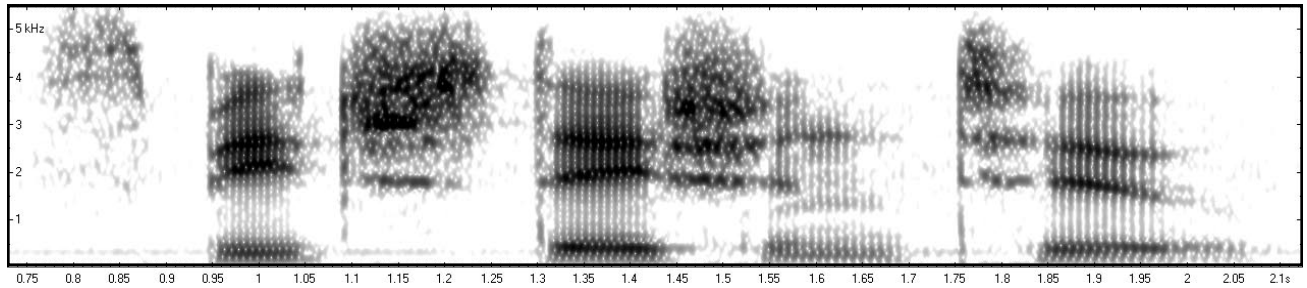


SpeechStation2



Our warranty...

Thank you for purchasing *SpeechStation2*

Please send us the owner identification card that we have enclosed. This card authorizes you to obtain technical support and allows us to send you software corrections and notices of upgrades.

Carefully read the license agreement for the *SpeechStation2* software. It is important that you understand your rights in the use of the software, as well as the restrictions on its distribution or sale to others.

If the *SpeechStation2* software that you have purchased does not work as described in this manual, we will either replace it with software that does, or take the software back and refund its purchase price to you.

If you believe that you are entitled to a refund or to return the software, you need to write to us, describing the problem exactly. If we are at fault, we will send you the authorization needed for return of the software.

You must respect copyright law. By installing the software, you agree that you will not copy it, that you will not give or sell a copy of it to anyone, and that you will not incorporate any of it into other work. That restriction includes all features of *SpeechStation2*, especially those covered by copyrights or patents.

Entire contents Copyright ©1997-2004 Sensimetrics Corporation. All Rights Reserved.

“Windows”, “Windows 95” and “Windows 98” are trade marks of Microsoft. Whenever capitalized, “Windows” refers to Microsoft’s operating system. Other trade marks cited are those of their respective owners.

Getting help from Sensimetrics

If you see anything that could be improved in *SpeechStation2*, we hope that you will tell us about it. We will also be happy to provide help, if we can, in using it in your research or other applications. The easiest way to do this, and one that gives us time to think about your question, is by e-mail. Send your question or comment to

help@sens.com

and we’ll see that it goes to the person who will be best able to help you.

Should your need be something of an emergency, our telephone number is

1-800-111-0101

Our fax number is

781-399-0858.

If you have something you want to say in a more leisurely manner, we look forward to receiving your letter at

**Sensimetrics Corporation
14 Summer Street, Suite 403
Malden, MA 02148**

One year of free technical support is provided (from the date of purchase) with *SpeechStation2*. Additional technical support may be purchased from Sensimetrics. When emailing Sensimetrics regarding a technical support inquiry, please be sure to include your product’s Serial Number.

Contents

Introducing <i>SpeechStation2</i>	4
Requirements for Use.....	5
Installation	6
The Main Screen	8
Overview.....	8
Main Screen Commands.....	10
Loading and Navigating a File.....	16
Playback and Recording	19
Spectrogram Analysis Options.....	23
Spectrogram Display Options	26
Pitch Tracking	28
Formant Tracking.....	30
Labels	31
Printing	32
Stereo Files.....	33
Spectrum Viewer	34
Spectrum Viewer Menus	35
Waterfall Plot	38
Real-Time Spectrogram	42
Vowel Space Plot	44
Filter	46
Appendix A: Recorded Phonetic Library	48
Appendix B: Making Better Speech Recordings	53
Index ...	56

Introducing *SpeechStation2*

SpeechStation2 is a software system for speech analysis. It has been especially designed for research in production, recognition, synthesis and pathology of speech, and linguistics. Its straightforward user interface and technical quality also make it very useful in studies of the singing voice, musical acoustics, sound and vibration analysis, bioacoustics, and forensics.

Historical Note

When the original *SpeechStation* was first introduced in 1989, the speed and computing power of desktop computers were modest compared to the requirements of practical digital signal processing. The system therefore included, in addition to its software, a specially-designed circuit board that contained analog-to-digital conversion, digital-to-analog conversion, a Motorola 56000 DSP processor, high-speed memory, and a microphone preamplifier with gain control. With these resources, the *SpeechStation* enabled users with inexpensive desktop computers to obtain results with the accuracy needed in serious research.

When the first *SpeechStation* was developed, desktop units offered breathtaking “turbo” speeds of 12 MHz; now, clock rates of more than 500 MHz are common, cheap and will likely be obsolete in a few years. This increase in CPU speed, together with the availability of new software development tools, has enabled Sensimetrics to produce an all-software version of the *SpeechStation*, the present *SpeechStation2*, that is considerably faster and better than the original product. Moreover, by eliminating the need for special hardware, we have lowered the cost of *SpeechStation2* substantially.

We wish to thank...

Development of the *SpeechStation* and *SpeechStation2* could not have been completed without support from a number of organizations and individuals. Most important, initial and major support came from the U.S. National Institutes of Health (NIDCD) in the form of an SBIR grant for the *Speech Synthesis Workstation* (DC00624).

Many individuals helped evaluate and test the system during its development. We owe special gratitude to Dr. Amália Andrade and Dr. Çeu Viana of the *Linguistics Centre of the University of Lisbon*.

Kenneth Stevens and James Pickett have been continuing sources of good advice and encouragement.

The software was created by Jason Carr. He was helped by Julio Ortiz, Eric Carlson, and Robert Beau-doin. Jens Jorgensen created some graphic elements of the software. Robert Berkovitz contributed to the functional and visual design of the software and, along with Patrick Zurek, wrote this manual.

Requirements for Use

What is needed to use *SpeechStation2*

There are two essential requirements for use of *SpeechStation2*.

1. A computer running Windows 98/Me/NT/2000/XP.
2. A Windows-compatible audio card, or the equivalent on the computer's main circuit board.

In order to make recordings, you will need a microphone or tape recorder that can be connected to the audio card.

To listen to computer files of sound recordings, you need to have headphones or loudspeakers connected to the audio card. Loudspeakers designed for computer use are supplied in pairs; a headphone connection is often provided on one of the loudspeakers. Headphones are usually preferable to loudspeakers when using *SpeechStation2*, because the listener is partially isolated from room sound and can hear the speech more clearly.

Printer

To print copies of spectrograms and other displays, you need a printer with graphics capability. *Speech-*

Station2 will work with any Windows-compatible printer, laser, ink-jet or other types. *SpeechStation2* has been designed to provide printouts of the highest quality permitted by the printer, including color where appropriate.

Computer speed

Although it has not been specifically designed for the fastest computers, *SpeechStation2*, like most programs, will run more rapidly on a faster computer. If speed is important to you, your computer should have a processor running at 133 MHz or faster.

Monitor

As with any software that makes intensive use of high-resolution graphics, the computer's monitor should be as large and as sharp as you can afford, as you will be looking at it closely, perhaps for long periods of time. The *dot pitch*, which defines the smallest separation of visual units that the monitor can display, should be no more than 0.28 mm, regardless of the monitor size. High-quality monitors are available with a dot pitch as small as 0.25 mm (e.g., *Sony Trinitron*). Most large computer manufacturers (*Micron, Dell, Compaq, HP*) offer only monitors with a dot pitch of 0.28 mm or less.

To use the Real-Time Spectrogram feature effectively, the clock rate on the computer's CPU should be at least 133 MHz. See the section of this manual on the **Real-Time Spectrogram.*

Installation

CD-ROM

Put the *SpeechStation2* disk in the CD-ROM drive of your computer. Click the Windows *Start* button on the task bar, then click Run, and in the resulting dialog box type “D:\Setup.exe” then click OK. If your computer’s CD-ROM drive has a different drive letter than D you will need to type its drive letter followed by “:\Setup.exe” instead. When instructed to do so, follow the instructions that appear on the screen.

Adobe Reader

This manual is also available on the CD-ROM as an Adobe Reader .pdf file. Adobe Reader is available free for download from the Adobe website at: www.adobe.com. After you have downloaded and installed *Adobe Reader*, you can view the instruction manual for *SpeechStation2* in color on your computer screen by double-clicking the **SSManual.PDF** file.

Using the mouse

Throughout the instructions in this manual, the operations **click**, **drag** and **select** are used. These are basic methods of using the mouse control device employed in Microsoft Windows, and users of *SpeechStation2* need to be familiar with them. Where no other description is given, references to the mouse button refer to the left button of the mouse.

The software provided with most computers includes demonstration and practice software that allows a user to become familiar with these mouse operations and to adjust their sensitivity.

Opening *SpeechStation2* Automatically

The Microsoft Windows operating system provides a way of linking files of the same kind to a specific program. For example, Windows is usually set so that when the name or icon of a file with the extension .WAV is clicked, the *Media Player* program starts.

This kind of linkage can be set up between *SpeechStation2* and .WAV files, so that whenever the icon or name of a .WAV file is clicked with the mouse, *SpeechStation2* opens automatically and loads the file. Although a .WAV file can still be loaded in the usual way, this replaces the steps of opening the program, finding the file to be loaded, selecting it, and clicking the “Open” button. Of course, if *SpeechStation2* is already started, you would use the latter method.

Linking .WAV files and *SpeechStation2*

The following steps will set up a connection between .WAV files and *SpeechStation2*.

1. Click **Start > Programs > Windows Explorer**.
2. A window labeled “Exploring” will appear. In the menu at the top of the window, select **View > Options**.
3. In the Options box, click the tab, “File Types.”
4. In the window “Registered file types,” scroll down to see the listing, “Wave Sound.”
5. Click once on “Wave Sound,” then click the button marked “Edit... ”

6. A window opens, "Edit File Type." Remove any items listed in the "Actions" window by clicking on each item and then on the "Remove" button. Click on the button marked "New."

7. A window opens, "New Action." In the "Action" window type the word, "Open."

8. In the window, "Application used to perform action:" you need to insert the path to the *SpeechStation2* program. To do this, click on the *Browse* button, find the program SPEECHSTATION2.EXE and click "Open." The program is automatically installed in the following path, "**C:\Program Files\Sensimetrics\SpeechStation2\SpeechStation2.exe**" unless you have chosen a different location during installation. The quotation marks should appear in the text box showing the path.

9. After the final quotation mark, add *one space*, then "%1" *including the quotation marks*.

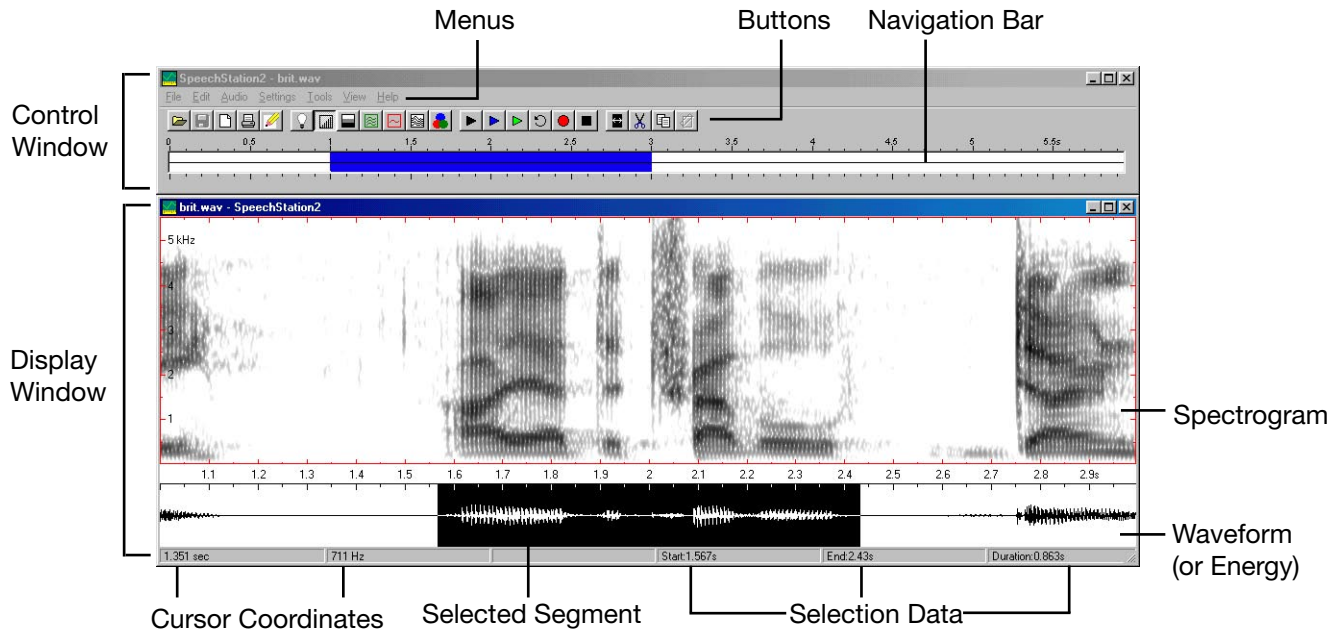
10. Do NOT check the box marked "Use DDE."

11. Click "OK." Click on the word, "Open" and then on the button marked "Set Default."

12. Click *Close*, then *Close* again, and from Windows Explorer, click **File > Close**.

If this procedure is carried out, clicking on any file with the .WAV extension will automatically open *SpeechStation2* with that .WAV file loaded.

The Main Screen



Overview

When you start *SpeechStation2* a small window at the top of the screen, called the control window, will be displayed. Loading a file brings up a larger window, called the display window, below the control window.

SpeechStation2 works with three types of signal segments: 1) the whole file; 2) the portion of the file displayed in the display window (referred to as the “displayed segment”); and 3) a portion of the currently-displayed segment called the “selected segment” or simply, the “selection.”

The Control Window

The control window contains the menus, the but-

tons, and the navigation bar, as well as a title bar. If a file is loaded, the name of the file is displayed in the title bar; if multiple files are loaded, the name of the currently-active file is displayed.

The **menus** provide access to all *SpeechStation2* commands. These menus are described on the following pages.

The **buttons** provide quick access to the most frequently-used commands. The button commands are described along with the menus, and are summarized on page 15.

Some commands also have **keystroke equivalents**. These are described along with the menus and are also summarized on page 15.

The **navigation bar** shows both the length of the current file and the portion of the file currently displayed in the display window. The length of the file is indicated by the numerical scale associated with the navigation bar. The location and extent of the displayed segment is indicated by the blue strip in the navigation bar. See *Loading and Navigating a File* for further description of the use of the Navigation Bar.

The Display Window

The display window has two main parts. The upper, larger, part shows the **spectrogram** of the displayed segment and the lower, smaller, part shows the **waveform/energy** of the displayed segment. The spectrogram and waveform/energy plots are always exactly time aligned. The display window also has a title bar (top) and a status bar (bottom) which gives certain numerical data.

The **spectrogram** shows the results of a temporal-spectral analysis of the signal contained in the displayed segment, with frequency plotted on the vertical axis, time on the horizontal axis, and intensity coded by level of gray. Many aspects of the spectrographic analysis can be controlled within *SpeechStation2* (see *Spectrogram Analysis Options*), as can the details of how the results are displayed (see *Spectrogram Display Options*).

When the mouse's cursor is placed anywhere in the spectrogram, the **cursor coordinates**, time and frequency, are shown in the status bar.

The **waveform/energy** display shows either the time waveform or the energy of the displayed segment.

A portion of the displayed segment can be selected for further analysis. This **selected segment** is most easily chosen by clicking and dragging over the desired portion of the waveform/energy display with the mouse. The selected segment will be highlighted. See *Loading and Navigating a File* for a more detailed discussion of segment selection.

Other Important Features

SpeechStation2 offers several tools to assist in the analysis of speech signals. See the relevant sections of this manual for more discussion of the following features.

Clicking anywhere in the spectrogram will bring up the **Spectrum Viewer**. This tool allows several options for detailed spectral analysis.

Formant tracking can be performed with visual display (or numerical output to a file) of the results.

Pitch tracking can also be performed with visual display (or numerical output to a file) of the results.

In addition to the default display, a spectrogram can be shown as a **waterfall plot**.

A **Real-Time Spectrogram** of any input signal can be shown as a continuously moving spectrogram.

Vowel-space plots show the joint distribution of first and second formant frequencies.

To allow signal filtering, a facility is provided for specifying the magnitude response of a **digital filter**.

Main Screen Commands

All of the commands for the main screen are accessible from the pull-down menus. In addition, many of the most frequently used commands are available as buttons and/or keystrokes. Commands that can be issued by a button are indicated by a picture of the button; keyboard equivalent commands, when available, are shown to the right of the main command. A summary of the button and keystroke commands is given on page 15.

File Menu

New

Ctrl-N



Creates a new spectrogram window. Previous spectrograms are not lost when a new spectrogram window is created.

Open

Ctrl-O



Produces a *File Open* dialog box. Opening a file causes it to be loaded into *SpeechStation2*.

The *File Open* dialog box contains a check box that allows a file to be marked as “Read Only.” If this box is checked, the file cannot be edited.

Opening a file with the Open command (without having created a new window first with the *New* command) will close the file in the most recently accessed display window and replace it with the file being opened. To open a file without losing the current displayed file, first use the **New** command to create a new display window and then use the **Open** command to open the file.

Note that *SpeechStation2* does not have a Close command. To close a file, click on the “x” in the upper right-hand corner of the display window. In either way of closing, if changes have been made to the file, a dialog box will appear to ask if the changes should be made before closing the file.

Convert .DAT File

The original *SpeechStation* saved speech recordings in a file format different from that of *SpeechStation2*. This function creates a new file with the .WAV extension but saves the older .DAT file.

Save



Saves the currently displayed file, with any modifications that have been made.

Save As...

Allows a file to be saved with a new name.

Display Information

Ctrl-I

This command brings up a dialog box showing the start, end and duration of both the displayed segment and the current selection, if any. These values can be changed in this dialog box. See *Loading and Navigating a File* for further information.

Print



Opens the print dialog box. See the section on *Printing* for more detailed information.

Exit

Closes the program.

Edit Menu

Undo

Cancels the last executed action.

Cut



Cuts the selected segment and puts it on the clipboard. The change becomes permanent when the file is saved.

Copy



Copies the selection to the clipboard.

Paste



Inserts the data currently on the clipboard at the location of the waveform cursor or replaces the current waveform selection with the clipboard data.

To create an insertion point, click once at the desired location on the waveform, using care to be sure that the mouse is not moved sideways while clicking. The line cursor blinks when this has been done correctly.

Delete

Removes the selected data from the file. The removed data cannot then be pasted, as when the *Cut* command is used.

Clear

Removes the selected data from the file and substitutes an equivalent period of silence.

Audio Menu

Ctrl-Z

Play Selection

F5



The black arrow plays the current selection.

Ctrl-X

Play Display

F6



The blue arrow plays the displayed segment.

Ctrl-C

Play File

F7



The green arrow plays the entire file.

Ctrl-V

Repeat Play



When selected, modifies the action of the other play commands, causing them to loop playback until the Repeat Play button is clicked again, the Repeat Play menu item is deselected, or the Stop Playing command is selected.

Record



Initiates recording setup, but does not actually start recording. See the section on *Playback and Recording* for a complete description of this function and its options.

Stop Playing



Stops playback and returns to the start of the file.

DEL

Settings Menu

Analysis Mode

Provides a choice between two forms of spectrogram, LPC (linear prediction coefficient) and FFT (fast Fourier transform).

LPC

This analysis method is used primarily for showing the spectral envelope of the signal. Spectral details are highly smoothed, unless the LPC order (the number of coefficients employed) is increased well above the default value of 12. This method of analysis is discussed in the *Spectrogram Analysis Options* section of this manual.

FFT

The standard method of spectral estimation by digital signal processing. The characteristics of this method are described in the *Spectrogram Analysis Options* section of this manual.

Window Size

Selecting this menu item offers the user a choice of four window lengths for analysis: 64, 128, 256 or 512 samples. The default is 128.

Window Type

Four types of windows are provided for analysis: *Blackman*, *Hamming*, *Hanning* and *Rectangular*. Their characteristics are discussed in the *Spectrogram Analysis Options* section of this manual.

LPC Order

Enables the user to select a new value for the order of the LPC analysis.

Pre-emphasis Filter



To make high-frequency information more readily visible in the spectrogram, a pre-emphasis of 6 dB per octave is applied to the signal displayed in the spectrogram. This pre-emphasis is implemented by default in SpeechStation2, but it can be disabled by clicking the button or selecting the menu item. The pre-emphasis is not applied to the signal displayed at the bottom of the screen (either Waveform or Energy).

Gray Scale



This command brings up a dialog box that allows adjustment of the mapping of dB levels to gray scale values. These adjustments are described in detail in *Spectrogram Display Options*.

Color



This command toggles the spectrogram display between color and black and white.

Tools Menu

Spectrum Viewer

Opens the *Spectrum Viewer* window. The window is normally empty when first opened in this way. After a mode is selected, spectrum and waveform data can be plotted in the viewer. This feature is described fully in *Spectrum Viewer*.

Filter

Opens the dialog box allowing a digital filter to be applied to the data. This function is described in detail in the *Filter* section of this manual.

Waterfall Plot



Makes a waterfall plot of the selected data. This type of plot is treated in detail in the section *Waterfall Plot* in this manual.

Vowel Space Plot

Uses LPC analysis to locate the two formants with lowest frequencies and plots their locus as it changes with time. This plotting method is described in detail in the section of this manual entitled *Vowel Space Plot*.

Formant Tracks



Uses LPC analysis to draw the center frequencies of formants as a set of green tracks across the spectrogram. If the *Clear Spectrogram* function is selected from the *View* menu first, the spectrogram is removed and the formant tracks are drawn on a white background. This analysis is described in detail in *Formant Tracking*.

Export Formant Tracks

This function automatically transfers formant tracks, whether or not they are currently displayed, to an ASCII text file. This file can then be printed or used as data by spreadsheets or other programs to perform analysis of the formant tracks.

Real-Time Spectrogram

Presents a continuously-moving spectrogram of an input signal. See the section of this manual entitled *Real-Time Spectrogram* for a detailed treatment of this function.

Pitch Track Plot



Draws a plot of pitch (F_0) as a function of time in the spectrogram space in red. To see the pitch value for any location on the track, place the mouse cursor at that time and read the pitch from the status bar. This and related commands are described fully in *Pitch Tracking*.

Export Pitch Track

Transfers pitch track data to an ASCII text file.

Pitch Multiplier

Allows a pitch track to be plotted at a multiple of its actual value for better visual resolution. The actual pitch value is displayed in the status bar.

Average Energy Threshold

If the *average* selection is checked, the pitch will be estimated whenever the signal energy exceeds the average energy level for the displayed portion of the

View Menu

waveform. For details on the calculation of signal energy, see *Spectrogram Display Options*.

Custom Energy Threshold

If this option is checked, the energy plot automatically appears, replacing the waveform in the window under the spectrogram. A movable horizontal line also appears, which can be raised or lowered by holding down the mouse button and dragging the line to set the threshold for pitch estimation. Threshold adjustment can be useful for avoiding false triggering of pitch estimates by background noise.

Selection



Replaces the current spectrogram with a spectrogram of the selection. The selection must be at least 512 samples in length for its spectrogram to be displayed.

Energy



This command toggles between a waveform plot and an energy plot in the lower part of the display window.

Redraw Spectrogram



The spectrogram is redrawn without formant tracks or a pitch track, using the current settings.

Clear Spectrogram

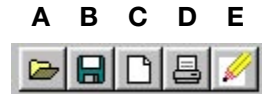
Removes the spectrogram from the screen, but leaves the navigation bar unchanged. Pitch and formant plots can be plotted on the clear background that results, and will be printed in this way if the *Print* command is invoked.

Spectrogram Grid

Places a red grid on the spectrogram that has lines at the major time and frequency divisions. The grid is removed by deselecting this menu item.

Summary of Button and Keystroke Commands

File control



A	open a file	Ctrl-O
B	save the current file	
C	open an empty display window	Ctrl-N
D	print the current display(s)	
E	draw the current display again	

Analysis and Display



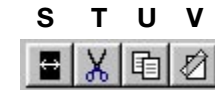
F	select energy or waveform plot
G	switch high-frequency pre-emphasis on/off
H	adjust mapping of level to gray scale
I	find and plot formant frequencies
J	extract and plot pitch (F_0)
K	make a waterfall plot from the selection
L	display in gray scale or color

Audio control



M	play the selected segment	F5
N	play the current display	F6
O	play the currently selected file	F7
P	play choice repeatedly	
Q	make a new recording	
R	stop playback	


Editing



S	expand selection to full screen width	
T	cut and hold selected segment	Ctrl-X
U	copy and hold selected segment	Ctrl-C
V	paste held segment at selected place	Ctrl-V

Loading and Navigating a File

Loading a file

The first button , at the extreme left in the row of control buttons, is the *Open File* control. If this control is selected, or the command *Open File* is selected from the *File* menu, a list of available files in the current folder will be displayed. If the folder shown is not the one you want, use the standard Windows commands in the dialog box to move to the correct drive and folder.

Double-click on the name of the file to be loaded, or click first on the file name, then on the *Open* button. If you want to open the file and, for safety, want it to be protected against accidental editing changes, click on the check box *Open As Read Only* at the bottom of the dialog box. If you intend to edit the file, make sure that the option is not checked. If you forget that the option is checked and later want to edit the file, you will need to close the file, and reload it with the box not checked.

Loading Multiple Files

An important feature of *SpeechStation2* is its ability to load more than one file at the same time. This makes it possible to compare spectrograms of different files, or of real and synthetic utterances, for example. The same facility can be used to examine and compare two locations in the same file by making a renamed copy of the file, and then loading both the original and the copy. Note that unless this procedure is followed, the same file cannot be loaded into two different windows.

In order to maintain a display window of a first file and then load a second file (or more), it is necessary to do so using the **File > New** command. If you use

the **File > Open** command to open a file, the new file will displace the one in the last accessed display window.

Loading .DAT Files

The original *SpeechStation* did not load or create .WAV files compatible with Microsoft Windows, but used its own special format. These files were saved with the extension .DAT.

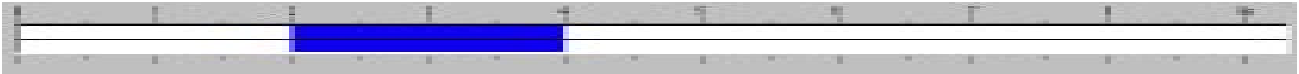
Users of *SpeechStation2* who wish to work with such .DAT files can do so by clicking the *File* menu and selecting the item *Convert .DAT file*. The .DAT file will be converted to a .WAV file and then loaded.

Initial display

When a file is loaded, or when a new file has been recorded, the spectrogram that appears shows the first two seconds of the file. The right end of the navigation bar near the top of the screen will show the length of the entire file in seconds or milliseconds (for files shorter than 1/2 second). The first two seconds of the navigation bar scale will be filled in blue. The spectrogram will show this segment of the data, as will the waveform display, which is synchronized with the spectrogram.



The navigation bar shows that 2 seconds are displayed in the spectrogram and that the file is 9.25 seconds long.



After the mouse is clicked in the white area (right of the blue bar) the blue bar and the displayed segment jump 2 seconds.



The blue bar can be stretched or shortened by dragging one of its edges with the mouse while pressing the Shift key..

Moving through a file

To see a different two-second segment of the file, drag the blue bar to the right with the mouse until the bar is at the desired place on the time scale, then release the mouse button.

To move through the file in fixed-size steps, click the mouse in the white area to the right of the blue bar. Each time this is done, the bar will jump to the right by exactly its length. To change the size of the jump, stretch the blue bar by the method described in the next section.

Another way of stepping quickly through a file is provided by the function keys **F2** and **F3**. These keys cause the blue strip to move left (**F2**) or right (**F3**) by one half its length. This results in only half the screen presenting new data, with the other half presenting data displayed prior to the jump. This feature is useful for providing some context, or memory, for your location within a signal.

Changing the size of the displayed segment

To see a segment of the file larger or smaller than two seconds (but with a minimum length of 512 samples), hold down the *Shift* key, put the cursor near the right or left edge of the blue bar, and drag right or left until the bar is the desired length. An-

other way to extend the bar is to place the mouse pointer in the navigation bar slot at the place where you want the blue bar to end. Hold down the *Shift* key and click, and the blue bar will snap to the new position.

Summary of navigation bar operations

Here is a summary of the operations used with the navigation bar.

Move right or left. Drag the blue bar by putting the cursor on it, holding down the mouse button, and sliding the mouse in the desired direction.

Jumping by one bar length. Click once in the white area at the right or left of the blue bar.


Jumping by one half bar length. Use **F2** to jump left, and **F3** to jump right, by one-half the length of the blue bar.

Stretching or shrinking the blue bar. Hold the *Shift* key down while dragging either end of the blue bar to make the bar longer or shorter.

Making a new blue bar. Put the cursor anywhere in the white space of the navigation bar and drag to create a new blue bar and eliminate the old one.

Selecting a segment of the spectrogram

Some operations in *SpeechStation2* are performed on a selected part of the displayed segment. To make a selection, put the cursor into the waveform or energy plot under the spectrogram, left-click and drag over the desired segment, then let go of the mouse button. The selection will appear with reversed colors. Only one selection at a time can be made.

The mouse can be used to slide, lengthen, or shorten the selection bar, as described for the navigation bar. However, clicking outside of the selection does not cause the selection to jump, as is the case with the navigation bar. Clicking outside the selection eliminates the selection and establishes a new cursor position. To expand the selected segment horizontally to fill the display window, click on .

Note that segment selection is not achieved by clicking on the spectrogram. Clicking on the spectrum causes the *Spectrum Viewer* window to appear for detailed spectral analysis (see *Spectrum Viewer*).

Cursors

As the mouse cursor passes over the spectrogram, the time-frequency coordinates of the cursor are displayed in the two left boxes in the status bar at the bottom of the display window.

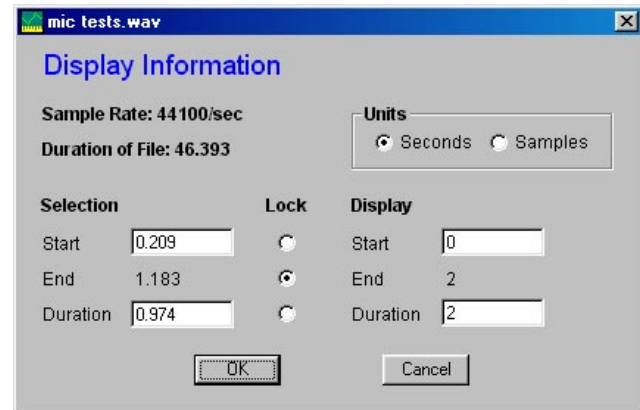
Display Information

Ctrl-I

This command brings up a dialog box showing the start, end and duration of both the display and the current selection, if any. These values can be changed in this dialog box. At any time one of the three values is fixed; use the 'Lock' button to select which. Either of the other two may be changed, and the third will be recomputed from the changed and the locked value.

Note that the endpoints for a selection cannot fall outside those of the display. In other words, the start and end of a selection each must be greater than or equal to the start of the display, and less than or equal to the end of the display. Changes in the display range that violate these constraints result in changes to the selection so that the constraints are satisfied.

A segment must be at least 512 samples in length to be displayed; if it is not, a blank display window will appear. If the segment has more than 512 samples but fewer than the width of the display window (in pixels), the spectrogram will only partially fill the display window horizontally and there will be a black area at the right. The width of the display window in pixels depends on its width in centimeters, the size of your monitor, and the screen resolution you have chosen in your Windows display settings.




The Display Information box. The time locations and durations of the displayed segment and selection can be observed and, if desired, modified by entering new values.

Playback and Recording

Playback

A direct check that your soundcard can playback audio signals can be made by clicking on the button

showing the blue triangle . You should hear the signal contained in the display window. Clicking on the green triangle button plays the entire file. Clicking on the black triangle will produce sound only if a selection has been made.

If you hear no sound...

If you cannot hear sound when the *Play Display* or *Play File* button is clicked, check carefully to be sure that the volume control for the loudspeakers or headphones is not turned off, that they are connected correctly, that the power switch for the loudspeakers is on, and that the loudspeakers are connected to a power source.

If the loudspeakers are not turned off, it is a good idea to set their volume low, so that when you fix the problem, the sound will not suddenly be played at a very high level. This is especially important if you are listening with headphones.

If turning up the volume and clicking *Play Display* still produces no result, you may need to adjust the Windows volume control. Before sound can be recorded or played, the playback and recording levels on the Windows *Volume Control Mixer* must be set correctly. If this is not done, playback and recording will not work even though all other settings and electrical connections are correct.

Setting the Windows volume control

To set the Windows volume controls, first look for a small loudspeaker icon in the lower right corner of the screen. Click on the icon *once* and a small volume control panel will appear with a control slider. Use the mouse to drag the slider to the top, then click on the desktop to make the control disappear. Try playing sound again. If there is still no sound, click *twice* on the loudspeaker icon to display a larger volume control mixer, with sliders for *Volume Control*, *CD Audio*, *Line In*, and other sources. Make sure that at least the *Volume Control*, *Line-In* and *Wave* sliders are at or near the tops of their ranges by dragging the sliders upward with the mouse, and that neither slider is muted in the check box below the slider.


If there is no small loudspeaker icon at the bottom right corner of your screen, you can access the *Volume Control* panel in another way. Click on the "Start" button at the lower left corner of the screen. When the choices pop up, click successively on **Programs > Accessories > Multimedia > Volume Control**.

If all of the controls are set correctly and there is still no sound, the loudspeakers or headphones may not be connected to the correct jack on the computer, or the power for the loudspeakers may not be switched on (worth checking again). There are usually three jacks on the back of the computer or the audio card: for a microphone; an input for a line-level external device (for example, a tape recorder); and an output from the audio card (for loudspeakers or headphones). All of the jacks have identical 1/8-inch (3.2 mm) openings, so that they look and feel the same, and as they are on the rear of the computer, they are often difficult to see to identify by their markings.

Recording

Your computer's sound card can accept a microphone, a tape recorder, an FM tuner, CD player or other source. The sound card contains an analog-to-digital (A-D) converter that will sample the analog signal onto the computer's hard disk. All tape recorders that use full-size cassettes, except for the very cheapest, are suitable for speech recording. See *Appendix B: Making Better Speech Recordings* for more information on selecting a microphone and recording medium.

The New Recording box

Start by clicking on , or select *Record* from the *Audio* menu. The *New Recording* box will then be displayed. (See example at right).

Choose a name for the new file

In the box marked *File Name* type the name to be given to the file. The extension *.WAV* will be added to the file name automatically. To change the directory in which the file is saved, click on the button marked "..."

Choose mono or stereo recording

Speech recordings are usually made on one channel only (mono); note that *SpeechStation2* selects the left channel and mono by default. If you choose to record in stereo, remember that *SpeechStation2* will treat each channel as a separate file during analysis. You can change the channel used for mono recording to the right channel if there is reason to do so, but it is probably best to make all single-channel recordings on the same channel to avoid confusion.



If Record is selected from the Audio Menu, or the button with the red circle is clicked with the mouse, the window shown above appears on the screen.

Select a sampling rate

The standard Windows sampling rates are 11025 Hz, 22050 Hz and 44100 Hz. The frequency range of the resulting recording will be half the sampling rate.

Note: Some sound cards may not be able to record at the three standard sampling rates. If you are unable to record at one sample rate, try one of the others.

There are trade-offs in the choice of a sampling rate. The highest sample rate of 44100 Hz will be sure to capture all of the high-frequency components in speech. However, this maximal sample rate results in a displayed frequency range of 0 to 22050 Hz, which reduces the region below 2200 Hz to only about 1/10 of the height of the spectrogram display. Many important patterns of speech spectrograms in the low-frequency region may be difficult to identify on this compressed scale.


Using 11025 Hz, on the other hand, produces clearly-resolved low-frequency features but no display at all of components higher than 5500 Hz.

Another issue that arises in the choice of a sample rate is storage. For a signal of a given duration, the highest sample rate will require four times the storage needed with the lowest sample rate. For typical speech analysis applications, we recommend a compromise of 22050 Hz.

(Note that although *SpeechStation2* provides only the three standard sampling rates for recording signals, it will accurately analyze and display signals that have been recorded or created at any sampling rate.)

Set the recording level

Above the *Record* and *Cancel* buttons in the *New Recording* box is a strip like a thermometer that functions as a recording level meter. If the strip is gray, no signal is being received, or its level is too low to register, and there should be no sound from the loudspeakers or headphones. If you are certain that the microphone or recorder are connected correctly, and a signal is being sent to the computer, but no signal is apparent, check the *Volume Control* used by Windows. This can be done by double-clicking on the small loudspeaker symbol at the lower right corner of the computer screen. If the loudspeaker symbol is not present, the volume controls can be displayed by clicking on **Start > Programs > Accessories > Multimedia > Volume Control**. When the volume control mixer is visible, click **Options > Properties** and click the circle next to "Recording." This will allow the recording settings to be checked to be sure that the required inputs are selected and the control sliders are high enough. To see the "meter" level in the *SpeechStation2* mixer,

you need to click the record button .

Your sound card may come with its own mixer in software form. If so, follow the manufacturer's instructions for setting the controls.

The recording level bar in the *SpeechStation2* recording window will flicker when a signal is being received, and will be blue if the signal level is within the normal recording range. Ideally, the flickering blue bar should reach to the middle of the slot, turning yellow occasionally as it goes higher. If the level is high enough to approach distortion, the bar will turn red. While it is permissible to reach this level occasionally, consistently excessive level will result in a distorted recording that cannot be analyzed properly.

Note that the mere presence of red in the level indicator does not indicate distortion, unless the level indicator also strikes the right-hand edge of the slot repeatedly. An easy way, and perhaps the best way, to check recording level is to make a test recording and examine the waveform under the spectrogram. If the recording level is too low, the waveform will be small and extend only slightly above and below the center (zero) horizontal line. A signal level that is too low will result in a poor signal-to-noise ratio and a loss of information at the lowest level.

If the level is too high, the waveform will meet the upper and lower edges of the waveform window. In this case, high level information will be lost, and distortion components will be added to the signal. This distortion is disturbing to hear and confusing during analysis. During speech recording, this is most likely to occur during brief bursts, especially if the talker's mouth is too close to the microphone.

Start then stop recording

Clicking on *Record* will begin the recording. The marking on the button will change from *Record* to *Stop*; the next time the button is clicked, recording will stop. Because the recording process is given priority over metering, the recording level indicator is updated less frequently during recording. The signal being recorded should be monitored through the computer's loudspeakers, if a tape recorder is being used as the source, or through headphones if a microphone is used.

Note that if the loudspeakers are on while recording with a microphone, the microphone may pick up the sound and feed it back to the loudspeakers. This can lead to a loud howling or whistling sound. To prevent this, use headphones to monitor when recording with a microphone and turn off the loudspeakers.

When the recording is complete, click on *Stop*. The recording control box will disappear and a display window will appear, showing the first two seconds of the file that has just been recorded.

Check the file and its location

When the recording is finished, step through the file using the navigator bar while examining the waveform window under the spectrogram. If there are unexpected loud parts of the recording that have exceeded the dynamic range of the system, the recording may need to be made again. You can test the sound of such segments by selecting them and playing them back.

If you are satisfied with the recording, it is then a good idea to check that you have given it an appropriate name and not left it with the default name ("Untitled-2.wav" for example). If you have forgotten to give the file the name you wanted to use, you can change the name with the Rename facility of Windows Explorer.

Spectrogram Analysis Options

SpeechStation2 offers two spectrographic analysis modes, one based on the Fast Fourier Transform (FFT) and the other based on Linear Predictive Coding (LPC).

FFT Mode

In the early days of speech analysis, spectrograms were made by passing the signal through a series of narrow bandpass filters. The magnitude of the output of each filter was plotted as a horizontal trace that varied in darkness as the level of the output varied over time. Today, computers achieve a similar analysis by sampling the signal and calculating the level in narrow frequency bands using the discrete Fourier transform. (The “Fast” Fourier transform, or FFT, is an algorithm for computing the discrete Fourier transform very efficiently). This analysis mode produces spectrograms like those obtained historically using analog filters.

Time-Frequency Trade-off

Consider the analysis of a segment of a signal sampled at F_s Hz that is N sample points in length. FFT analysis of this N -point segment, which is N / F_s seconds in duration, produces an estimate of amplitude¹ in each of $N / 2$ spectral bins evenly spaced from 0 to $F_s / 2$ Hz, with spacing between adjacent bins equal to F_s / N Hz. We can therefore speak of the time resolution as being N / F_s and the frequency resolution as being F_s / N .

This reciprocal trade-off between frequency and time resolution is an important property of FFT analysis.

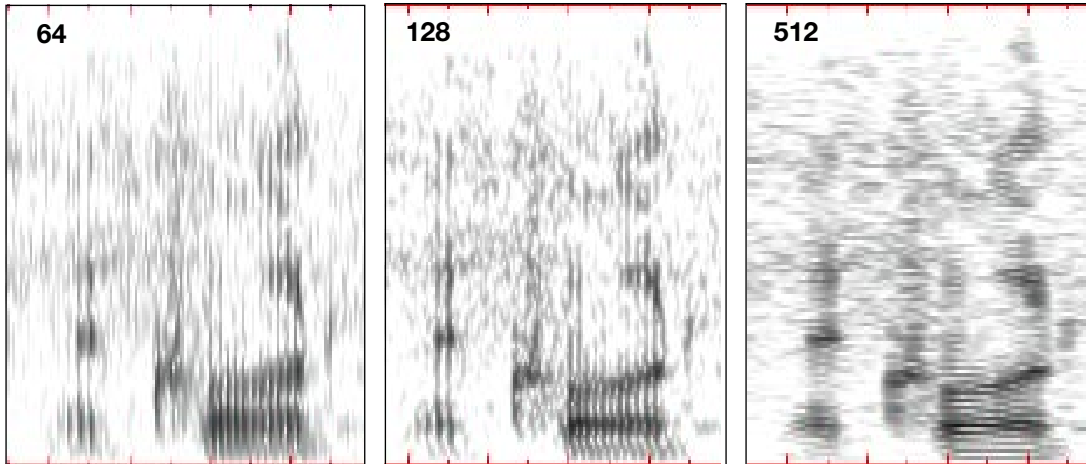
¹The FFT also gives the phase in each bin, but phase is not used in *SpeechStation2*.

Greater frequency resolution (i.e., finer separation of frequency components) can be achieved only by analyzing a longer segment of the time signal (i.e., increasing N), assuming the sample rate is fixed. Conversely, increased time resolution (i.e., showing more rapid temporal changes), can be achieved only by reducing frequency resolution (i.e., decreasing N). Of course, the same trade-off characterizes the spectral analysis achieved by analog filters as well. The trade-off between resolution in time and resolution in frequency has implications for the kind of analysis you wish to perform on a speech utterance.

Selecting an Analysis Window

SpeechStation2 provides several choices for both the length and shape of the analysis window. The effect of changing the length of the analysis window can be readily appreciated by examining spectrograms of the same utterance made using windows of different sizes, such as the examples shown on the following page. Use of a 64-sample window (left) accurately resolves the time of occurrence of each glottal pulse, but shows none of the harmonics in the voiced segment. A 512-sample window (right) resolves harmonics, but fails to show individual glottal pulses. Neither one of these extremes shows the formants as clearly as does the compromise 128-sample window (center), which is used as the default in *SpeechStation2*.

FFT analysis can be applied to a data segment directly, or it can be applied after the data segment has been “windowed,” or multiplied by a specific “window” function in time. The purpose of windowing is to restrict the range of frequencies that contribute to a spectral bin. If no window is used, which is equivalent to using a “rectangular” window, then



Representations in three spectrograms of the words, "the quote," using windows of 64, 128 and 512 samples. The 64-sample window shows good time resolution, and the 512-sample window good frequency resolution; the 128-sample window is a compromise.

the effective width of the bin is wider than if a window is used that smoothly tapers the beginning and end of the segment to zero.

SpeechStation2 offers users a choice of four window shapes: Hamming, Hanning, Blackman, and rectangular. You can view the effects these window shapes have by analyzing a pure sine wave with each in the *Spectrum Viewer*. You will see that there are not large differences among the three non-rectangular window shapes. However, there is a very large difference between those three and the rectangular window, which causes substantial spreading of energy.

LPC mode

Identifying formant peaks in an FFT spectrum is often difficult. It is for this reason that *SpeechStation2* provides Linear Predictive Coding (LPC) as an alternative to the Fourier Transform for spectrogram or spectrum analysis.

LPC analysis assumes that a signal is the output of a causal linear system, and therefore that each sample can be expressed as a weighted sum of previous samples. When LPC is used to model a speech signal, it is usually assumed that the vocal-tract system is an all-pole filter and that the input to the system is an impulse train. As a result, the filter's frequency response derived by LPC will include the shape of the voice-source spectrum, in addition to the vocal-tract resonator characteristics. Because of these assumptions, LPC analysis is usually most appropriate for modeling

vowels, which are periodic and for which the vocal-tract resonator does not usually include zeroes.

The LPC spectrum

The frequency response of the LPC filter is referred to as the LPC spectrum. Note that because it is a frequency response, if the order of the filter is chosen correctly (see below), it should not contain harmonics of the voice pitch; it should only represent the combination of the spectral shape of the glottal source and spectral shaping imposed by the vocal tract. The LPC spectrum (when the order is appropriately chosen) usually approximates the envelope of the (Fourier) magnitude spectrum.

SpeechStation2 uses Burg's algorithm, a so-called "lattice method", to compute the LPC coefficients and, from them, the LPC spectrum. The time/frequency and analysis window concerns that apply to FFT analysis apply to LPC analysis as well. For a relatively stationary speech signal, a longer window will result in improved resolution of the formant peaks. A shorter window is necessary to resolve rapid time variations in the spectrogram. The window should always be long enough to include at least two pitch periods. As discussed in the previous paragraph, the LPC spectrum should not contain pitch harmonics, and thus, unlike the FFT, the length of the window will not affect resolution of the harmonics.

Order of the LPC filter

The order of an LPC model is the number of poles in the filter. Usually, two poles are included for each formant, and 2-4 additional poles are included to represent the source characteristics. The number of

formants depends on the sampling rate and the average formant spacing. For adult speakers, average formant spacing is in the 1000-Hz range for males and in the 1150-Hz range for females (assuming the speed of sound is 34,000 cm/s, and vocal-tract lengths of 17 cm and 15 cm for males and females, respectively). For example, if the sampling frequency of a speech signal is 10 kHz and the cutoff frequency is 5 kHz, five formants would be expected in vowel spectra from male speakers, but only four in vowel spectra from females. Therefore, an LPC model should have an order of 12-14 for males and 10-12 for females. For a 22050-Hz sampling rate and an 11025-Hz cutoff frequency, the order would have to be increased to 24-26 for males and 22-24 for females. In these cases, the poles of the models will be reasonable estimates of the formant frequencies.

In *SpeechStation2* the order of the LPC analysis can be set through the Settings menu. The default value is 12, which is appropriate for signals sampled at 8-11 kHz. For higher sampling rates, the order should be increased.

For a given signal, as the order of the model is increased, the LPC spectrum becomes a better and better match to the speech spectrum and, if the window is long enough, will appear to have harmonics. Although that might seem to be a good thing, the poles of the model are no longer good estimates of the formant frequencies, and it will no longer be easy to identify formant peaks. Therefore, it is best to follow the recommendations given above when setting the LPC order. For the Formant Tracking feature, described later, it is essential that the LPC order not be set too high.

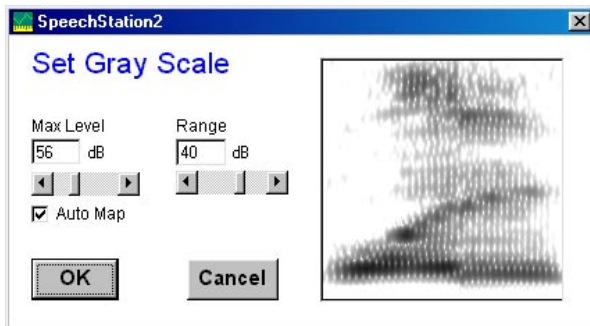
Spectrogram Display Options

Gray scale adjustment

To make the spectrogram as clear as possible, the gray scale can be adjusted. Default settings can be applied automatically and work well for most speech signals. However, they can easily be changed whenever necessary.

To experiment with the gray scale settings, bring up the gray scale control by clicking on *Settings > Gray Scale* in the *Settings* menu, or the gray scale push-

button .

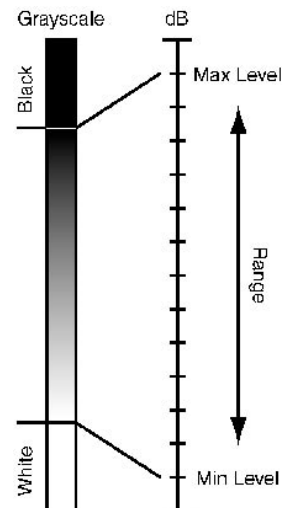


The gray scale control enables two values to be set: the *maximum level* and the *dynamic range*. Levels *above* the maximum level will be displayed as black, and will therefore be indistinguishable from the maximum level. The maximum level should ideally be chosen so that it corresponds to the highest level reached in any of the spectra in the current display. This level is automatically detected and set by *SpeechStation2* in the *Auto-Map* mode, which is the

default. To switch off *Auto-Map*, click on  and click the box marked *Auto-Map* to remove the check mark.

If the *Auto-Map* box is checked, the highest detected signal level in the current displayed segment is automatically mapped to the maximum gray scale value. When auto-mapping is in effect, comparisons of signal levels across different spectrograms should not be made on the basis of gray scale because the level-to-gray mapping is likely to be different from one spectrogram to the next.


The dynamic range sets the difference in dB between the maximum level (black) and the minimum level below which the spectrogram is white. Changes in level below minimum will be invisible. It is obviously desirable to set the dynamic range so that all useful speech information is visible in the spectrogram, while excluding as much low-level noise as possible. As this cannot readily be determined automatically, adjustment of this level should be made using the preview window provided in the gray scale dialog box. The default value of dynamic range applied by *SpeechStation2* is 40 dB.



Waveform/energy plot

The display under the spectrogram is a plot of the signal waveform, and is synchronized with the spectrogram. Although the waveform is usually so compressed that only its envelope is visible, the waveform can always be seen by selecting the *Spectrum Viewer* or clicking on the spectrogram. A separate section of this manual discusses the *Spectrum Viewer*.

The scale of the waveform display extends from the minimum to maximum values possible for a signal sampled with 16-bit resolution: -32767 to 32767. There is no rescaling of this vertical axis when signals are small.

Select *Energy* from the *View* menu or click on  to shift from the waveform display to an energy plot. The plot will change to one showing the relative energy, in dB, for the displayed segment. To return to the waveform display, click the same button again.

The value displayed in the energy plot is $10 \log_{10} E$, where E is the energy of the signal. *SpeechStation2* calculates the energy E at time sample n by the formulas:

$$E_n = \sum_{i=n}^{n+m/2} s^2(i) / E_w$$

$$E_w = \sum_{i=n}^{n+m/2} w^2(i)$$


where s is the signal and w is the window of length m in effect. Note that the energy is normalized by a window-dependent normalization factor E_w .

Grid Display

Selecting *Spectrogram Grid* in the *View Menu* will cause a grid to be superimposed on the spectrogram. The grid is aligned on the major frequency and time divisions. This grid can be useful when the time and frequency of events in the spectrogram are to be estimated visually.

Pre-emphasis Control

To aid in examination of the higher frequency components of speech, a high frequency pre-emphasis of 6 dB per octave is normally applied to spectrograms. This pre-emphasis, which *SpeechStation2* applies by default, can be switched off and on by selecting or deselecting “Pre-emphasis Filter in the

Settings menu, or using the button . This pre-emphasis is not applied to the waveform/energy plot at the bottom of the screen. The *Spectrum Viewer* has its own pre-emphasis control that is independent of the one applied to the spectrogram.


Resizing Windows


Because of technical constraints on displaying spectrograms accurately on the monitor screen, the display window cannot be resized with the freedom that is typical of most Windows applications. The display window can be resized horizontally to any desired size. However, attempting to resize vertically will initially compress the waveform section of the window. Further reduction in the vertical dimension will result in clipping, rather than compression, of the spectrogram.

The *Maximize* button can be clicked to have the display window fill the screen. However, this results in no vertical expansion of the spectrogram; it is only the waveform section that is increased in height.

Pitch Tracking

Displaying a Pitch Track

A pitch track is automatically computed and plotted when the *Pitch* button  is clicked, or the selection *Pitch Track Plot* is made from the *Tools* menu. The color of the button, which is red, serves as a mnemonic for the pitch track, which is plotted in red on the spectrogram. If the pitch track has been plotted, it will also appear on printed spectrograms in red, or in gray on a black-and-white printer.

If it is necessary to see or print the pitch track plotted without the spectrogram, that is, against a plain white background, this can easily be done. First, from the *View* menu, select *Clear Spectrogram*. Then click the *Pitch* button. The pitch track will be drawn in red on a clear background, and will be printed on the same background. Use of the *Redraw Spectrogram* function  will bring back the spectrogram, but without the pitch track.

In *SpeechStation2*, the pitch track is computed by an algorithm that uses center clipping and autocorrelation.¹ This method rapidly produces accurate pitch tracks from speech recordings, and works well with many poor or noisy recordings.

Adjusting Energy Threshold

During noisy or silent parts of the signal, an accurate pitch estimate may be difficult to make. To avoid reacting to the noise and filling the spectrogram plot with unreliable data, pitch estimates are only made

when the energy in the signal exceeds a threshold. This is reasonable because voiced segments of the speech signal are usually those with the highest energy.

By default, the threshold is set somewhat arbitrarily at the mean energy (computed over the displayed segment), a setting that works well in most cases when the speech recording is free of excessive distortion and noise. By selecting *Custom Energy Threshold* from the *Pitch Track Setup* in the *Tools* menu, it is possible to drag the threshold up and down on the energy plot. When a recording is low in noise, adjustment of the threshold may allow an increase in pitch track coverage. If the recording is a poor one, or there is strong background noise present, the threshold can be raised to avoid artifacts.

If there is substantial noise in the recording and adjustment of the energy threshold fails to produce a satisfactory pitch track, there is an alternative. The waveform can be examined in the *Spectrum Viewer* and the pitch in that region estimated from the time interval between successive glottal pulses, provided that these can be seen clearly. The method consists of sliding the vertical red cursor first to one clearly recognizable feature of one cycle, then to the same feature in the next cycle. The pitch estimate is then the reciprocal of the time difference in seconds. It is easiest to do this if the *Spectrum Viewer* has been maximized.

¹Hess, Wolfgang (1983), *Pitch Determination of Speech Signals: Algorithms and Devices*. (New York & Heidelberg: Springer-Verlag). p. 359.

Resolution of the Pitch Estimate

The pitch detection algorithm used in *SpeechStation2* estimates pitch values at discrete intervals. The relation between these intervals and the sampling rate can be seen in the following example.

At a sample rate of 11025 Hz, individual samples are 90.7 microseconds apart. If the autocorrelation maximum is at an interval of exactly 120 samples, for example, the estimated pitch will be $11025/120 = 91.875$ Hz. If the maximum autocorrelation is found at a shift of 119 samples, that is, one sample less, the estimated pitch will be 92.65 Hz; if 121 samples, or one sample higher, the pitch estimate will be 91.12 Hz. In this case, the resolution of the pitch detection method is 0.76 Hz in either direction, down or up.

In general, the percentage frequency resolution of a pitch estimate of f_0 at a sample rate of f_s is approximately $(f_0/f_s) \times 100$ percent.

Setting the Pitch Multiplier

When the pitch track is plotted on the spectrogram, it is plotted at 10 times the actual frequency to allow pitch values and variations to be seen clearly. If the pitch is not too high, this multiplier can be increased to make details of the pitch track clearer. On the other hand, if the pitch is very high, the multiplier should be reduced to keep the pitch track from going above the top of the displayed frequency scale, because *SpeechStation2*'s pitch-detection algorithm rejects pitch estimates that would fall off-scale.

In addition to the graphic display, the pitch can always be read from the status bar under the waveform plot by moving the mouse pointer to the desired point on the spectrogram.

Exporting Pitch Track Data

To process pitch data with another program, it is useful to convert the data to a simple text file listing time and pitch. The menu item *Export Pitch Track* in the *Tools* menu does this, creating an ASCII file listing the data.

The start of the file has the appearance shown below.

```
8/3/99 5:45:14 PM
SpeechStation2 v.1.1.0
Pitch Track
F:\EXAMPLES\Talker01.wav
From:0 sec
To:2 sec
Sample Rate:11025 Hz
Samples per index:17.334
Seconds,Frequency
*****
0,0
.0015,0
```


The first entry on each data line (following the introductory information) gives the time index and the second entry the pitch estimate.

Formant Tracking

SpeechStation 2 can compute formant tracks using linear predictive coding (LPC). (See the section on *LPC mode* in *Spectrum Analysis Options* for a description of linear predictive coding.) Formant estimates are obtained by peak-picking on the LPC spectrum and the tracks formed from these estimates are plotted in the spectrogram area.

To get appropriate formant estimates, the order of the LPC model must be chosen correctly. A guide to determining the order is given in the *LPC Mode* section of *Spectrum Analysis Options*.

Displaying Formant Tracks

Formant tracks are automatically computed using LPC when the  button is clicked, or when *Formant Tracks* is selected from the *Tools* menu. The color of this button is green to serve as a mnemonic for the formant tracks which are plotted in green on the spectrogram. Like the pitch track, the formant tracks can be plotted in a cleared area, instead of on the spectrogram, by giving the *Clear Spectrogram* command in the *View* menu before drawing the tracks.

Exporting Formant Tracks

The data in the Formant Tracks can be exported to a text file by selecting *Export Formant Tracks* in the *Tools* menu.

The start of the file has the appearance shown below.

```
8/3/99 5:44:59 PM
SpeechStation2 v.1.1.0
Formant Tracks
F:\EXAMPLES\Talker01.wav
From:0 sec
To:2 sec
Sample Rate:11025 Hz
Samples per index:17.334
Seconds,Frequency1,Frequency2,...
*****
0.0000,1329.86,2187.84,3431.90
0.0015,1308.41,3431.90
```

The first entry in a data line (following the introductory information) is the time in seconds, and subsequent entries are the estimated formant frequencies.

Labels

SpeechStation2 allows labels to be attached to files. The labels may be placed anywhere on the spectrogram; if the file is then saved, the labels will appear at the same locations whenever the file is loaded. When a file is saved with labels attached, a label file is created with the same name as the data file but with the extension .LAB.

Creating a Label

To create a label, hold down the CTRL key and click where the label is to be placed. The position does not need to be very accurate, as the label can easily be moved after it has been created. Next, type the text of the label, which can consist of any ASCII characters available in the *Arial* font (a *True-Type* font supplied with Windows).^{*} When the text is complete, press the *Enter* key. If *Enter* is not pressed, the label is not complete and cannot be moved.

Accurate positioning for printing

When a spectrogram is printed, only the time placement of the first character is preserved. Therefore, spaces and tabs should not be used to position text in a label. Do not try to combine labels that are to appear at different places on the time scale. Click, type and enter each one.

Moving the label


To move a label after it has been placed and entered, position the mouse cursor on it and hold down the mouse button, and then drag the label to its new position.

Removing a label

To remove a label, hold down the *Ctrl* key and click the mouse on the text.

^{*}To see the characters available, click *Start* in the lower left corner of the screen, then *Programs*, *Accessories* and *Character Map*, and select the *Arial* font.

Printing

SpeechStation2 will print its displays on any Windows-compatible printer. When the Print button  on the main screen is clicked, or Print is selected from the File menu, a dialog box appears (like that shown at right) to direct the printing of main screen display windows.

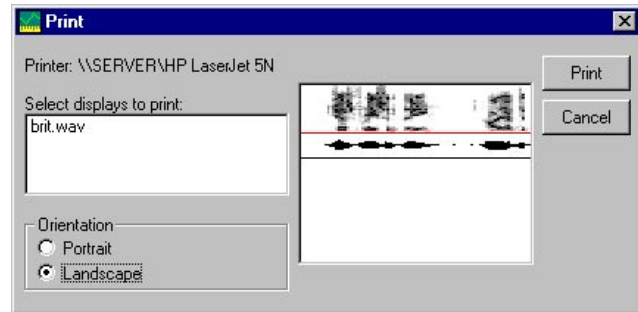
In the left panel of the dialog box is a list of the files currently loaded. The display window for any or all of these can be printed. The user selects each display to be printed by clicking the left mouse button on the corresponding file name. Below the names of the files are option boxes for the choice of portrait or landscape orientation.

In the window at the right side of the dialog box is a small representation of the spectrogram and waveform as these will appear on the page. If only one display window is to be printed, the size of the image, as well as the proportions of the spectrogram and waveform parts can be adjusted. The adjustment of spectrogram/waveform proportion is made by clicking and dragging the red line dividing the spectrogram and waveform in this small image. The same procedure can be followed with the black line at the bottom of the image to control the overall size of the image.

If multiple displays are selected for printing, *SpeechStation2* automatically divides the page appropriately for the number selected. Relative spectrogram/waveform proportion can still be adjusted with the red line, and this proportion is applied to all the displays printed. Dragging the line at the bottom of the image will adjust the overall size of each spectrogram within its portion of the page. A maximum

of four displays will be fitted on a page in portrait mode, three in landscape.

To provide spectrograms of the highest possible quality, those printed by *SpeechStation2* are not bit maps transferred from the screen display, but are constructed expressly for printing. This allows the best use of the gray scale and typographic capabilities of each printer. In particular, type used for time and frequency scale labels is rendered accurately and consistently during printing, without the jagged edges often produced when fonts are mapped to the screen and printed.



The print dialog box, showing the name of the one file currently loaded. A representation of the layout for printing is shown in the right panel.

Stereo Files

SpeechStation2 provides the ability to record, analyze and edit stereo files. Although many applications call for analysis only of single-channel (mono) recordings or files, there are some instances in which the analysis of two-channel, or stereo, signals is desired.

When you open a stereo .WAV file, or make a stereo recording within *SpeechStation2*, a dialog box appears that asks if you want to open the left channel, the right channel, or both in separate windows. If you select “Left”, for example, a display window would appear showing the spectrogram of the first two seconds of the signal in the left channel, with an ‘L’ appended to the filename in the title bar. If you select “Both” channels to be opened, then two display windows appear, one showing the left channel signal and the other showing the right.

For purposes of analysis and display, the two display windows for the left and right channel signals can be treated completely independently. They can have different analysis modes, windows, gray or color scales, etc.

For purposes of manipulation, however, the two channels of a stereo file are linked. **Whenever an operation calls for a change to be made to one channel of a stereo file, that operation is applied automatically to both channels - even if only one channel is opened for display.** The commands that change signals are those in the Edit menu as well as the application of a filter (under the Tools menu). If any of the commands under the Edit menu (Undo, Cut, Copy, Paste, Delete, and Clear) is given for a selection in one channel of a stereo signal, that command is executed equally on both channels.

Although it is possible to have different segments selected in the displays for the two channels, the selection that applies to any Edit or Filter operation is the one in the currently-active display window (the file name of which is shown in the title bar of the control window). Further, even if only one channel is displayed, any changes to it are also made to the non-displayed channel and saved if the changes are accepted when the file is closed.

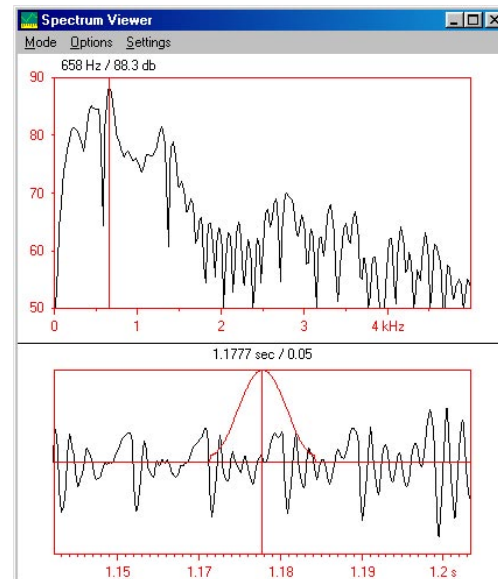
Spectrum Viewer

If you click anywhere on the spectrogram, two red lines appear, one horizontal and one vertical, intersecting at the place that was clicked. A new window will appear, the *Spectrum Viewer*. The upper half of the *Spectrum Viewer* shows the single spectrum that corresponds to the time location selected by the click; a vertical red line marks the frequency that has been selected on the spectrogram. If the spectrum is too low or too high in its frame to show details you need to see, you can use the up and down arrow keys on the keyboard to move the spectrum up or down. The display also shows the frequency and level (in dB) of the selected point above the spectrum. In the bottom half of the viewer, a waveform display shows 512 samples along with the analysis window centered on the selected time.

The frequency and time cursors (red lines) can be moved by dragging them to the left or right. When the cursors are moved, their new values are displayed. After moving the time cursor, selecting *Redraw* from the *Options* menu causes the *Spectrum Viewer* to be redrawn with a new spectrum and waveform centered on the new time location. The cursor can be used to measure the time interval between two consecutive peaks in the waveform to calculate a precise local pitch estimate ($1 / \text{interval} = \text{frequency}$).

The decibel value reported in *Spectrum Viewer* is $20 \log_{10} x$ where x is the magnitude of the Fourier coefficient for that time slice and analysis frequency computed from the windowed signal. *SpeechStation2* normalizes its window functions so that changing window type or length leaves the *total* energy of the spectrum display unaltered; changing from a window whose frequency response has a narrow

center lobe to one with a wider center lobe (e.g., from a rectangular window to a Hamming window of the same length, or from a longer to a shorter window of a given shape) will tend to lower and broaden peaks in the spectrum.



Spectrum Viewer: Mode Menu

The *Mode* menu allows the user to select the type of display shown in *Spectrum Viewer*, which can be one of three kinds.

Single Spectrum

This display is the default (see previous page). The upper part of the viewer window is a single spectrum, and the lower part shows the waveform from which the spectrum has been derived. Also shown is the analysis window in effect.

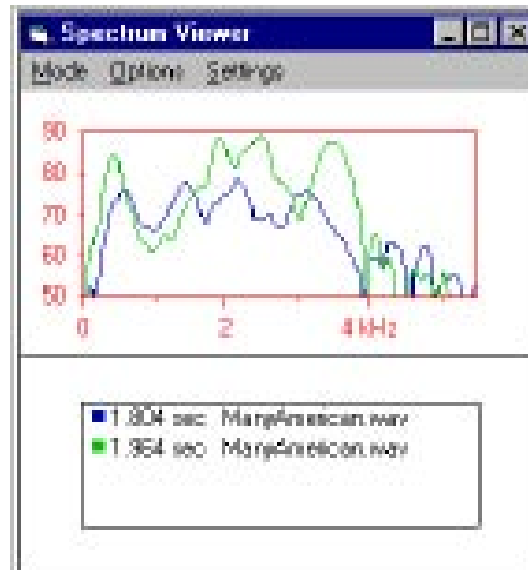
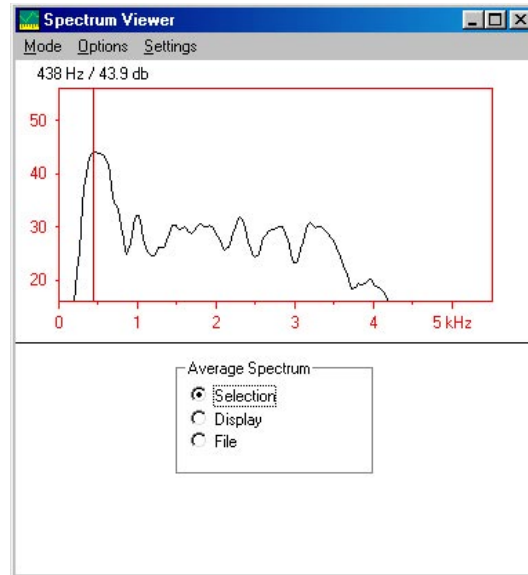
Average Spectrum

This selection produces a display like that shown in the upper figure at right. If a segment of the waveform has been selected, the spectra in this segment will be averaged to produce the spectrum in the upper display. If either of the other possible segments (*Display* or *File*) are selected, the average spectrum for the spectrogram segment or the entire file will be calculated. RMS averaging is performed in all cases.

Multiple Spectra

If this selection is made, up to six individual spectra can be displayed in the spectrum plot area by clicking on the appropriate time location in the spectrogram for each spectrum desired. Spectra can be chosen from different spectrograms in a file, or from different files loaded at the same time. However, all spectra to be compared must come from files having the same sample rate. An example of a Multiple Spectra display is shown in the lower right figure.

Each spectrum is shown in a different color. The colors are changed automatically every time a different place on the spectrogram is clicked. The selected time and the file name are printed alongside a small color square that identifies the spectrum.



Spectrum Viewer: Options Menu

Clicking on the file name in the information box will display the spectrogram from which the spectrum was taken with the selection cursor visible in its appropriate color. This cursor can be moved within its file by clicking and dragging it. A cursor can be erased from a spectrogram, and its corresponding spectrum removed from the *Spectrum Viewer*, by holding down the *Shift* key while clicking on the cursor.

Animate Spectrum

With this command, you can view the spectrum change as a function of time. This option sweeps through the time segment displayed in the lower window of the *Spectrum Viewer*. A red line cursor and time display show the time at the center of the window used to calculate the current spectrum.

Animate Selection

Like the above, but a selection made in the waveform of the display window is animated.

Print

Ctrl-P

Prints the entire *Spectrum Viewer* display.

Redraw

The time cursor, a vertical red line, can be dragged from its central position to a new position using the mouse. The *Redraw* operation puts the new time at the center of the plot and draws a new spectrum and waveform for the selected time.

Windowed Waveform

Controls the display of the waveform in *Single Spectrum* mode. If this option is selected, the waveform is multiplied by the window before display, showing the data actually used to compute the FFT. If this option is not checked, the unaltered waveform is drawn and the window superimposed.

Spectrum Viewer: Settings

The *Settings* menu for *Spectrum Viewer* includes some of the *Settings* functions of the main control window, and offers additional choices. Note that the settings made in *Spectrum Viewer* are independent of those made in the Main Screen.

Analysis Mode: FFT

The spectrum shown in the upper half of the viewer is computed as an FFT of the signal samples in the analysis window. This mode and the LPC mode are explained in detail in *Spectrogram Analysis Options*.

Analysis Mode: LPC

The spectrum shown in the upper half of the viewer is computed from an LPC analysis of the signal in the analysis window.

Analysis Mode: Both

If this selection is made, both LPC and FFT spectra are displayed simultaneously, in different colors.

Window Size, Type, LPC Order, Pre-emphasis

These have the same functions as the corresponding spectrogram commands in the Main Screen. For more detail, see *Spectrogram Analysis Options*.

Note that in specifying the LPC order, a small dialogue box appears. After entering the order number, type the 'Enter' key for the entry to take effect. Clicking on the 'x' in the corner of the dialogue box will close the box without making a change to the LPC order.

Y-Axis

This command allows the maximum level and range of the vertical scale of the spectrum to be set. It can be thought of as a numerical version of the command *Gray Scale*. The maximum level of the y-axis can also be changed with the up/down arrow keys.


Average Interval

When an average spectrum is calculated, the spectra that are averaged are taken from data locations separated by a fixed time interval. The length of the interval, which has a default value of 64 samples, can be specified by the user.

With this parameter, the user can reach a compromise between over-sampling, which doesn't miss any events but takes a long time to compute, and under-sampling, which computes rapidly but may have missed some events in the averaging process. An average interval that is about one-fourth to one-half the size of the analysis window will be sure to include all brief events in the average.

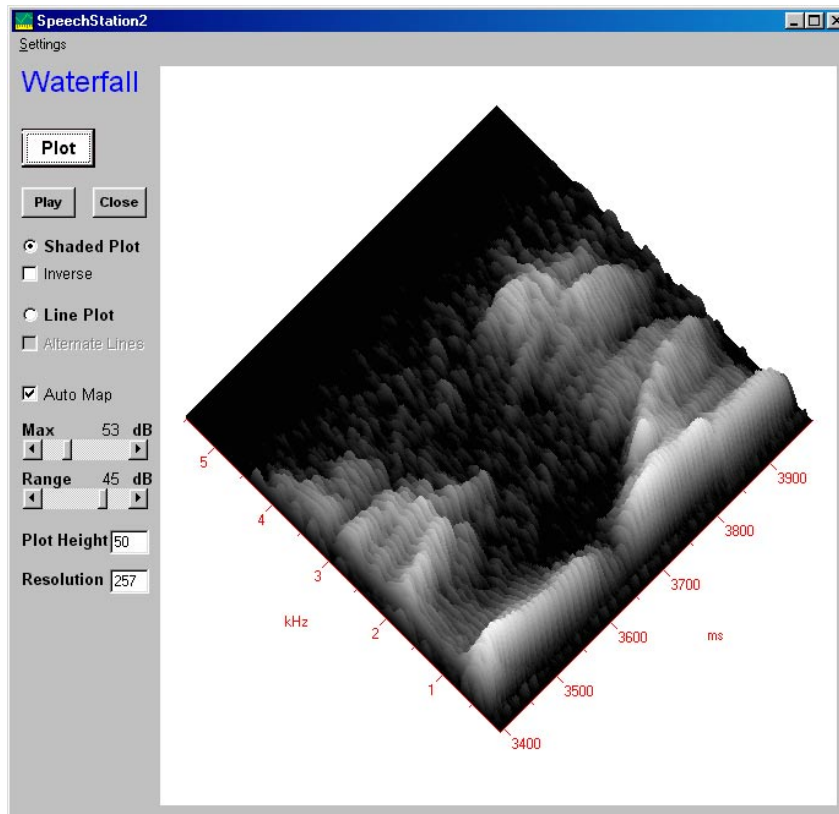
In the dialogue box for Average Interval, type the 'Enter' key to have the new entry take effect. Click 'x' to cancel.

Waterfall Plot

Clicking on the *Waterfall Plot* button  after a selection has been made produces the *Waterfall Plot* display window shown below. If no selection has been made, the button is inoperative. A selection must be made before a waterfall plot can be made.

Settings Menu

The *Waterfall Plot* display window has its own *Settings* menu, separate from those of the Main Screen and *Spectrum Viewer*. The menu allows analysis mode, window size and type, LPC order, pre-emphasis filter and gray scale or color display to be selected. These commands have the same functionality here as they do in the Main Screen (see *Main Screen Commands*). The Print command is also listed in this menu (see *Printing Waterfall Plots*, page 41).



On-Screen Controls

Plot

After controls have been adjusted as desired, click this button to display the waterfall plot.

Play

The *Play* button can be used to play the part of the file selected to make the waterfall plot.

Close

The *Close* button closes the *Waterfall Plot* window and returns to the Main Screen.

Plot type selection

The four kinds of waterfall plots available are:

1. shaded
2. inverse shaded
3. line
4. alternate-line

The next page illustrates each of these plot types using the same segment of a speech signal. The choice of plot type generally depends on whether or not the plot is to be reprinted, whether good half-tone reproduction will be available in a publication, and the final size of a published or printed plot. If the plot is to be duplicated using low-cost copying methods, the alternate line plot (4) may be the best to use.

Auto-Map, Max and Range adjustments

These controls operate like those in the Main Screen, setting the mapping from dB level to gray scale (see *Spectrogram Display Options*).

Plot height

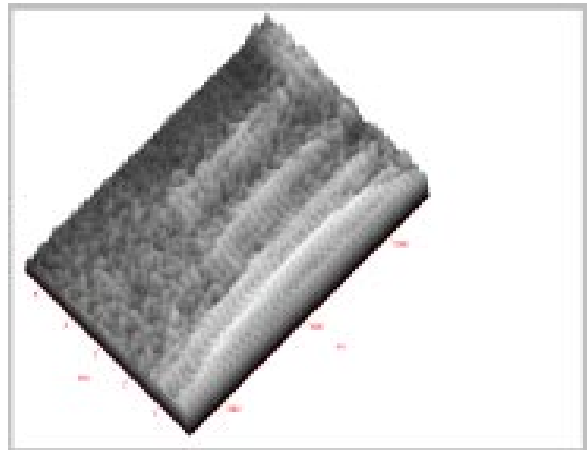
The apparent height of the plot is given a nominal default value of 50. This can be adjusted by the user to make the terrain appear taller or flatter.

Resolution

Normally, the frequency axis of the waterfall plot is made of 256 plotted points. The plot can be made narrower by changing this value.

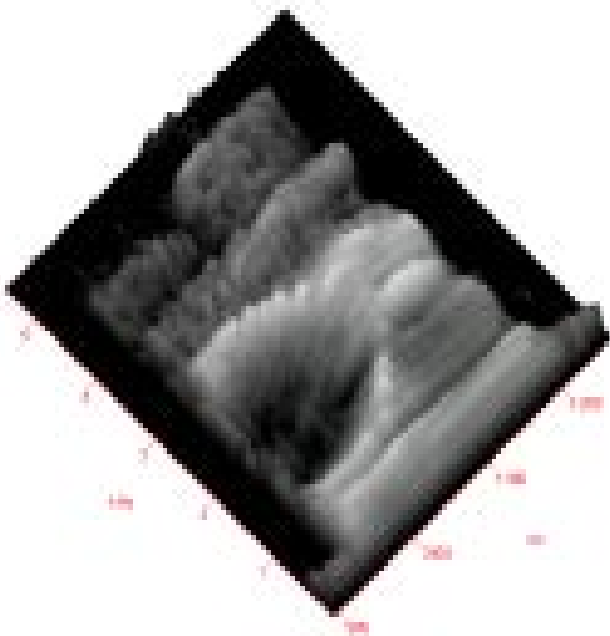
Enlarged window

The plot is automatically enlarged when the window is maximized. This is done by clicking on the *Windows maximize* button at the upper right corner of the window. In most cases, this will also change the aspect ratio of the plot.

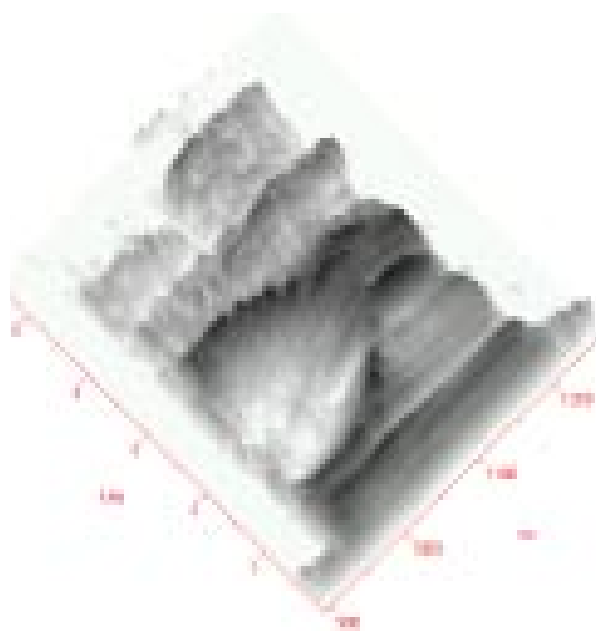


Aspect ratio of the waterfall plot display window when the window has been maximized. The space at the right of the plot is set by the clearance needed at the top of the plot, which the program automatically maintains.

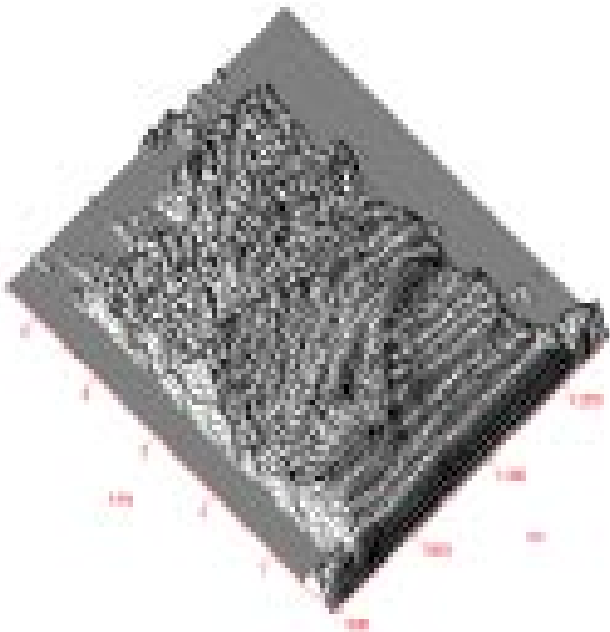
1



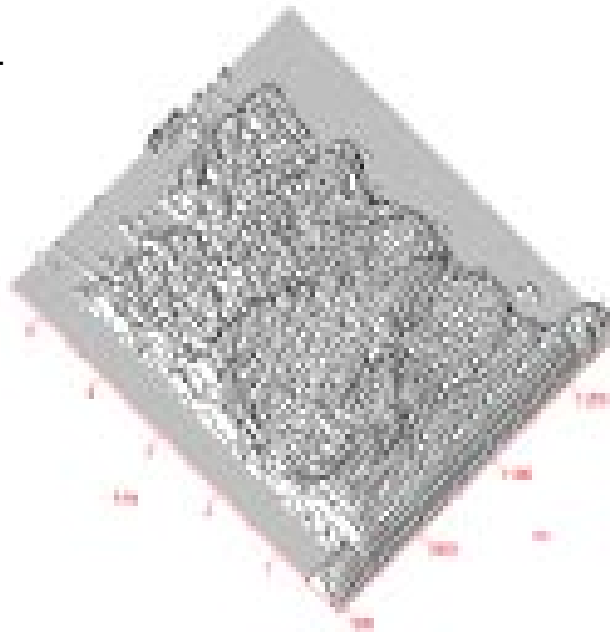
2



3



4



Printing waterfall plots

To print a permanent copy of a shaded plot (1 or 2) you need a printer that can print gray scale images. Most ink-jet and many laser printers will yield good gray scale reproduction, producing their own form of halftone. Color ink-jet printers that print photographs well generally produce excellent shaded waterfall plots.

An ink-jet printer uses much more ink when large dark areas are printed. This is one reason that the inverse shaded mode (2) is provided. In this mode, the same information is shown as in the shaded mode (1) but the plot has a light background and can be printed using much less ink. This avoids the smearing and wrinkling that can occur when the paper is soaked with ink. While laser printers do not wet the paper, more toner is used to print plots made in the shaded mode.

Real-Time Spectrogram

The Real-Time Spectrogram presents a continuously-moving spectrogram of an analog input signal. Before starting the *Real-Time Spectrogram*, check the color setting of the computer's monitor screen. This is very important, and is explained in the box on the following page.

To bring up the window shown in the illustration below, click on *Tools* in the menu bar, then select *Real-Time Spectrogram*. Play the tape recording, or introduce a signal from another source, and listen to be sure that sound is coming from the loudspeakers or headphones connected to the computer's audio output. If everything is connected correctly, sound will be heard.

If sound is heard at normal volume, click the *Go* button and the spectrogram will begin scrolling from right to left across the screen. Next, try speeds 2 and 4. If you have set your computer monitor as described earlier, and your computer is fast enough, you can set the sampling rate to 22050 Hz. Many users find that 2 is a useful speed, fast enough to see important details, but slow enough to allow easy observation.

The *Max* and *Range* settings work in the same way as those in the Main Screen for controlling the mapping of dB levels to gray scale.

Recording

To record while observing the spectrogram of the incoming signal, stop the moving spectrogram and click on the check box *Save*. Then give a name to the file that is to be recorded.

Settings

The settings used in *Real-Time Spectrogram* are the ones in effect for the Main Screen, except for the *FFT Size* selection. Normally, the moving spectrogram display is 257 pixels high, the number of frequency points resulting from an FFT operation on 512 consecutive data samples. Switching to a 256-point FFT reduces the height of the displayed spectrogram by a factor of two, but also speeds the plotting on the screen.



Note

Desktop computers with CPU clock speeds greater than 133 MHz should be able to operate the *Real-Time Spectrogram* with all of its features. However, this may require setting the monitor display to 16-bit color or less (rather than 24-bit). This adjustment is a standard Windows operation, accomplished as follows:

1. Close all programs.
2. Click the right mouse button on the background of the desktop.
3. Select "Properties".
4. Click the tab entitled, "Settings".
5. On the left side of the panel, set the color palette to *High Color (16 bit)*.
6. Click OK.

Note that computers with CPU clock rates below 133 MHz may not be able to use this feature of SpeechStation2.

Vowel Space Plot

The *Vowel Space Plot* window is accessed by selecting *Vowel Space Plot* from the *Tools* menu. The plot is a display of the two lowest formant frequencies (F1 and F2). Such plots are of interest because articulatory behavior results in repetitive, characteristic relations between formants.

Setting Parameters

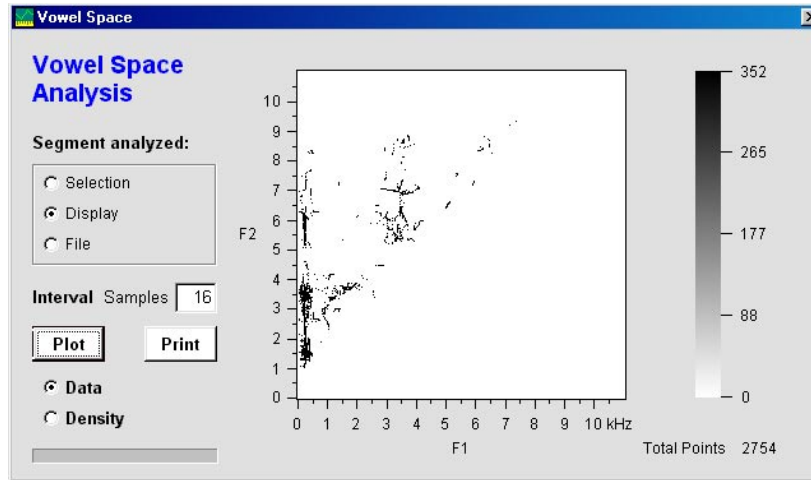
The plot is created by scanning the data segment (either the selection, the display, or the entire file) using LPC analysis to find the center frequencies of the two lowest formants. The analysis parameters used are those in effect in the Main Screen. From the window, the user can select the interval at which measurements are made and points plotted; 16 samples is the default. Reducing the interval increases the number of calculations made for an equivalent signal duration, and increases computation time proportionally. A progress bar at the bottom of the panel shows the progress of the plot toward completion.

Data and Density plots

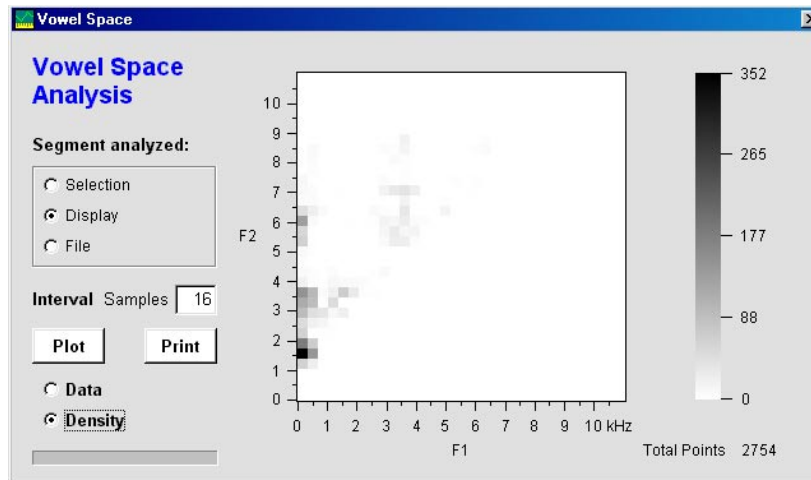
Two plotting formats are provided; these are illustrated on the following page. In the *Data* format, every point is plotted. In the *Density* format, the field is divided into a 32x32 grid of squares. The gray density of each square represents the number of data points detected within its boundaries. The format can be selected before, during, or after plotting.

To the right of the F1/F2 plot is a key relating the gray-scale values in the density plot to the number of data points in the F1/F2-cell, and a tally of the total number of data points plotted. Clicking the left mouse button while the mouse is over the F1/F2 plot will bring up horizontal and vertical cursor lines inter-

secting at the clicked spot. Clicking the right mouse button while the mouse is over the F1/F2 plot will cause the F1/F2-cell at the mouse location, all other cells with the same number of data points, and the corresponding gray level in the key to be highlighted. Also, the number of points in each of these cells will be displayed next to the key. Clicking either mouse button while the mouse is over the key will highlight the gray level at the mouse location, display the corresponding number of data points, and highlight all cells containing this number of points. (To remove the highlighting, hold the shift key down while clicking on the plot or the key.)



The Vowel Space Plot window, showing the data plot format.



The density plot format of a Vowel Space Plot.

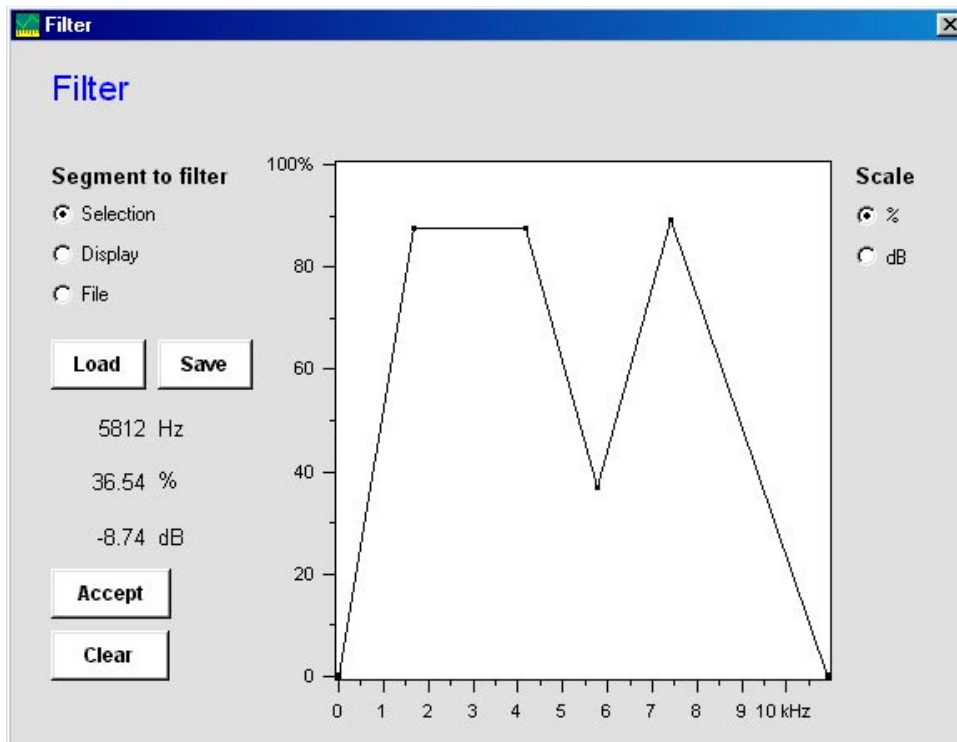
Filter

When *Filter* is selected from the *Tools* menu, a window appears like the one shown below. The first choice to be made is the segment to be filtered, which can be (1) a selection from the waveform, which must already have been made, (2) the segment displayed on the screen, or (3) the entire file.

If you are not sure of the filter's shape or want to try it to see what happens, you may want to rename a copy of the file you are working with and use the copy for your tests.

Adjusting the filter characteristics

The filter's magnitude response is adjusted by placing the cursor at the points that define the filter, one at a time, and clicking once with the left mouse button. The cursor's coordinates are shown at the left. Each time the cursor is placed and the mouse clicked, another point in the filter response will be defined. The filter can be saved and used again whenever it is needed. The vertical scale of the filter characteristic can be switched between percent and decibels of attenuation by using the controls at the right of the display.



Correcting mistakes

The *Clear* button completely removes the filter. If only one point needs to be changed, the incorrect point can be dragged to the right place, or removed by clicking on it while holding down the *Shift* key.

Accepting the filter

Clicking on the *Accept* button will implement the filter; that is, the filter will be applied to the designated segment.

Saving and reloading a filter

Click on *Save* if you want to save the filter, then give it a name. It will be saved with the extension *.FLT*. A filter that has been saved can be recalled with the *Load* button.

Appendix A: Recorded Phonetic Library

The *SpeechStation2* CD-ROM includes 125 sound files forming a recorded phonetic library containing most of the phonemes heard in the world's spoken languages. The recordings were made by Dr. Peter Ladefoged to accompany the Sensimetrics course on CD-ROM, *Speech Production and Perception I*. The files are listed here and on the pages that follow.

All of the file names start with the letters "PL," followed by the IPA number assigned to the phonetic symbol shown in the list.* To hear one of the recordings and examine its spectrogram, load the file from the *Phonetic* directory on the *SpeechStation2* CD-ROM.

*Esling, John H. and Harry Gaylord (1993) "Computer codes for phonetic symbols," *Journal of the International Phonetic Association* (1993) **23**: 2.

Consonants

Filename	Phonetic Symbol	Voicing	Place	Manner
PL101.wav	p	voiceless	bilabial	plosive
PL1011.wav	p'	voiceless	bilabial	ejective
PL1014.wav	p ^h	voiceless/aspirated	bilabial	plosive
PL102.wav	b	voiced	bilabial	plosive
PL103.wav	t	voiceless	alveolar	plosive
PL1031.wav	t'	voiceless	alveolar	ejective
PL1034.wav	t ^h	voiceless/aspirated	alveolar	plosive
PL104.wav	d	voiced	alveolar	plosive
PL105.wav	ʈ	voiceless	retroflex	plosive
PL1051.wav	ʈ'	voiceless	retroflex	ejective
PL1054.wav	ʈ ^h	voiceless/aspirated	retroflex	plosive
PL106.wav	ɖ	voiced	retroflex	plosive
PL107.wav	c	voiceless	palatal	plosive
PL1071.wav	c'	voiceless	palatal	ejective
PL1074.wav	c ^h	voiceless/aspirated	palatal	plosive
PL108.wav	ɟ	voiced	palatal	plosive
PL109.wav	k	voiceless	velar	plosive
PL1091.wav	k'	voiceless	velar	ejective
PL1094.wav	k ^h	voiceless/aspirated	velar	plosive
PL110.wav	g	voiced	velar	plosive
PL111.wav	q	voiceless	uvular	plosive

Filename	Phonetic Symbol	Voicing	Place	Manner
PL1111.wav	q'	voiceless	uvular	ejective
PL1114.wav	q ^h	voiceless/aspirated	uvular	plosive
PL112.wav	ɕ	voiced	uvular	plosive
PL113.wav	ʔ	voiceless	glottal	plosive
PL114.wav	m	voiced	bilabial	nasal
PL115.wav	ɱ	voiced	labiodental	nasal
PL116.wav	n	voiced	alveolar	nasal
PL117.wav	ɳ	voiced	retroflex	nasal
PL118.wav	ɲ	voiced	palatal	nasal
PL119.wav	ŋ	voiced	velar	nasal
PL120.wav	ɴ	voiced	uvular	nasal
PL121.wav	ʙ	voiced	bilabial	trill
PL122.wav	r	voiced	alveolar	trill
PL123.wav	ʀ	voiced	uvular	trill
PL124.wav	ɾ	voiced	alveolar	tap/flap
PL125.wav	ɽ	voiced	retroflex	tap/flap
PL126.wav	ɸ	voiceless	bilabial	fricative
PL127.wav	β	voiced	bilabial	fricative
PL128.wav	f	voiceless	labiodental	fricative
PL129.wav	v	voiced	labiodental	fricative
PL130.wav	θ	voiceless	dental	fricative
PL131.wav	ð	voiced	dental	fricative
PL132.wav	s	voiceless	alveolar	fricative
PL133.wav	z	voiced	alveolar	fricative
PL134.wav	ʃ	voiceless	postalveolar	fricative
PL135.wav	ʒ	voiced	postalveolar	fricative
PL136.wav	ʂ	voiceless	retroflex	fricative
PL137.wav	ʐ	voiced	retroflex	fricative
PL138.wav	ç	voiceless	palatal	fricative
PL139.wav	ʝ	voiced	palatal	fricative
PL140.wav	x	voiceless	velar	fricative

Filename	Phonetic Symbol	Voicing	Place	Manner
PL141.wav	ɣ	voiced	velar	fricative
PL142.wav	χ	voiceless	uvular	fricative
PL143.wav	ʁ	voiced	uvular	fricative
PL144.wav	ħ	voiceless	pharyngeal	fricative
PL145.wav	ʕ	voiced	pharyngeal	fricative
PL146.wav	h	voiceless	glottal	fricative
PL147.wav	ɦ	voiced	glottal	fricative
PL148.wav	ɸ	voiceless	alveolar	lateral fricative
PL149.wav	β	voiced	alveolar	lateral fricative
PL150.wav	ʋ	voiced	labiodental	approximant
PL151.wav	ɹ	voiced	alveolar	approximant
PL152.wav	ɻ	voiced	retroflex	approximant
PL153.wav	j	voiced	palatal	approximant
PL154.wav	ɰ	voiced	velar	approximant
PL155.wav	l	voiced	alveolar	lateral approximant
PL156.wav	ɭ	voiced	retroflex	lateral approximant
PL157.wav	ʎ	voiced	palatal	lateral approximant
PL158.wav	ɮ	voiced	velar	lateral approximant
PL159.wav	ɓ	voiceless	bilabial	implosive
PL160.wav	ɣ	voiced	bilabial	implosive
PL161.wav	ɸ	voiceless	alveolar	implosive
PL162.wav	ɓ	voiced	alveolar	implosive
PL163.wav	ɸ	voiceless	palatal	implosive
PL164.wav	ɓ	voiced	palatal	implosive
PL165.wav	k	voiceless	velar	implosive
PL166.wav	g	voiced	velar	implosive
PL167.wav	q	voiceless	uvular	implosive
PL168.wav	ɢ	voiced	uvular	implosive
PL169.wav	ɱ	voiceless	labial-velar	fricative
PL170.wav	w	voiced	labial-velar	approximant
PL171.wav	ɥ	voiced	labial-palatal	approximant

Filename	Phonetic Symbol	Voicing	Place	Manner
PL172.wav	h	voiceless	epiglottal	fricative
PL173.wav	ʔ		epiglottal	plosive
PL174.wav	ɸ	voiced	epiglottal	fricative
PL175.wav	ɸ̥	voiceless	postalveolar-velar	fricative
PL176.wav	ʘ		bilabial	click
PL177.wav			dental	click
PL178.wav	!		postalveolar	click
PL179.wav	ɰ		palatoalveolar	click
PL180.wav	l̥		alveolar	lateral/click
PL181.wav	ɺ	voiced	alveolar	lateral/flap
PLbj.wav	bj	voiced	bilabial-palatal	plosive
PLgb.wav	gb	voiced	labial-velar	plosive
PLkp.wav	kp	voiceless	labial-velar	plosive
PLpc.wav	pc	voiceless	bilabial-palatal	plosive

Vowels:

Filename	Phonetic Symbol	Height	Backness	Manner	Roundness
PL301.wav	i	close	front	tense	unrounded
PL302.wav	e	close-mid	front	(tense)	unrounded
PL303.wav	ɛ	open-mid	front	(lax)	unrounded
PL304.wav	a	open	front		unrounded
PL305.wav	ɑ	open	back		unrounded
PL306.wav	ɔ	open-mid	back	(lax)	rounded
PL307.wav	o	close-mid	back	(tense)	rounded
PL308.wav	u	close	back	tense	rounded
PL309.wav	y	close	front	tense	rounded
PL310.wav	ø	close-mid	front	(tense)	rounded
PL311.wav	œ	open-mid	front	(lax)	rounded
PL312.wav	ɶ	open	front		rounded
PL313.wav	ɒ	open	back		rounded
PL314.wav	ʌ	open-mid	back	(lax)	unrounded
PL315.wav	ɤ	close-mid	back	(tense)	unrounded
PL316.wav	ʊ	close	back	tense	unrounded
PL317.wav	ɨ	close	central		unrounded
PL318.wav	ɤ̞	close	central		rounded
PL319.wav	ɪ	close	front	lax	unrounded
PL320.wav	ɪ̞	close	front	lax	rounded
PL321.wav	ʊ̞	close	back	lax	rounded
PL322.wav	ə	mid	central		unrounded
PL323.wav	ɵ	mid	central		rounded
PL324.wav	ɐ	open	central		unrounded
PL325.wav	æ	open/open-mid	front	(tense)	unrounded
PL326.wav	ɜ	open-mid	central		unrounded
PL396.wav	ɘ	open-mid	central		rounded
PL397.wav	ɚ	close-mid	central		unrounded

Appendix B: Making Better Speech Recordings

Selecting a microphone

The factors that are likely to have the greatest effect on the quality of a recording are the microphone and its placement. Although it is not necessary to buy an expensive microphone to get good results, some care in making the choice can ensure that one will obtain good recordings suitable for analysis.

Types of microphones

Two types of microphones are available in electronic supply stores. One is the *dynamic* type, and the other the *condenser* type. Both can produce high-quality recordings, although a dynamic microphone of good quality generally costs more than a suitable condenser microphone.

Dynamic microphones

The dynamic microphone has a membrane attached to a small coil, suspended in the field of a magnet. This type of microphone works like a small loudspeaker in reverse. Sound waves striking the membrane make the coil vibrate in the magnet's field. This movement generates weak electric currents in the coil which can be amplified and used to make a recording. Good dynamic microphones generally cost between \$50 and \$100; the best cost much more.

A common type of microphone supplied with many computer audio cards is an inexpensive dynamic microphone. Unfortunately, such microphones tend to have relatively low output, so that they must usually be placed very close to the talker's mouth, and can pick up hum from external electric fields. Better dynamic microphones will not present this problem, but need a desk stand because of their substantial shielding and heavier magnets.

Condenser microphones

A condenser microphone typically consists of an extremely thin plastic membrane carrying a permanent electric charge, stretched tightly and very close to a flat metal plate. The membrane and plate form the two plates of a capacitor, and are connected to a battery. Sound waves striking the membrane make it vibrate, passing small electric current through the capacitor. Condenser microphones usually work best if they are oriented so that the microphone diaphragm is parallel to the acoustic path, that is, the microphone faces upward rather than directly at the mouth of the talker. This can reduce the tendency of such microphones to emphasize high frequencies, which produces exaggerated hissing or spitting sounds when frication is recorded.

Almost all professional recording is now done with condenser microphones that may cost thousands of dollars. Condenser microphones are also used in laboratories to make precise acoustic measurements. On the other hand, the principle of the condenser microphone is simple enough so that tens of millions are made each year for use with consumer tape recorders. A useful, good quality condenser microphone can be bought for \$20 to \$50.

Noise-cancelling microphones

One kind of microphone specially designed for speech is called "noise-cancelling," and is supplied with some speech recognition software packages. Noise-cancelling microphones are attached to a headband, and have a flexible mounting that holds the microphone in a fixed position relative to the talker's mouth. Room noise and reverberation have little effect on the speech signal when such a microphone is used because the microphone is so close

to the talker's mouth. These microphones have the added advantage that head movements by the talker do not change the position of the microphone relative to the mouth. A drawback of a close-talking microphone is that if it is placed too close to a talker's lips, there may be a deficiency of nasal sounds and radiation from the throat in obstruent sounds.

Microphone placement

In the case of speech recorded for later analysis, it is important to choose a microphone type and placement that will enable you to acquire all of the sounds of speech in a balance that is natural. If your objective is to present speech as it would sound to a listener, a good place for the microphone is directly in front of the mouth of the talker at a distance of 10 to 20 cm. Unless the microphone is specifically designed for close placement, bursts of air produced while speaking will overload closely placed microphones.

Choosing the best recording medium

Modern equipment that is low in cost allows researchers to make recordings with quality and stability that were nearly impossible to achieve a few decades ago. There are, in fact, enough choices that the main reason for a selection is as likely to be convenience as cost or fidelity.

Cassette recorders

The conventional tape cassette ("Compact Cassette") is an excellent choice for speech recording, despite the wide availability of comparatively exotic digital media. The cassette has several excellent advantages: the media are universally available, low in cost, and generally more than adequate for speech recording and analysis. In addition, portable,

battery-operated cassette recorders are inexpensive and technically adequate. Using good cassette tape provides a dynamic range of about 50 dB; using Dolby noise reduction, an additional 10 dB (B-type) or as much as 20 dB (C-type) can be added. If recording is done with reasonable care, this is probably the best choice for general-purpose speech recording.

Digital audio tape (DAT)

Digital audio tape (DAT) recorders provide recordings of extremely high quality, almost always limited only by conditions in the recording environment. Most locations convenient for speech recording, especially at schools or universities, do not provide an ambient noise level low enough to allow full use of the dynamic range of a DAT recorder, which is about 90 dB. An unusually quiet room is likely to have an ambient noise level of about 30 dB SPL; speech at a conversational level will not usually have peaks larger than 80-85 dB SPL. Although these figures are only approximate, they indicate that equipment with a dynamic range of 50-60 dB should be adequate for high-quality speech recording.

The frequency response characteristics of DAT recorders and tapes are far superior to those of analog cassettes. In one area, the ability to record high frequencies at very high levels, for example, trumpet sounds, analog cassettes have great difficulty where digital media, including DAT tapes, have no problem. If the recorder is to be used to save musical instrument sounds, this may be a consideration.

It should be noted that DAT recorders and the tape that they use are much more expensive than their standard cassette counterparts. However, both kinds of tape must be carefully stored in cool, dry

environments if they are to be preserved for more than a few years. This is especially true in climates where heat and humidity are abnormally high.

The compact disc

Still another generally available medium is the compact disc or CD, which is different from the above media in that standard (CD-R) media cannot be erased and reused. A more expensive blank disk, CD-RW, can be erased and reused. Currently the standard marketing medium of the commercial recording industry, CDs can be made at relatively low media cost when a permanent, non-erasable digital recording of high quality is needed. Current costs are a dollar or so for CDs that hold up to 80 minutes of stereo recording at a 44100 Hz sample rate. A useful characteristic of the CD is the ease and speed with which a specific location in the recording can be found by its time index, or by its track or index number if these have been placed on the CD during recording.

The MiniDisc

A newer medium that appears to combine the best attributes of all of the above is the MiniDisc. These tiny disks, about 60 mm in diameter, hold up to 74 minutes of stereo recording sampled at 44100 Hz. If the recording is monophonic, a disc will store nearly 2½ hours with the same fidelity. MiniDiscs can also be recorded and replayed with relatively inexpensive equipment. At the time of writing, a desktop recorder-player can be purchased in the U.S. for about \$275; tiny pocket units that record and play are about the same price.

MiniDiscs are recorded by a method that is both magnetic and optical. One manufacturer claims that a disk can be recorded and played more than a million times. All of the devices record over the same very wide frequency range.

MiniDiscs achieve such compact storage through a form of compression sometimes referred to as “psychoacoustic-based compression.” When played back, the compressed sound should be indistinguishable from the original. However, some listeners claim that when certain program material is played, they are able to hear the effects of the compression.

Index

A

- Adjusting Energy Threshold 28
- Adobe Reader 6
- Analysis Mode 12
- Andrade, Amalia 4
- Animate
 - Selection 36
 - Spectrum 36
- Audio Card 5
- Audio Menu 11
- Auto-Level mode 26
- Average
 - Energy threshold 13
 - Interval 37
 - Spectrum 35

B

- Beaudoin, Robert 4
- Berkovitz, Robert 4
- Blackman Window 24
- Burg's Algorithm 25
- Button and Keystroke Commands 15
- Buttons 8

C

- Carlson, Eric 4
- Carr, Jason 4
- Cassette recorders 54
- CD-ROM 6
- Changing size of displayed segment 16
- Clear Spectrogram 14
- Clipboard 11
- Close 10
- Color 12
- Commands, Main Screen 10
- Compact discs 55

- Computer speed 5
- Condenser Microphone 53
- Contents 2
- Control Window 8
- Copy 11
- Copyright 3
- Cursor Coordinates 8
- Custom Energy Threshold 13
- Cut 11

D

- DAT Files 16
 - Converting 10
- Data Plots 44
- Decibel value, in Spectrum Viewer 34
- Delete 11
- Density Plots 44
- Digital Audio Tape (DAT) 54
- Disk, SpeechStation2 6
- Display
 - Formant Tracks 30
 - Information 10, 18
 - Segment 8
 - Window 8, 9
- Dolby, noise reduction 54
- Dynamic Microphones 53

E

- Edit Menu 11
- Energy (or Waveform) Plot 14
- Energy Threshold
 - Adjusting 28
 - Average 13
- Esling, John H. 48
- Exit 10
- Export
 - Formant Tracks 13, 30

Pitch Track 13, 29

F

Fast Fourier Transform (FFT) 23

Mode 12, 23

File 8

Menu 10

Open 10

Filter 13, 46

Formant Tracks 13, 30

Frequency resolution 23

G

Gaylord, Harry 48

Gray Scale 12, 26

Adjusting 26

Grid Display 27

H

Hamming Window 24

Hanning Window 24

Headphones 5, 19

Help, from Sensimetrics 3

Historical Note 4

I

If you hear no sound 19

Initial display 16

Insertion point 11

Installation 6

J

Jorgensen, Jens 4

K

Keystroke equivalents 8, 15

L

Labels 31

moving 31

removing 31

Ladefoged, Dr. Peter 48

License agreement 3

Linear Predictive Coding (LPC) 23

Linking .WAV files and SpeechStation2 6

Loading .DAT Files 16

Loading a file 16

Loading and Navigating A File 16

Loading Multiple Files 16

Loudspeakers 5, 19

LPC 12

LPC mode 24

LPC Order 12

LPC spectrum 25

M

Main Screen 8

Main Screen Commands 10

Making Better Speech Recordings 53

Maximize 27

Menus 8

Audio 11

Edit 11

Settings 12

Tools 13

View 14

Microphone 5

condenser 53

dynamic 53

noise-cancelling 53

- placement 54
- MiniDisc 55
- Mode menu 35
- Monitor 5
- Mono 20
- Mono recording 20
- Mouse 6
- Moving through a file 17
- Multiple Spectra 35

N

- Navigation Bar 8, 16
 - Operations 17
- New Recording 20
- New spectrogram 10
- NIDCD 4

O

- Open 10
- Opening SpeechStation2 Automatically 6
- Order, LPC filter 25
- Ortiz, Julio 4
- Overview 8

P

- Paste 11
- Phonetic Library 48
- Pickett, James 4
- Pitch 28
 - Multiplier 13, 29
- Pitch Track
 - Plot 13
 - Setup 28
- Play
 - Display 11
 - File 11
 - Selection 11

- Playback and Recording 19

Plots

- Data 44
- Density 44
- Pitch Track 13, 28
- Vowel Space 44
- Waterfall 13, 38
 - Waveform/Energy 14, 27
- Pre-emphasis Control 27
- Pre-emphasis Filter 12
- Printer 5
- Printing 10, 32

R

- Read Only 10
- Real-Time Spectrogram 13, 42
- Recording 11, 19, 20
 - Level 21
 - Level Bar 21
 - Medium 54
- Rectangular Window 24
- Redraw Spectrogram 14
- Registration with Owner identification card 3
- Repeat Play 11
- Requirements for Use 5
- Resizing Windows 27
- Resolution of the Pitch Estimate 28

S

- Sampling rate 20
- Save 10
- Save As... 10
- SBIR 4
- Scale
 - Energy Plot 27
 - waveform display 27
- Selected segment 8, 9

- Selecting
 - a microphone 53
 - an Analysis Window 23
- Selection 8, 14
 - Data 8
- Settings Menu 12
 - for Spectrum Viewer 37
- Signal segments 8
- Single Spectrum 35
- Spectral analysis 23
- Spectrogram 8, 9
 - Analysis Options 23
 - Display Options 26
 - Grid 14
- Spectrum Viewer 14, 34
 - Mode Menu 35
 - Options Menu 36
 - Settings 37
- SpeechStation 4
- SpeechStation2 4, 6
- SSManual.PDF 6
- Stereo recording 20
- Stevens, Kenneth 4
- Stop Playing 11

T

- Tape recorder 5
- Technical support 3
 - Upgrades 3
- Time resolution 23
- Time-Frequency Trade-off 23
- Title bar 8
- Tools Menu 13

U

- Undo 11

V

- Viana, Çeu 4
- View Menu 14
- Vowel Space Plot 13, 44

W

- Warranty 3
- Waterfall Plot 13, 38
- Waveform/Energy 8, 9
- Waveform/Energy Plot 27
- Window
 - Size 12
 - Type 12
- Windowed Waveform 36
- Windows
 - 3.1. 5
 - Volume Control 19

Z

- Zurek, Patrick 4

SENSIMETRICS