pad.

2.1   Main menu

`Screen Files Analysis Play Mixer Quit  hi Curs`

`Screen` (submenu)

`Files` (submenu)

`Edit` (submenu)

`Analysis` (submenu)

`Play` Play the cursor-selected interval of the selected channel through the D/A converter. The Demo version requires a sound card fully compatible with the Creative Labs Sound Blaster 16. Be warned that many 16 bit sound cards that are Sound Blaster compatible are not Sound Blaster 16 compatible, including the new PCI-bus sound card from Creative Labs. The Creative Labs AWE64 Value and Gold cards are Sound Blaster 16 compatible.

`Mixer` Sets playback sources and levels. You may have to invoke the Mixer from the Windows program supplied with your sound card.

`Quit` Exit CSpeech without saving the waveform buffer and waveform display configuration.

`hi/lo Curs` Press `c` to change between high and low increments of cursor movement. Placing cursors with the mouse selects `lo Curs`.

The playback command `Play` does not appear if you select a pitch trace, a non-playable channel.

2.2   Screen menu

`Screen: Mark Find Label Del`

> `Mark` Mark the cursor-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.
>
> `Del` Remove any text display at the bottom of the screen (same as pressing `Del` key).

2.3   Files menu

`Files: Open Exit Text.lbl`

Open Open a waveform buffer by retrieving a waveform from a file. The buffer duration is adjusted to the length of the data, and channels not receiving waveforms are zeroed. Accepts data in the CMU, CSpeech, CSRE4.0, ILS, Kay .NSP, NCVS92, RIFF, Speech Station, SPHERE, or UW Microbeam DF file formats. The sampling frequency is determined by the data file. Options include

Single Input a single waveform from the file *fname.ext* into Ch 1. If a pitch trace is stored in ASCII (alphanumeric text) format in *fname.*`f0`, input that trace into Ch 2.

Lbl Open the specified label file. If the label file does not exit, it will be created. If you specify *lname*, all changes to text labels will be saved to *lname.*`lbl`, and pitch, formants, and other analyses will be saved to *lname.*`f0`, *lname.*`fbw`, and so on. This command is especially useful for analyzing waveforms input from CD-ROM (such as TIMIT).

Exit Exit CSpeech, optionally saving analysis results such as the pitch trace.

Text.lbl Invoke the `edit` command to view or modify text labels made by `Screen Mark` (requires DOS 5.0 or later).

2.4 Analysis menu

```
Analysis: Filter Spec Display Time/freq Ncoef Measure Voice Beam
```

`Filter` Options include

`Hipass` 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.

`Phase` Apply a phase-correcting high-pass filter to the screen interval of the selected channel. This filter corrects for the phase of recordings made with a 20 Hz high-pass cutoff microphone and the 12 Hz high-pass cutoff of the M-ACPA card.

`FIR` Apply a Hamming window lowpass FIR filter to the cursor-selected interval of the selected channel. The cutoff frequency is designated in kHz, and the maximum number of coefficients is 255.

`fiR_hp` Apply a Hamming window highpass FIR filter to the cursor-selected interval of the selected channel. The cutoff frequency is designated in kHz, and the maximum number of coefficients is 255.

(the Analysis menu listing continues on the next page)

`Spectrum` Options include

  `Fourier` Compute a Hamming window Fourier spectrum of the
     cursor-selected interval of the selected channel.

  `LPC` Compute LPC coefficients and spectrum of the cursor-selected
     interval of the selected channel.

  `Zero spec` Zero the spectrum display buffer.

  `Time_avg` Compute the long-term average (LTA) Fourier spectrum
     of the cursor-selected interval of the selected channel.

  `Save` Output the last two computed spectra and their difference to
     the file `spec.dat` in alphanumeric text (ASCII) format. These
     data can be viewed with a plot program.

`Display` Display computed spectra.

`Time/freq` Compute and display a gray scale time/frequency spectro-
   gram. Press `A` or `B` to select top or bottom of a two-channel spec-
   trogram. Press `P` to select better screen colors for printer output
   (using Pizazz+ software). Values for analysis bandwidth, frequency
   range, minimum displayed intensity, and screen interval may also be
   modified. Invoking this command activates the `Time/freq` menu of
   the time/frequency spectrum display.

`Ncoef` Change the number of coefficients for LPC analysis and the pre-
   emphasis coefficient for LPC, Fourier spectra, and time/frequency
   spectrograms.

`Measure` Options include

  `Decibels` Compute the decibel (dB) level of the cursor-selected in-
     terval of the selected channel.

  `Compute_stat` Compute the mean, min, max, standard deviation
     (and percent voiced for pitch) of the cursor-selected interval of
     the selected channel – results are saved in the temporary file
     `stat`.

  `Edit_stat` View, edit, or copy the temporary file `stat` (requires DOS
     5.0).

  `Save_stat` Copy the temporary file `stat` to `stat.sav`.

`Voice` Options include

> `Newjit` Compute jitter, shimmer, and aperiodicity SNR for the cursor-selected interval of the selected channel. SNR is computed using a two-tap pitch predictor (hence Newjit).
>
> `RMS` Compute the RMS amplitude envelope of the cursor-selected interval of the selected channel.
>
> `DB` Same as RMS, but gives amplitude envelope in decibels. Zero dB is the same as `Analysis Measure Decibels`.
>
> `Pitch` Compute the pitch contour of the cursor-selected interval of the selected channel using the short-term autocorrelation function with center clipping.
>
> `Edit_f0` Edit the pitch contour computed from the acoustic waveform in the selected channel. With the cursors marking a pitch period, convert period to frequency and place the value in the destination channel – permits manual pitch measurement on a cycle by cycle basis. Invoke `Analysis Voice Save.f0` to save these changes.
>
> `Import.f0` Retrieve the pitch contour that goes with the acoustic waveform in the selected channel. CSpeech will prompt for the channel to put the pitch. The pitch contour is input from an alphanumeric text (ASCII) file into the cursor-selected interval of the destination channel. This command retrieves data stored with the `Analysis Voice Save.f0` command.
>
> `Save.f0` Save a pitch contour. Before invoking this command, you may select either the acoustic waveform used to compute the pitch contour or the pitch contour itself; CSpeech will indicate the channel where it found the pitch. This command saves the pitch contour in the entire waveform buffer into an alphanumeric text (ASCII) file.

`Beam Xy` Displays x-y positions of pellet markers tracked with the x-ray microbeam system. If the waveform is `tp001.acc`, displays pellet data from the file `tp001.xyd`. Drag the left mouse button over the waveform display to position the x-y display in time. Commands include

> `Play` Plays the cursor interval of the waveform and animates the x-y vocal tract display. To change the cursor interval, `Exit`, reposition the cursors in CSpeech, and reinvoke `Analysis Beam Xy` to get back this display.
>
> `Speed` Select normal or /2 playback speed.
>
> `Initial` Locate the initial marker at the current position of the x-y display.
>
> `Final` Locate the final marker at the current position of the x-y display.
>
> `Mark` Write a label marker to the label file (such as `tp001.lbl`) using the times of the initial and final markers (see above) and the text label you enter with this command.
>
> `Trace` Display colored glow worms of the paths of individual pellets over time.
>
> `Clear` Remove the pellet path glow worms.
>
> `Write` Save pellet and optionally formant values for the current time position in the x-y display to a file. The file is automatically generated using the waveform name and time code. For waveform `tp009.acc` and time 1950 ms, the file name is `tp00901.950`. The format is
>
> | | | | | | | | | |
> |---|---|---|---|---|---|---|---|---|
> | $T1_x$ | $T1_y$ | UL_x | UL_y | $MANI_x$ | $MANI_y$ | F_1 | F_2 | F_3 |
> | $T2_x$ | $T2_y$ | LL_x | LL_y | $MANM_x$ | $MANM_y$ | | | |
> | $T3_x$ | $T3_y$ | | | | | | | |
> | $T4_x$ | $T4_y$ | | | | | | | |
>
> where positions are in cm and formant frequencies in kHz.
>
> `Exit` Return to the CSpeech main menu.

2.5   Time/frequency menu

This menu is selected by invoking `Analysis Time/freq` to display a time/frequency spectrogram. Invoking `Main menu` restores the main menu and waveform display.

`Time/freq: Screen Files Analysis Play Main menu  hi Curs`

`Screen` Options include

  `Color` Toggle between a gray scale rendering and a color rendering of the time/frequency spectrogram (VGA or super VGA only).

  `Bias` Change the vertical position of the selected channel.

  `Mark` Mark the cursor-selected interval with a text label.

  `Label` Show the text labels for the marked intervals.

`Files` Options include

  `Exit` Exit CSpeech, optionally saving the waveform buffer for later redisplay

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

`Analysis` Options include

  `Spectrum` Options are the same selections as `Analysis Spectrum` activated from the waveform display.

  `Display` Display computed spectra.

  `Time/freq` Recompute and display the time/frequency spectrogram.

  `Ncoef` Change the number of coefficients for LPC analysis and the preemphasis coefficient for LPC, Fourier spectra, and time/frequency spectrograms

  `Formants` Compute, display, and edit formant tracks,

  `Beam` Display and analysis of x-ray vocal tract measurements.

`Play` Play the cursor-selected interval of the selected channel through the Sound Blaster 16 compatible sound card.

`Main menu` Exit the time/frequency display and return to the main menu of the waveform display.

`hi/lo Curs` Change between high and low increments of cursor movement.

2.6   Find prompt

The command `Files Open` employs the Find prompt (Figure 1) to locate and select files. The Find prompt may be used to 1) change to a new drive or directory, 2) search the contents of a directory, 3) specify a file name, or 4) take no action.

The Find prompt indicates the path: the drive and directory you are "in." Change the path by typing a valid DOS directory or drive name and pressing `Enter`. The `Path` indication will reflect the change. Home base is the CSpeech current directory: the drive and directory selection in effect when CSpeech was invoked (also contains the CSpeech work files). You may always return to the current directory by changing to directory "." (type a period and press `Enter`).

Search the indicated path for files and directories by typing a DOS wildcard specification and pressing `Enter`. You can find information on DOS wildcards in the DOS manual. An example wildcard is `*.wav`, meaning search for any file name with the extension `.wav`. The wildcard `*.*` means search all files. The last wildcard entered can also be selected by pressing ↓. To search only for directories and drives, press the `PgDn` key. You can also select these search options with the mouse.

The file and directory names turned up in the search will be listed below the Find prompt. Use the cursor arrow keys to scroll through the list and select one of the names. To accept a selection, press the `Enter` key. Optionally, you can double click the selection with the mouse, or you can use the mouse to select `Confirm`. Selecting a directory name in this manner will cause CSpeech to change to that directory. Selecting a file name in this manner will select that file name for `Open`, `Save`, `Get`, `Put`, or `Read`.

You may also specify a file by typing the file name and pressing `Enter` at the Find prompt. Exit the Find prompt without specifying a file by pressing the `Esc` key or by using the mouse to select `Exit`.

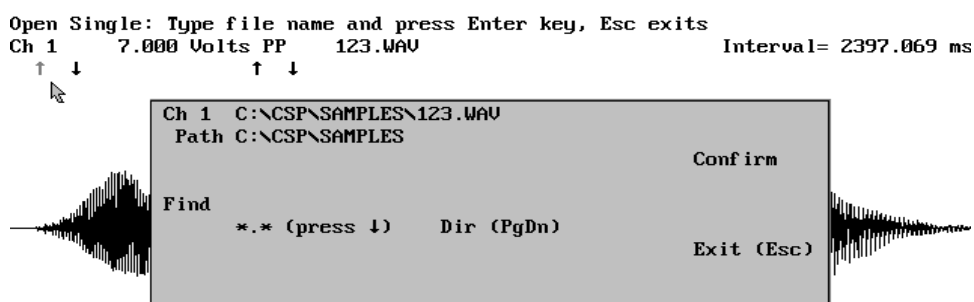Figure 1: Find prompt