

Windows EDW (WEDW)

1. Introduction

WEDW displays speech (or other sound) files stored in Microsoft .wav (RIFF) format. It provides a variety of standard speech displays and analyses like spectrograms, f0 contours, RMS amplitudes, spectral cross sections, etc. WEDW allows editing of the speech data such as cutting and pasting regions of a displayed waveform and inserting waveform data from other files into the displayed waveform. WEDW also recognizes and allows editing of segment (region) labels embedded in waveforms as an extension to RIFF format to delimit specific regions within the waveform. Waveforms may contain multiple channels of data, some or all of which may be simultaneously displayed.

2. Differences from Unix/DOS EDW.

WEDW is a fundamentally new program which attempts to provide similar functionality to the Unix/DOS version (EDW), but with a very different user interface. WEDW functions in a way that is most similar to EDW's "mark" mode, i.e., a cursor is always present and single-key segment marking and window positioning functions are always active. WEDW replaces EDW's command line interface with menu items for the selection of optional settings, extended mouse functions for segment label manipulation, and dialog boxes where text input is necessary. WEDW retains some of the appearance of EDW in that a waveform display region is always present while spectrogram and pitch marking windows can be toggled on and off as desired. Both EDW and WEDW read and write waveforms in an extended RIFF (Microsoft .WAV) format that includes waveform segment definitions and both are also able to read an older .WAV format that was the original format used by EDW. WEDW is still under development with several planned features presently inactive.

3. Appearance

Figure 1 shows the screen appearance of WEDW when a waveform and spectrogram window are displayed. The waveform display region is always present. The lower portion of the window can display a spectrogram, f0 contour, or RMS amplitude contour, and a small region above the waveform display can show pitch marker locations. The overall WEDW window can be resized using the mouse to click and drag a window corner. The horizontal scale of all windows and the vertical scale of the waveform display window will be adjusted to the new overall window size. However, the vertical size of the lower display region used for spectrograms and other functions is a constant number of pixels as determined by a parameter which can be changed as desired.

In addition to the standard window controls, features to note are:

- The waveform file being viewed is displayed in the window title bar along with an indication of the file format.
- A menu bar contains pull-down menus for File access, Editing, analysis/display Options, Play mode, spectral Modeling, and (for multi-channel files) Channel settings.
- A status/toolbar contains readouts of waveform and spectrogram information and several buttons to control the display format and play the viewed signal.

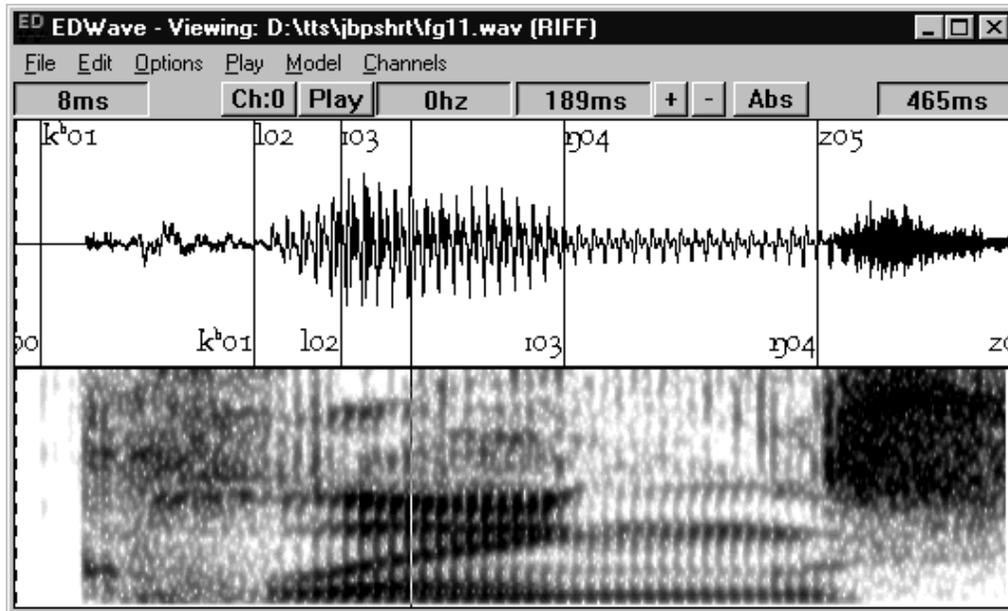


Figure 1 Typical WEDW display of a labeled waveform with spectrogram enabled

All these features are described below in more detail.

4. Using the Keyboard in Wedw

When the mouse pointer is within one of the graphic display windows, the following keys are bound to events as follows (case is ignored):

- **B** - The current vertical cursor location marks the beginning of a segment. A dialog box appears to accept the segment name.
- **E** - The current vertical cursor location marks the end of a segment. A dialog box appears to accept the segment name.

Note: when either the B or E marker of a new segment is first entered, both markers appear at the same location. WEDW allows one of the marks to be changed to a new location to generate a segment rather than location marker. Thereafter, if either begin or end marker is to be moved, WEDW will request confirmation before allowing the change.

- **L** - The waveform is repositioned within the display window so that the sample at the current vertical cursor location is moved to the left border of the display. The time-extent of the window is unchanged.
- **R** - The waveform is repositioned within the display window so that the sample at the current vertical cursor location is moved to the right border of the display. The time-extent of the window is unchanged.
- **^L- (CTRL+L)** The waveform is repositioned within the display window so that the sample at the current vertical cursor location is moved to the left border of the display. The waveform sample located at the right border of the display is unchanged and time-extent is reduced to "stretch" the waveform to fit within the display window.
- **^R- (CTRL+R)** The waveform is repositioned within the display window so that the sample at the current vertical cursor location is moved to the right border of the display. The waveform sample located at the left border of the display is unchanged and time-extent is reduced to "stretch" the waveform to fit within the display window.
- **V** - Set the selected pitch marker to indicate a voiced pitch event.
- **U** - Set the selected pitch marker to indicate an unvoiced event.

Note: V and U function only when the pitch marker to modify has been selected by clicking the left mouse button on the marker.

5. Region-specific functions.

In addition to the standard use of the mouse with pull-down and pop-up menus, WEDW uses mouse events within regions of the display as described in the following sections

5.1. Pitch Marker Region -

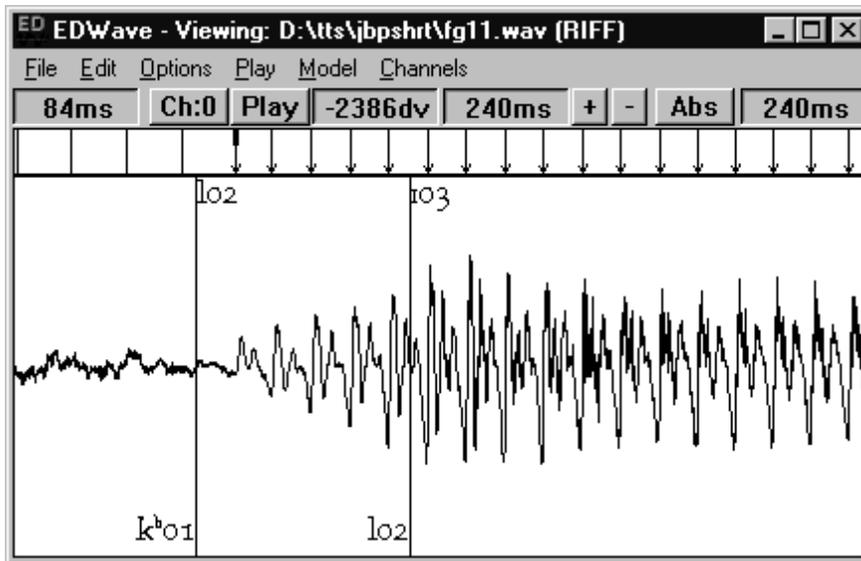


Figure 2. Waveform display with pitch marking window active.

The pitch marker region as illustrated in Figure 2 is a small window above the waveform display window in which markers are placed to indicate pitch and other events associated with the waveform. Markers that are associated with pitch periods in voiced regions of

the signal have a small arrowhead at their base and are called *voiced markers*. Other markers may be placed to divide voiceless regions of the signal into smaller epochs of a size similar to a pitch period. These are termed *voiceless markers*.

When the pitch display is first enabled, WEDW searches the current working directory for a file that has the same base name as the waveform file being viewed but with the extension PPS. If a PPS file is found, it is read and information from the file is used to build the pitch marker display. However, WEDW is also able to estimate pitch marker information directly from the waveform (a process called *pitch tracking*) and use its estimates for the pitch marker display. Section 8 below provides more details on using the pitch tracking in WEDW.

When the pitch marker window is visible, the following mouse actions are used to edit the information:

- Left_click on a marker to select it. The selected marker has a broader than usual top, as for example, the first voiced marker in the display above. Pressing the **U** key will set a selected marker to unvoiced; Pressing the **V** key will set a selected marker to voiced.
- Left_double_click on a marker to delete it.
- Left_double_click on white space to insert a marker. (Hint: unless the waveform is zoomed in so that pitch periods are well separated, it can be very difficult for WEDW to distinguish between a double click intended to insert a new mark and one intended to delete a nearby existing mark. Be sure to zoom in enough to make the difference obvious.)
- Left_press and drag to move a marker to a new location.

5.2. *Waveform/Spectrogram Windows -*

- Left_click to place a marker for subsequent Paste/Insert.
- Left_press and drag to select a region (it will be bounded by dashed lines when the mouse button is released). This region becomes the object of CUT/COPY/+Play selections or buttons.
- Left_press and drag on a segment label to drag the segment boundary to a new location. By default, any segment boundaries which lie at the same sample location as the selected label will be moved as well (see Label Grouping below). (Hint: Be careful in selection a region that you don't accidentally capture a segment and drag the boundary instead.)
- Left_double_click on a segment label to bring up a dialog box which allows the segment to be deleted or the segment name to be changed.
- CTRL+Left_press and drag to select and move only a single marker from a group of markers assigned to the same sample location. (see Label Grouping below).

Note: At all times, status windows in the toolbar report the time-coordinate of the mouse pointer. When the pointer is in the Pitch or Waveform window, the waveform amplitude

is also displayed as an unscaled digital sample value (designated dv units). When the mouse pointer is in the spectrogram window, the status window displays the frequency coordinate corresponding to the vertical pointer location.

5.3. *F0/RMS Windows* –

- **Left_click** (when line drawing mode is enabled) at a frequency or amplitude by time coordinate to place a sketch marker (a green X) at that point. A subsequent **left_click** at a new location within the window will draw a straight line from the location of the previous sketch marker to the present location and the sketch marker will be moved to the present location. This only occurs when *line drawing* mode is enabled in the Edit menu. (See Prosodic Modification)

6. **Toolbar Buttons**

- **Play** - clicking the play button will play the signal in the active channel for either the selected region, the waveform subsumed by the window, or the entire file depending on which mode is selected in the Play menu item. Note that play will always play the signal associated with the active channel whether it is visible in the display or not.
- **Ch:n** - This button always displays the currently *active* channel number, that is, the channel which will be heard when play is pressed and from which a spectrogram will be computed (if the spectrogram is turned on). Pressing the channel button will pop up a dialog box which allows a different channel to be selected as the active channel. When viewing single channel files, the button always displays **Ch:0** and pressing the button will have no effect.
- **+ -** If a selected region is smaller than the present display window size, the + button zooms into that region. Otherwise, the display is zoomed by a factor of two (i.e., the temporal resolution of the display is doubled and displays half as much "real" time). Pressing the Control key while clicking the + button will undo the last zoom (i.e., unzoom).
- **- -** The display is unzoomed by a factor of two. Double clicking the - button will display the entire file in the display window.
- **Abs** - Indicates that cursor time readout and end-time are in "absolute milliseconds" offset from the beginning of the waveform file. Clicking the **Abs** button will switch it to **Rel**.
- **Rel** - Indicates that cursor time readout and end-time are in milliseconds relative to the begin time at the left edge of the display. Thus, the end-time is the duration of the current window and the cursor time is its position offset from the left edge of the display. Clicking the **Rel** button switches it to **Abs**.

7. **Menu Items**

Other items may appear in the menus, the following mentions only those which are active.

7.1. *File:*

- **View...** - Open a waveform file to view and/or edit.
- **Save** - save any changes to the current file.
- **Save As...** - save current file under a new name.
- **Stats** - show waveform file information (sampling rate, number of channels, and segment definitions).
- **Quit** - Exit WEDW without saving changes.
- **Exit** - Exit WEDW after saving changes.

7.2. *Edit:*

- **Cut** - Delete the currently selected region, but save it in a paste buffer.
- **Copy** - Copy the selected region to the paste buffer.
- **Paste** - Insert paste buffer contents at selected location.
- **Select all** - As though the entire window was selected with the mouse.
- **Rename** - Change the name of a segment.
- **Delete** - Delete a segment definition (waveform is unchanged).
- **Insert** - Insert the contents of another waveform file into the present file at the selected location. The inserted file also replaces the contents of the paste buffer.
- **Line Drawing** - Selects line drawing mode for modifying F0 and RMS contours. (See Prosodic Modifications)
- **Modify** - Applies changes to F0 and RMS contours. (See Prosodic Modifications)
- **Global...** - Reports duration, average F0, and average RMS amplitude in the selected region and allows these to be altered. (See Prosodic Modifications)
- **Smooth** - Runs a low pass filter over the F0 or RMS data and applies the smoothed contour to the speech. (See Prosodic Modifications)
- **Invert Active** - Inverts the waveform for the active channel.

7.3. *Options:*

- **Waveform...** - Select among LINE|BAR|DOT waveform display formats and set scale multiplier for waveform display.
- **Labels:**

Group - assign neighboring label boundaries in a selected region to a single sample location.

Settings... - Bring up dialog box for label settings to select the label font, size, and mapping (see section 10. Label Settings).

- **Spectrogram:**
 - Toggle On/Off** - toggle spectrogram window.
 - Settings...** - Bring up a dialog box to set various parameters of the spectrogram (see section 8.1 Spectrogram Settings).
- **Pitch:**
 - Toggle On/Off** - Toggle pitch marker display.
 - Settings...** - Adjust properties for pitch tracker (also reports F0 statistics). (See Pitch Tracking)
 - Track** - Estimate pitch period locations for waveform visible in the window. (See Pitch Tracking)
 - Clear** - Erase all existing pitch marks for waveform visible in the window. (See Pitch Tracking)
 - Tweak** - Adjust voiced pitch marks to “best” nearby location. Useful for marks originally inserted by hand to improve the regularity of the pitch contour.
- **F0 Contour:**
 - Toggle On/Off** - toggle F0 window.
 - Settings...** - Bring up dialog box to set parameters of the F0 contour display including the height of the display window and the range of f0 values within the display window.
- **RMS Contour:**
 - Toggle On/Off** - toggle RMS window.
 - Settings...** - Bring up dialog box to set parameters of the RMS contour display including the height of the display window and the range of amplitude values within the display window.
- **Spectral Cross Section:**
 - Toggle On/Off** - toggle RMS.
 - Settings...** - Adjust settings for spectral cross section display (see Spectral Cross Section).
- **Save State** - Save current parameter settings to wedw.ini file in the current directory or in the directory specified in the EDWPATH environment variable. These settings will be in effect whenever edw is started in the directory.

7.4. *Play:*

- **Window** - Play button plays window contents (DEFAULT).

- **Region** - Play button plays selected region regardless of the present window contents (even if the region is no longer in the display window).
- **File** - Play button plays entire file regardless of the display window or region markers.

7.5. Channel:

The Channel menu only appears in the main menu bar when viewing a multi-channel file. By default, all channels of a multi-channel file are simultaneously displayed in the waveform window by vertically subdividing the window into as many regions as there are channels. However, the Channel menu allows any combination of channels to be selected for display. To select a single channel, simply select that channel in the display; all other selected channels will be deselected. To add a channel to one or more channels already selected, press the control key *while the mouse is in the waveform window* then while still pressing the control key, select the channel to be added from the Channel menu. This awkward arrangement is because at the moment WEDW will not detect the control key press unless the mouse pointer is in the waveform window. The items that appear on the Channel menu are:

- **All** - When checked (the default) all channels of a multi-channel waveform file will be displayed simultaneously within the waveform window.
- **Channel n** Select channel n for display in the waveform window. There will be N (=the number of channels of data in the file) similar entries in the menu.

8. Spectrograms

The spectrogram displays speech information in a time-by-frequency-by-amplitude format. The time (X) axis is aligned with the waveform window time axis. The frequency (Y) axis indicates ascending frequency from a low frequency cutoff value (at the bottom of the display) to a high frequency cutoff value at the top of the display. Both these cutoff values can be set in the spectrogram settings dialog box (see below). Finally, sound amplitude at every time-by-frequency coordinate is displayed in shades of gray with white being the lowest amplitude and black the highest amplitude. The lowest and highest amplitude values can

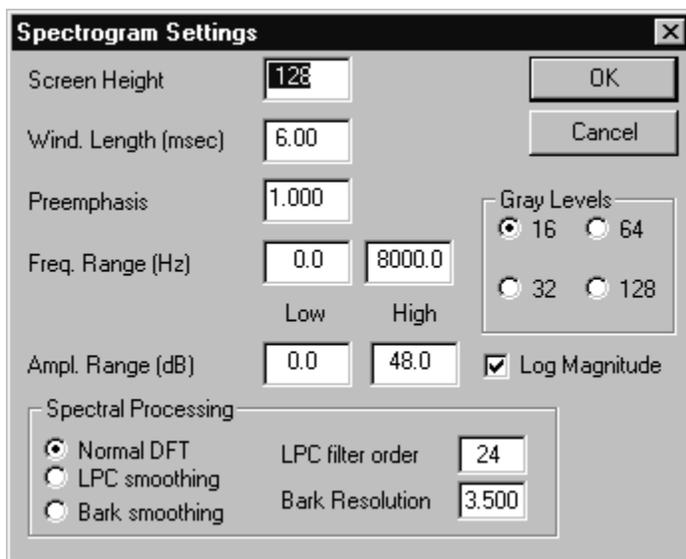


Figure 3 Dialog box for setting spectrogram parameters.

be set in the spectrogram settings dialog box.

In addition to standard spectrograms based on successive windows of short time Fourier transforms of the waveform, WEDW can display spectrograms based on LPC and Bark feature analyses of the speech signal. The LPC spectrogram can provide narrow band formant frequency resolution without harmonics. The Bark feature analysis displays spectral information with the frequency resolution of the human auditory system, giving fine spectral resolution at low frequencies and coarse spectral resolution at high frequencies.

8.1. Spectrogram Settings Dialog Box

Select “settings...” under the Options → Spectrogram menu to bring up the dialog box for adjusting spectrogram parameters. Figure 3 shows this dialog box with its default settings as follows:

- **Screen Height** -- the vertical extent of the spectrogram window in pixels. The spectrogram drawing is fastest when this number is a power of two and equal to the number of frequency values in the FFT used to compute the spectrogram. The size of the FFT is determined by a combination of the analysis window length (the second parameter in the display box) and the waveform sampling rate. The FFT length is the smallest power of two that is at least 64 and greater than or equal to the number of speech samples in the analysis window. For instance, if the sampling rate is 10 kHz and the window length is between 6.4 and 12.8 msec, the FFT length will be 128. When the screen height is not a power of two or not equal to the FFT length, EDW must interpolate values of the spectrum to the screen pixel locations and this slows down the drawing process.
- **Window Length** – the length of the FFT analysis window in msec. This parameter determines whether the spectrogram will be a narrowband or wideband spectrogram. Specifically, the bandwidth in Hertz of the spectrogram is approximately $2.0/\text{Length}$ in msec of the window. Thus, a window length of 6.0 msec corresponds to a bandwidth of about 333 Hz ($2/.006 = 333$) which is approximately the standard for a wideband spectrogram. For a narrow band spectrogram (45 Hz bandwidth) the window length would need to be about 44 msec, however, values in excess of 25 msec usually produce reasonable narrow band spectrograms.
- **Preemphasis** – a coefficient in the range 0.0 – 1.0 which has the effect of emphasizing high frequency energy and de-emphasizing low frequency energy. Normally speech is heavily weighted to the low frequency end of the spectrum. A coefficient of 1.0 creates a highpass filter that removes all DC energy and has a 6 dB/Octave slope. Values of the preemphasis coefficient closer to zero produce less high frequency preemphasis.
- **Frequency range** – specifies the frequency coordinate at the bottom (low frequency) and top (high frequency) of the spectrogram. By default, the low frequency cutoff is 0 and the high frequency is 5000 Hz or one half the sampling rate (the Nyquist frequency), whichever is smaller. You may change these parameters to see more detail

of a region within the spectrum, however, the high frequency value cannot exceed the Nyquist frequency.

- **Amplitude Range** – the decibel range from lowest (white) to highest (black) in the spectrogram. All amplitudes below the low value will appear white and all amplitudes above the high value will appear black. Note that these values are in decibels whether or not the Log Magnitude check box is selected.
- **Log Magnitude** – the scaling applied to the amplitude range. When this box is checked, the grayscale for the spectrogram will be assigned in logarithmic steps which has the effect of allowing small differences at low amplitudes to be more obvious in the display. Removing the check from this box will cause amplitude to be displayed in linear grayscale in which most lower amplitude regions of the spectrogram will appear white. On displays that allow only a small number of shades of gray, the linear scale can often provide a more readable spectrogram.
- **Gray Levels** – the number of shades of gray between white and black. This only has an observable effect on low resolution palette-based displays that show 256 or fewer colors. If the display color mapping is set to 16-bit or higher color resolution, WEDW always uses 128 gray shades.

9. Pitch Tracking and Settings

To perform pitch tracking in WEDW, the pitch window must be toggled on. This enables the Track item in the Pitch submenu of the Options menu. There are two modes of operation for the pitch tracker. For both modes, pitch tracking is only applied to the portion of the waveform that is visible in the waveform window, and pitch marks, if found, will replace all previously defined pitch marks for the visible portion of the waveform. This makes it possible to use different pitch tracking settings for different parts of the waveform if necessary. In the *supervised* pitch tracking mode, an example pitch period must be located manually by using the mouse to select a region of the waveform corresponding to one pitch period. This example is then used to *seed* the pitch tracker which will search the portion of the waveform visible in the waveform window in both directions starting at the location of the example pitch period. For the other, *unsupervised*, mode of pitch tracking, it is not necessary to manually select a seed; WEDW automatically selects a seed period using waveform amplitude and the F0 Mean value given in the Pitch Settings dialog box (Figure 4). Once a seed is selected either manually or automatically, the pitch tracking algorithm (described below) is the same for both pitch tracking modes. A checkbox labeled *Require Seed* in the Pitch Settings dialog box determines which method WEDW uses.

By default, the *Require Seed* box is checked indicating that the region markers delimit a single pitch period for use as the seed. When the Track item is selected in the Pitch menu, WEDW checks to see that the selected region corresponds to a period associated with an F0 between the Min and Max F0 values given in the Pitch Settings dialog box. If this condition is not met, WEDW displays an error box and you must either adjust the markers by reselecting the region, or adjust the values for F0 Min or Max in the dialog box then select pitch tracking from the options menu again. Note that the F0 Min and

Max values are used both to screen the seed selection, and to report statistics after pitch tracking has completed. Therefore their values will generally change after each call to the pitch tracker and in some cases, especially after a failure in tracking, they may inherit unrealistic values that would need to be adjusted by hand before a new seed will be accepted.

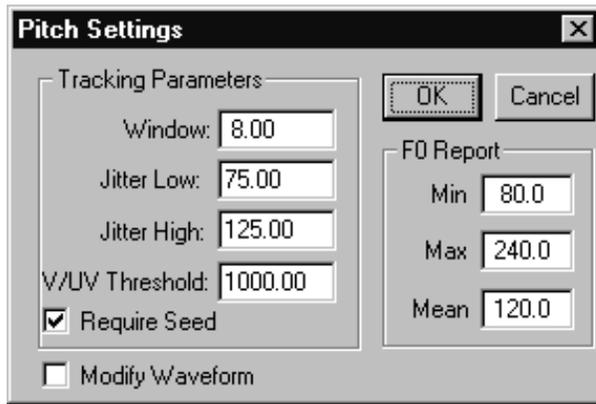


Figure 4 Dialog box for setting pitch tracking parameters.

When the *Require Seed* box is unchecked, the selected region, F0 Min, and F0 Max are all irrelevant, however, the F0 Mean value is used in estimating the duration of the internally generated seed period. Because of this it is important to set a realistic value in the F0 Mean field before starting unsupervised pitch tracking. As with F0 Min and Max, the Mean value is updated after each call to the pitch tracker to report the mean value obtained for the voiced pitch periods in the waveform. Thus, this field may also need to be reset to

a realistic value if the tracker failed to run correctly.

The pitch tracking algorithm in WEDW uses a raised cosine windowed portion of the waveform centered around the onset of the seed period as the initial search template for an adjacent pitch period. The onset is assumed to be a positive-going zero crossing preceding the first (usually strongest) F1 fluctuation in the seed period. The search template is compared to the structure of the waveform within a reasonable range of distances away from the seed period using a correlation statistic. This range is determined by the amount of allowable jitter (period-to-period fluctuations in F0) in successive periods. The location of the subsequent period is taken as the location at which the search template correlates most strongly with the waveform in the region being searched. If this correlation is above a voiced/unvoiced threshold value, it is assumed that another pitch period has been detected and the onset of the new pitch period is windowed and averaged with the previous search template to form a new template that is used in searching for the next pitch period. If the maximum correlation value in the search region is below the voiced/unvoiced threshold, the present location is assumed to be unvoiced and an unvoiced marker is placed at a location corresponding to the average pitch period following the last pitch marker. When an unvoiced region of speech is encountered, the search template is replaced with a windowed inverted sine wave having a period corresponding to the duration of the window. The same algorithm is applied in both directions starting with the seed period to identify all predecessor as well as all successor periods to the seed period.

The pitch settings dialog box allows the parameters of this tracking algorithm to be adjusted to improve the performance of the algorithm with various talkers. The tracking parameters that can be adjusted are displayed in the pitch settings dialog box:

- **Window** - The duration in msec of the cosine window used to select the search template. Values slightly less than the duration of an average pitch period seem to work fairly well.
- **Jitter Low** - Shortest likely adjacent pitch period expressed as a percentage of the current period. The default value of 75% seems to work most of the time.
- **Jitter High** - Longest likely adjacent pitch period expressed as a percentage of the current period. Something around 120% to 125% seems to work fairly well. Note that jitter values are always relative to the duration of the most recently identified pitch period. This makes it possible to set a fairly restricted amount of jitter and still track large variations in F0, as long as F0 changes gradually enough (i.e., the change from one period to the next does not exceed the allowed amount of jitter). The disadvantage of having jitter relative to the current pitch period is that errors, once made, tend to compound.
- **V/UV Threshold** - Voiced/Unvoiced threshold in arbitrary units. If this value is 0.0, the pitch tracking algorithm sets a value based on the seed period, otherwise, the value in the edit box is used. Values in the neighborhood of 1000 seem to work reasonably well for speech of moderate amplitude.
- **Modify Waveform** – Presently inactive.

10. Spectral Cross Sections

Toggling on the Spectral Cross section brings up a window to display the spectrum of the speech signal at the cursor location. As the cursor moves, the spectrum changes. This is illustrated in Figure 5. The cross section window is always on top of the waveform spectrogram display when it is over the main Wedw window, but it can be moved off the region of the main window. The window may be resized as desired.

When the mouse cursor is in the cross section window, a crosshair cursor follows its position and the coordinates of the crosshair cursor

are shown in the window title bar. This is illustrated in Figure 5 where the cursor is indicating a peak at 1619.8 Hz and 46.9 dB. Note that the precision of this readout is a function of the size of the display window. The precision of the measurement is improved by enlarging the window.

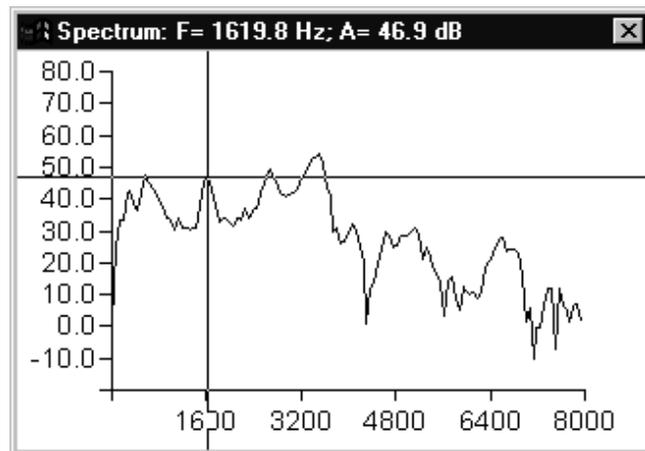


Figure 5 Spectral Cross Section window indicating crosshair cursor measurement of a spectral peak. Cursor coordinates are shown in the title bar.

Characteristics of the cross section display are controlled with the spectral cross section settings dialog box as shown in Figure 6. Many of the fields in the dialog box are similar to those in the Spectrogram Settings Dialog box and have equivalent meanings. The meaning of the fields in the dialog box are as follows:

- **Wind. Length (msec):** The length of the waveform region surrounding the location of the cursor that will be used for computing the spectrum. Wide windows result in narrow band spectral resolution (e.g., individual harmonics can be seen in voiced speech). Short windows result in broad band spectra that often do not show individual harmonics, especially with low pitched talkers. The default window length of 15 msec is chosen as a compromise between the two general types. However, if you wish to display the spectrum envelope, a much wider window is recommended; try about 40 msec for male talkers.

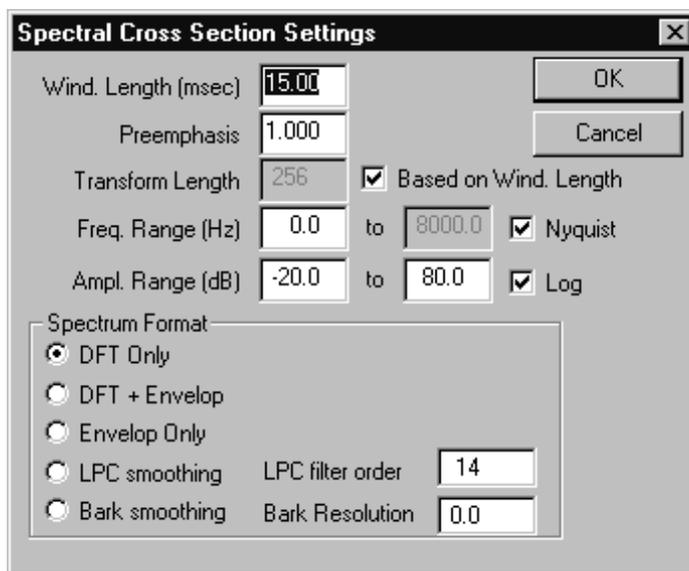


Figure 6 Spectral cross section settings dialog box.

- **Preemphasis:** This is a filtering coefficient which determines the amount of high frequency preemphasis to be applied to the speech.
- **Transform Length:** The length of the Fourier transform. This must be a power of two and large enough to contain all the samples in the window. The number of samples in the window is a function of the sampling rate (i.e., the window length in msec times the sampling rate in samples per second divided by 1000). To make life easy, you may check the box next to this edit field which indicates that Wedw should calculate the transform length based on the window length. In this case, the transform length box is switched to read-only as shown in Figure 6.
- **Freq. Range:** Specifies the low and high frequency cutoffs for the X axis of the display in Hz. If the Nyquist box is checked, Wedw will use the highest frequency available at the waveform sampling rate for the high frequency limit.
- **Ampl. Range:** Specifies the amplitude range for the Y axis of the display. The Log checkbox indicates that the display should be logarithmic.
- **Spectrum Format section:** The radio buttons in this section allow selection among different spectral display formats. These include: DFT – the raw Fourier transform spectrum; DFT+Envelop – the spectral envelope is traced along with the raw

spectrum; Envelop – The envelop is shown without the raw DFT; LPC – the LPC smoothed spectrum is shown instead of the DFT or its envelop; and Bark – the bark spectrum is shown (this is not presently implemented). In addition to the radio buttons, the number of coefficients for the LPC analysis can be set (the more coefficients, the more peaks can be found in the LPC spectrum, and the Bark resolution can also be set (but the bark display isn't working).

The envelop which can be shown in the spectral cross section displays is estimated by (a) using a cepstrum analysis to derive an estimate of f_0 , and (b) using a tunable comb filter to improve the estimate by finding the exact comb spacing that provides the best fit to the lower frequency portion of the spectrum. Lines are then drawn connecting the tips of the teeth in the best fitting comb filter. On slow computers, this can be very slow indeed.

11. Label Settings

The Label Settings dialog (Figure 5) provides control of the grouping of segment boundaries, the font used to display segment labels, and whether, by default, segment boundaries that share a location move together when a segment is moved by dragging the

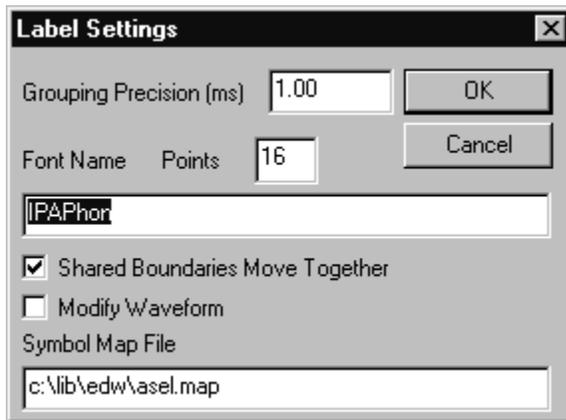


Figure 7 Dialog box for setting parameters of the segment labels.

boundary label with the mouse.

Generally, when a segment boundary is moved by dragging its associated label with the mouse, only the selected segment boundary changes. However, as a special case, if more than one segment has a boundary at exactly the same location as the selected boundary, all segment boundaries which share the location are moved together. This is especially useful when, for example, the boundary between adjacent phonemes in a labeled waveform is represented simultaneously by the end marker for the earlier segment and the begin marker for the

subsequent segment. Logically, it is the boundary comprising both segment markers that one is probably trying to move.

The grouping feature is provided because it can sometimes be difficult to place two boundary markers at exactly the same sample locations. By selecting the Group item in the labels menu segment markers which are quite close, but not exactly overlapped can be automatically adjusted to overlap. This in turn will ensure that they will normally move together when any boundary is dragged.

The *Modify Waveform* checkbox is presently disconnected.

- **Grouping Precision** - How close two boundary markers must be (in msec) to be collapsed to a single location when the Labels Group menu item is selected. This parameter works in conjunction with the selected region of the waveform to

determine which segment boundaries are to be grouped by assignment to a single sample location.

- **Font Name** - Font used for displaying segment labels. Obviously, this can be a phonetic font if one is available.
- **Points** - Point size for the label font.
- **Shared Boundaries Move Together** - Check box which determines the default behavior of overlapping boundary markers when one is dragged. The status of this checkbox can be inverted by pressing CTRL while dragging a segment label. That is, when the box is checked, CTRL frees a boundary to move alone, and when the checkbox is clear, CTRL causes boundaries to move together.
- **Symbol Map File** - File which specifies a mapping between characters or character strings in label names (which are typically standard ASCII) and a special symbol font such as an IPA font (see below). The file must be in the current directory or in the directory indicated by the EDWPATH environment variable.

11.1. Symbol Mapping for Segment Labels

WEDW provides a way to display special symbols such as IPA phonetic symbols when a font for the symbols is available. This is done by mapping between standard ASCII letters in segment labels and special symbol codes. The mapping interprets strings of alphabetic characters as tokens which can be replaced by characters or character strings from a specific font. For example, one might map the letter 'x' to the character code for schwa in an IPA font, or the sequence of letters 'ae' to the character code for the æ character in an IPA font. The mapping itself is read by WEDW from a user-constructed file which specifies the name of the special symbol font and the mappings from input (ASCII label characters) to output (character **codes** in the symbol set) using the format:

FontName

<input letter>[<input letter>...] <output code>[,<output code>...]

<input letter>[<input letter>...] <output code>[,<output code>...]

For example, Table I shows a portion of a symbol mapping table for a font called IPAPhon. In some cases, single characters map to single codes. This is the case for p, t, k among others in Table I. Sometimes, multiple characters map to a single code as for dx and ng in this example. Sometimes, two or more codes are needed to represent a given input sequence as for em, en, and eng in the example.

When symbol mapping is enabled by specifying the name of a mapping file in the Labels Settings dialogue box, WEDW will break every segment label into one or more string tokens and search the mapping table for a matching token. If a match is found, the token is replaced by the output codes for the token, otherwise, the input token is assumed to correspond exactly to the output sequence. That is, tokens which are not found in the mapping table are displayed (in the symbol font) using the ASCII code of the input character(s). As a result of this strategy, Table I contains a number of single character mappings which are actually unnecessary. In particular, the mappings for p, t, k, b, d, g,

m, and n are redundant since the symbol code for these letters is the same as their ASCII code. However, the ASCII letter 'q' maps to the phonetic symbol for a glottal stop and its code does not correspond to the character code for 'q, consequently, that entry is not redundant.

Table I - Example Mapping Table
IPAPhon

p	112
t	116
k	107
dx	228
b	98
d	100
g	103
q	214
m	109
n	110
ng	247
em	109,164
en	110,164
eng	247,164

WEDW tokenizes segment labels using a very simple set of rules. All adjacent alphabetic characters (i.e., the letters a-z and A-Z) are assumed to be part of a single token and all other characters are treated as token delimiters. With one exception, delimiting characters are displayed in the output without mapping. The exceptional case is the '-' character which may be used to introduce a diacritic. WEDW allows for the possibility that - introduces diacritics, but it makes no assumptions about the diacritics themselves which must be specified (including the '-') in the mapping table if they are to be mapped. For instance, we use '-n' to mean nasalized and therefor would specify a nasalized schwa with the sequence ax-n. Our mapping table contains the lines:

```
ax 92
-n 126
```

allowing the sequence ax-n to appear as \tilde{a} in the

WEDW mapped screen display.

12. Prosodic Modification

As of September, 1997 WEDW provides the capability to modify the prosodic structure of speech. This feature is based on and requires pitch marker information and, if the prosodic modifications are to be successful, the pitch marker information must be accurate. When pitch marker information is available, WEDW is able to display either F0 or RMS amplitude data in place of the spectrogram display. All three of these display types (F0, RMS, and Spectrograms) inherit the same screen height parameters and selecting any one of these will cause the lower display window to appear (if it was not present) or to have its contents replaced by the selected display type. The pitch marker window may be present simultaneously with any of these, but it need not be as long as pitch marker data is available (e.g., in a .PPS file) for the waveform being edited.

All data in the F0 and RMS contour displays are linked to the pitch marker data such that changing the location of a pitch marker will immediately change the value of F0 and potentially the RMS value associated with the pitch epoch. Changing the location of a pitch marker does not change the speech waveform. Editing the F0 or RMS contour using the line drawing or other features described below does not immediately change either the waveform or the pitch marker data, however, once the F0/RMS changes have been applied, the speech waveform will be modified and the resulting modifications will then be reflected in the pitch marker data.

Prosodic features of duration, F0, and amplitude can be changed in three distinct ways. Two of these depend on selecting a region of the waveform over which modifications are applied. Either F0 or RMS contours can be smoothed (i.e., low pass filtered) by selecting a region of the displayed contour and then selecting *smooth* in the edit menu. The second method which applies to a selected region allows additive changes to duration, F0, or amplitude. When a region is selected, the *global* selection under the edit menu brings up a dialog box which reports the duration, average F0, and average RMS amplitude for the selected region. Within the dialog box, any of these values can be changed and the changes applied to the selected region. Duration is changed by replicating or deleting pitch periods from the selected region to approximately achieve the desired duration (note that the actual duration after applying the changes will be within one pitch period length of the requested duration). F0 is changed by altering the duration of each pitch period. Generally, when F0 is changed it is also necessary to add or delete pitch periods to maintain approximately the specified duration, and as a result, duration will also change slightly when F0 is changed. RMS amplitude is changed by increasing or decreasing the amplitude of each pitch period (taken separately) to achieve the requested amplitude. Because there is some interaction between adjacent pitch periods, the resulting amplitude changes are also likely to be nearly, but not exactly, those requested. Moreover, when RMS and duration or F0 are changed simultaneously, these changes interact. RMS changes are applied first, and then duration/F0 changes are applied. This can result in deviations from the specified RMS value.

The third method for prosodic modification applies only to F0 and RMS contours since it involves changing the shape of the contour by sketching a new contour. For this method, it is first necessary to enable *Line Draw* mode (under the Edit menu) and thereafter, displayed values of F0 or amplitude are altered using the mouse. Once the desired contour has been drawn, the specified changes can be applied to the speech waveform by selecting the *modify* option in the Edit menu. In Line Drawing mode, clicking the left mouse button when the mouse pointer is within the lower window will place a *sketch mark* (a green X) at the pointer location. Moving the pointer to a new location and pressing the left button again will draw a new F0 or RMS contour by linear interpolation between the present pointer location and the sketch mark, and the sketch mark will then be moved to the present pointer location. Using this method, the desired contour is drawn by piece-wise linear approximation.

13. Analysis by Synthesis (AbS)

As of February, 1999 Wedw includes features for estimating formant synthesis parameters using an analysis by synthesis approach. These components of the program are highly experimental and are undergoing frequent adjustment and revision. The basic algorithm for the analysis by synthesis was presented at the Spring 1999 meeting of the Acoustical Society of America and the presentation slides are available on the speech research lab web site. The controls for analysis by synthesis are primarily located in two dialog boxes (one for source characteristics and one for filter (formant) characteristics) in the Model menu. However, the AbS processing also depends on settings in the spectral cross section dialog box which must be enabled to run the AbS.

13.1. Source settings

The Lilijenkrants, and Fant source model is used for the AbS. This model is parameterized by 6 features in Wedw (all are illustrated in the diagram shown in the source settings dialog box within Figure 8):

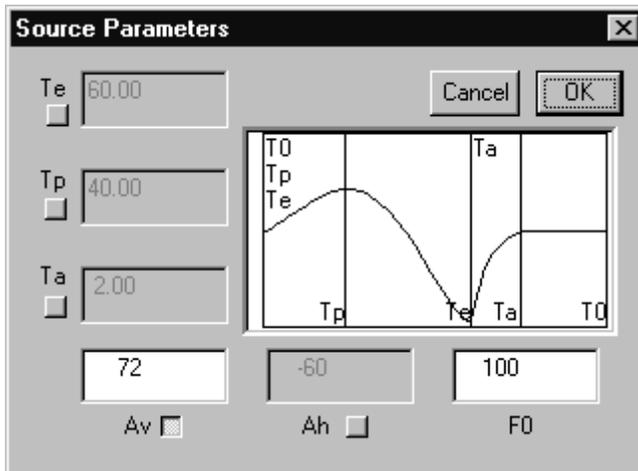


Figure 8 Dialog box for setting source parameters for AbS. The graph illustrates the meaning of each of the LF model features.

dB.

5. Ah – The amplitude of aspiration in dB. In Wedw, the aspiration is added in to the source waveform during the open phase of the glottal cycle only.
6. F0 – The fundamental frequency in Hz which determines T0, the pitch period. F0 is reported in the dialog box, but setting it has no effect since f0 is actually determined by the waveform being analyzed.

Each of the source parameters except f0 can be enabled or disabled by selecting/deselecting the push button next to the feature name in the dialog box. When a feature is disabled, its value is held constant at the value shown in the dialog box. It is not adjusted in the AbS processing, and it cannot be changed by the user as long as the button is checked. This allows one to design simpler AbS models by holding some features constant for the AbS while allowing the program to continue to adjust others for a better fit.

1. Tp – The proportion of the total pitch period from the start of the period to the time of maximum increasing change in airflow.
2. Te – The proportion of the total pitch period from the start of the pitch period to the time of maximum decreasing change in airflow.
3. Ta – Roughly the time from Te to the point where airflow ceases. Ta is actually calculated by projecting the initial slope of the return phase of the waveform to the point it crosses zero change in airflow.
4. Av – The amplitude of voicing in

13.1.1. Filter Settings

The synthesis filter for AbS is comprised of seven poles and two zeros. Each pole or zero

Formant	F1	F2	F3	F4	F5	P1	Z1	P2	Z2
Frequency	500	1500	2500	3500	4500	5500	250	6500	4000
Bandwidth	100	120	180	240	300	350	5000	5000	250
Active	<input type="checkbox"/>								

Buttons: Cancel, Use Higher Formant Correction, OK

Figure 9 Parameter settings dialog box for AbS vocal tract features comprised of 7 poles and two zeros.

is in turn specified by its frequency and bandwidth. The initial values or default values for each of the poles and zeros are set in the Filter settings dialog box (Figure 9). By convention, five of the poles are called formants and are labeled F1 – F5, the other two are simply called P1 and P2. The zeros are labeled Z1 and Z2. Each of these may be activated or deactivated using push button switches in the dialog box. When disabled, a pole or zero is held constant at the value shown in the dialog box. If active, the AbS procedure will use the setting in the dialog box as an initial setting.

In addition to the pole and zero settings, the dialog box contains a push button switch to enable or disable use of a higher formant correction to spectral shape. This may or may not be useful depending on sampling rate and the nature of the anti aliasing filter used to record the speech.

13.2. *Spectral Cross Section interactions*

The settings of the two dialog boxes just described interact with the settings of the spectral cross section. In particular, the AbS search range for comparing synthetic to original spectra is limited by the settings used for the spectrum in the Spectral cross section.

13.3. *Doing an AbS*

Once the initial settings have been established, and the synthesis model specified by selecting or de selecting parameters, the analysis is run by selecting *Fit Model* from the Model menu. All of the waveform displayed in the current window is analyzed in a pitch synchronous manner. Because the analysis is pitch synchronous, accurate pitch tracking information should be established before running AbS.

14. **Bugs, Gotchas, and other Beasties**

As of 7/10/99 -

- Multi-file waveform displays are still not enabled.

- There is no way within wedw to write a portion of the waveform to an external file.
- Pitch tracker cannot identify pitch periods at the edges of the display. This can cause markers to get dropped when retracking a portion of a vocalic interval.
- There are probably numerous small bugs and problems in the prosodic modification features.
- The analysis by synthesis modeling is extremely experimental and prone to failure. Use at your own risk!
- There should be a better way to print images from the program.