Marquette XBAT Filters

Summer 2009 Research Experience for Undergraduates

Corey Bost
Dylan Conley
Dex Delfrate

Contents

XBAT Filters

Installation

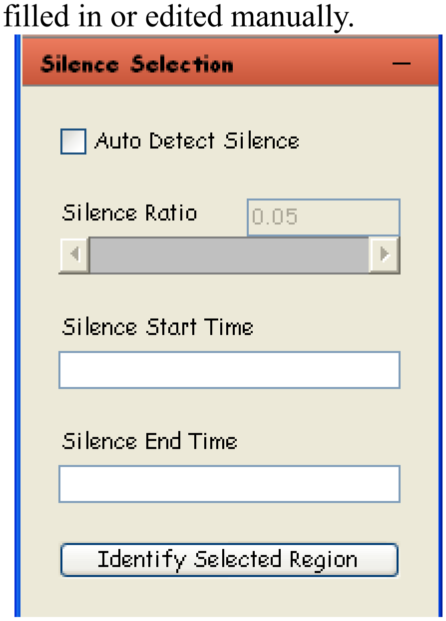
In order to use these filters, you must first have XBAT installed on the computer. XBAT can be downloaded freely at http://www.xbat.org. Once XBAT is installed, run the "Marquette Filters Installer.exe". When prompted for an install location, you must specify the main directory of XBAT (the directory which contains "xbat.m").

General Use

In order to access the filters, open a sound signal in XBAT and from the browser (spectrogram view) go to the Filters menu, followed by the Marquette submenu.
There are several features which appear identically across all five filters. The following sections explain the two groupings which will be common to all filter GUIs, the Silence Selection and Save Output features.

Silence Selection

The figure below shows the Silence Selection portion of all filters. When the filter is opened, the Auto Detect Silence box is checked by default. This check box serves as a toggle for the automatic silence detection algorithm (for info on how this works, please see Ephraim Malah in the function reference section). With the auto detection turned on, the only parameter to adjust is the Sampling Frequency which essentially sets a threshold at the given percentage of the overall signal strength (portions of the signal which fall below this threshold will be considered as silence). The default value for Silence Ratio is 5%. It should be noted that while auto detection is convenient, making a manual silence selection will often yield better results.
In order to manually select a region to use as silence, first uncheck the auto detect box. This will deactivate the Silence Ratio slider and in turn activate the Silence Start/End Time boxes. The next step is to make a selection (click and drag a region) on the spectrogram. Now, clicking the Identify Selected Region button will fill in the Start Time and End Time boxes. Alternately, the numbers in these boxes can be filled in or edited manually.

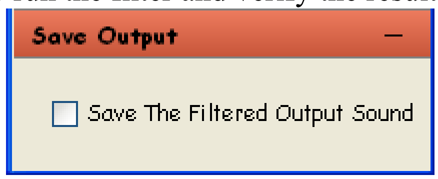


Save Output

As the figure below shows, the Save Output portion is quite straightforward. This allows the filtered sound signal to be saved. This is useful when applying more than one filter to a signal, or once a signal has been cleaned to a satisfactory level.

Turning on the filter with the save box checked will pop up a file save dialog once the filter has completed its computation. The output file will be of the .wav format, and can be saved to any location desired. Once the output file is saved, it will need to be manually added to the XBAT library.

NOTE: It is always advisable to run the filter and verify the results before saving the output.

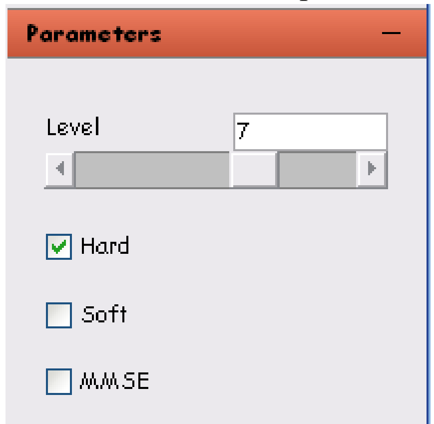


Direct Wavelet Decomposition Filter

In wavelet-based methods, the signal is decomposed using a wavelet transform instead of an FFT, which has certain beneficial properties in terms of time-frequency resolution. The wavelet coefficients are then thresholded (possibly using a noise model as with the above methods) and the signal reconstructed. There are a number of variations on the thresholding method.
This filter calls the following sub-functions: Direct Wavelet Decomposition, Wavelet OMLSA Cohen.

Parameters

The Direct Wavelet Decomposition filter has two main parameters: Level, Thresholding Type.



Level

The end level of wavelet decomposition tree. A typical value for this is eight.

Thresholding Type

Hard Thresholding: Where all coefficients below a predefined threshold value are set to zero.
Soft Thresholding: Where the remaining coefficients are linearly reduced in value.
Minimum mean square error: A technique designed to minimize the variance of the estimation or forecast error. This can also be referred to as OMLSA or nonlinear thresholding.

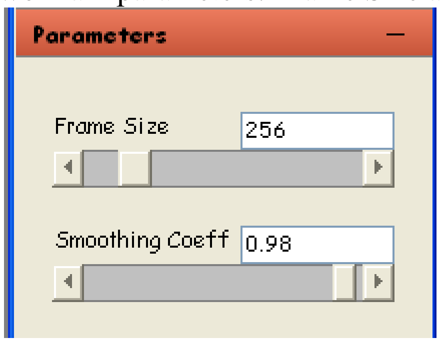
Ephraim Malah Filter

Ephraim and Malah in 1984 and 1985 derived formulas for optimal estimators of the magnitude spectrum of a clean signal given noise estimates and a statistical model for the signal and noise, for two different mean-square-error error criteria. The formulas are complicated but straightforward to implement.

This filter calls the following sub-functions: Ephraim Malah and Find Silence.

Parameters

The Ephraim Malah Filter has two main parameters: Frame Size and Sampling Frequency.



Frame Size

The Frame Size is simply the number of samples per signal frame. Each sound file has a certain sampling rate (for example: 44,100 Hz), which is the number of samples (data points) per second of signal. This means that one frame is a minute fraction of a second. The typical value for frame size is 256.

Smoothing Coefficient (Alpha)

The Smoothing Coefficient (typically about 0.98) represents the value of α (alpha) used in the filtering equation. Alpha is used as a percentage in a recursive approach to estimating the signal to noise ratio, where α% of the previous frame's information is used along with (1-α)% of the current frame's information. If this is confusing, leaving the default value of 0.98 is perfectly acceptable.

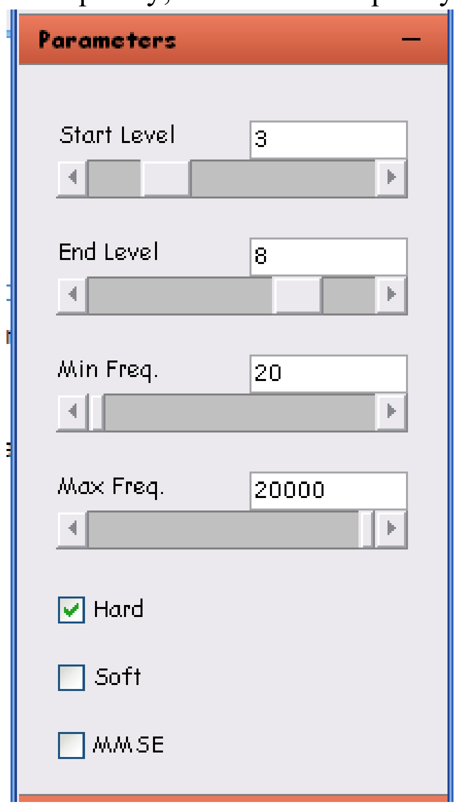
Perceptual Wavelet Decomposition Filter

Marquette University's Speech and Signal Processing Laboratory has developed a method for wavelet thresholding based on a perceptual auditory model, which can be applied to human speech or any species of animals (as needed for the Dolittle project). This method uses a special scaling on the frequency axis and an optimal thresholding based on modifications of the Ephraim Malah filter formulas.

This filter calls the following sub-functions: Automatic Down Decomposition, Calculate Frequency, Calculate Greenwood Frequency, Greenwood, Inverse Greenwood.

Parameters

The Perceptual Wavelet Decomposition Filter has five main parameters: Error: Reference source not found, End Level, Minimum Frequency, Maximum Frequency, Thresholding Type



Start Level

The start level of wavelet decomposition tree. This is normally set to one for the first level.

End Level

The end level of wavelet decomposition tree. A typical value for this is eight.

Minimum Frequency

The low threshold of hearing range for the selected animal.

Maximum Frequency

The high threshold of hearing range for the selected animal.

Thresholding Type

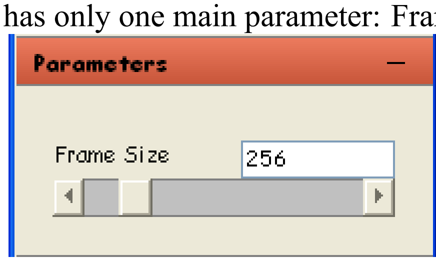
Hard Thresholding: Where all coefficients below a predefined threshold value are set to zero.
Soft Thresholding: Where the remaining coefficients are linearly reduced in value.
Minimum mean square error: A technique designed to minimize the variance of the estimation or forecast error.

Spectral Subtraction Filter

Spectral subtraction is the oldest and simplest technique for signal enhancement. In this method, the magnitude spectrum of the noise, estimated from a silence region, is literally subtracted from the magnitude spectrum of each frame, then recombined using the noisy phase spectrum to get the enhanced time domain signal.
This filter calls the following sub-functions: Spectral Subtraction and Find Silence

Parameters

The Spectral Subtraction Filter has only one main parameter: Frame Size.



Frame Size

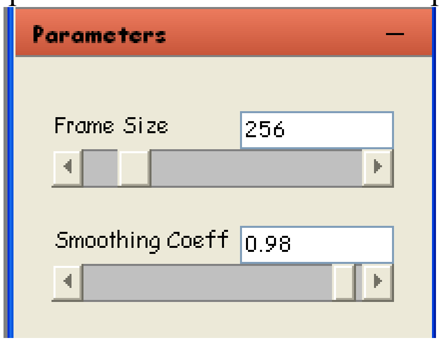
The Frame Size is simply the number of samples per signal frame. Each sound file has a certain sampling rate (for example: 44,100 Hz), which is the number of samples (data points) per second of signal. This means that one frame is a minute fraction of a second. The typical value for frame size is 256.

Wiener Filter

Wiener filtering is a method for frequency domain filtering based on estimates of both the signal and noise components. Since the signal component is obviously unknown, this is typically implemented through an iterative process where the noisy signal is run through the filter and then the result is used to improve the filter for the next iteration.
This filter calls the following sub-functions: Wiener and Find Silence

Parameters

The Wiener Filter has two main parameters: Frame Size and Sampling Frequency.



Frame Size

The Frame Size is simply the number of samples per signal frame. Each sound file has a certain sampling rate (for example: 44,100 Hz), which is the number of samples (data points) per second of signal. This means that one frame is a minute fraction of a second. The typical value for frame size is 256.

Smoothing Coefficient (Alpha)

The Smoothing Coefficient (typically about 0.98) represents the value of α (alpha) used in the filtering equation. Alpha is used as a percentage in a recursive approach to estimating the signal to noise ratio, where α% of the previous frame's information is used along with (1-α)% of the current frame's information. If this is confusing, leaving the default value of 0.98 is perfectly acceptable.

Reference

Alphabetic List of Parameters

End Level

The end level of wavelet decomposition tree. A typical value for this is eight.

Frame Size

The Frame Size is simply the number of samples per signal frame. Each sound file has a certain sampling rate (for example: 44,100 Hz), which is the number of samples (data points) per second of signal. This means that one frame is a minute fraction of a second. The typical value for frame size is 256.

Level

The end level of wavelet decomposition tree. A typical value for this is eight.

Minimum Frequency

The low end threshold of vocalizing range for the selected animal.

Maximum Frequency

The high end threshold of vocalizing range for the selected animal.

Sampling Frequency

The Sampling Frequency (or sampling rate) is the number of samples (data points) per second of signal. This is determined at the time the sound is recorded. An average sampling frequency is about 44,100Hz.

Silence Ratio

The Silence Ratio is only used when the Auto Detect Silence box is checked. It essentially sets a threshold at the given percentage of the overall signal strength (portions of the signal which fall below this threshold will be considered as silence). The default value for Silence Ratio is 5%. It should be noted that while auto detection is convenient, making a manual silence selection will often yield better results.

Silence Start/End Time

These values are only used in manual silence selection (Auto Detect box is unchecked). These values hold the timestamps of the beginning and end of the user-defined silence region. They can be entered manually by typing in the boxes or auto-filled by making a selection on the waveform and clicking Identify Selected Region button.

Smoothing Coefficient (Alpha)

The Smoothing Coefficient (typically about 0.98) represents the value of α (alpha) used in the filtering equation. Alpha is used as a percentage in a recursive approach to estimating the signal to noise ratio, where α% of the previous frame's information is used along with (1-α)% of the current frame's information. If this is confusing, leaving the default value of 0.98 is perfectly acceptable.

Start Level

The start level of wavelet decomposition tree. This is normally set to one for the first level.

Thresholding Type

Hard Thresholding: Where all coefficients below a predefined threshold value are set to zero.
Soft Thresholding: Where the remaining coefficients are linearly reduced in value.
Minimum mean square error: A technique designed to minimize the variance of the estimation or forecast error.

Functions

Automatic Down Decomposition

This function, "auto_down_decomp.m", takes as inputs minimum frequency, maximum frequency, low and high frequency (relating to the animal's hearing range), level start, and level end.

Calculate Frequency

This function, "calf.m", takes as inputs composition level, node index and maximum frequency. Its output is the resulting frequency.

Calculate Greenwood Frequency

This function, "cal_green.m", takes as inputs the number of filter banks used in calculating the Greenwood function, the minimum frequency, maximum frequency, low and high frequencies (relating to the animal's hearing range) and gives as output the calculated Greenwood frequency.

Direct Wavelet Decomposition

In wavelet-based methods, the signal is decomposed using a wavelet transform instead of an FFT, which has certain beneficial properties in terms of time-frequency resolution. The wavelet coefficients are then thresholded (possibly using a noise model as with the above methods) and the signal reconstructed. There are a number of variations on the thresholding method.
This filter calls the following sub-functions: Direct Wavelet Decomposition, Wavelet OMLSA Cohen and takes as input parameters Level and Thresholding Type.

Ephraim Malah

The function which handles the calculations in the Ephraim Malah filter is called "lsaenh.m". This function takes as input a signal, the Frame Size, the Smoothing Coefficient (Alpha), a Boolean value for auto silence detection, the Sampling Frequency, Sampling Frequency, and the Silence Start/End Time. The output is a filtered signal.

Find Silence

The Find Silence function (findsil.m) contains the algorithm used to automatically detect silence in the signal. This function takes as input a signal, the Frame Size, and the Sampling Frequency. The output consists of the indices of the frames which have been determined to be silence.

Greenwood

This function, “greenwood.m” takes as input the frequency in hertz and calculates the Greenwood domain frequency.

Inverse Greenwood

This function, “Igreenwood.m” takes as inputs the minimum and maximum frequencies as well as the real frequency in the Greenwood domain and outputs the Greenwood frequency in hertz.

Spectral Subtraction

The function which handles the calculations in the Spectral Subtraction filter is called “spec_sub.m”. This function takes as input a signal, Sampling Frequency, Frame Size, an event structure packed with Silence Start/End Time, a Boolean value for auto silence detection, and the Silence Ratio. The output is a filtered signal.

Perceptual Wavelet Decomposition

Marquette University’s Speech and Signal Processing Laboratory has developed a method for wavelet thresholding based on a perceptual auditory model, which can be applied to human speech or any species of animals (as needed for the Dolittle project). This method uses a special scaling on the frequency axis and an optimal thresholding based on modifications of the Ephraim Malah filter formulas.

This filter calls the following sub-functions: Automatic Down Decomposition, Calculate Frequency, Calculate Greenwood Frequency, Greenwood, Inverse Greenwood and takes as input parameters, Error: Reference source not found, End Level, Minimum Frequency, Maximum Frequency and Thresholding Type.

Wavelet OMLSA Cohen

This function, “wavelet_OMLSA_cohen” outputs enhanced wavelet coefficients using OMLSA thresholding.

OMLSA (MMSE)-Optimal modified least squared amplitude

Wiener

The function which handles the calculations in the Wiener filter is called “WF2.m”. This function takes as input a signal, Sampling Frequency, Frame Size, Smoothing Coefficient (Alpha), an event structure packed with Silence Start/End Time, a Boolean value for auto silence detection, and the Silence Ratio. The output is a filtered signal.