Song Scope

Bioacoustics Software Version 3.3 Documentation

Table of Contents

Introduction to Song Scope

by Wildlife Acoustics, Inc.

Be sure to visit our on-line user group forum for product announcements, questions, answers and tips. Click on the "Forum" link on our website at http://www.wildlifeacoustics.com/.

Welcome to version 2 of Song Scope! We have made major improvements to our classification algorithms (see the Release Notes for details). We are also working on new and improved documentation complete with examples to help you get the most out of Song Scope. Meanwhile, please contact us if you require any assistance. We have helped others by looking at recordings and making suggestions for how to improve sensitivity and accuracy, and would be happy to do the same for you.

Song Scope is a sophisticated digital signal processing application designed to quickly and easily scan long audio recordings made in the field and automatically locate vocalizations made by specific bird species and other wildlife.

Field biologists interested in the study of population density, migration patterns and behavior of specific animal species, especially those considered threatened or endangered, can use Song Scope to efficiently analyze long field recordings made by autonomous recording devices.

Song Scope was designed to be easy to use and does not require the user to be an expert in digital signal processing. That being said, some basic understanding is helpful. See Signal Processing Basics for more information.

Using Song Scope is as simple as 1 - 2 - 3:

1. Open and view recording files containing known vocalizations of the species of interest.
2. Manually find and mark (annotate) the known vocalizations
3. Song Scope automatically generates a "Recognizer" at the push of a button.

Once a Song Scope Recognizer has been created, you can use it to automatically scan new recordings for likely matches at the push of a button.

Release Notes

Version 3.3

We encourage your feedback, especially if you notice any bugs or have suggestions for improvements and new features. Please visit our on-line forum by visiting our website at http://www.wildlifeacoustics.com/ and clicking on the "Forum" link. You can also contact us at *songscope@wildlifeacoustics.com* or call toll-free +1 (888) 733-0200.

System Requirements

- All platforms require an Internet Connection during installation (for product activation after purchase). If a sound card is present, Song Scope can play audio portions of recordings.
- Windows 95/98/2000/NT/XP/Vista/7
- Linux RedHat i386
- Mac OS X 10.4 universal binary (ppc/i386)

Limitations

- Only PCM (uncompressed) .WAV files (RIFF/WAVE) and Audio IFF files are supported in addition to the Wildlife Acoustics Audio Compression (WAC) format. Anabat .??# files can also be opened, but they are converted from zero-crossing to 16-bit samples internally.

Version 3.3

Windows 7 Installer

> Windows version now packaged to support Windows 7

Bug fix: Hang on high sample rate with small frame size

> There was a bug in which Song Scope could hang when using small FFT sizes with large sample rates when viewing signal detection or normalized spectrograms

Version 3.2

Feature: Next button to skip gaps in triggered recording

> You can now press the "Next" button from inside the gap of a triggered .wac file recording to skip ahead to the next triggered event.

Enhancement: playback scrolling

The playback scrolling position was changed from 90% to 75%

Bug fix: anabat file processing

There was a bug in which Song Scope would incorrectly underestimate the duration of an anabat file so it would be displayed truncated.

Version 3.1

Bug fix: crash generating recognizer after changing sample rate

Sometimes changing the sample rate and then regenerating a recognizer would cause the Song Scope application to crash.

Version 3.0

Support for WAAC version 2

Added support for version 2 of Wildlife Acoustics Audio Compression (WAC) format. This includes lossy compression through dynamic range reduction to less than 16 bits, and zero-frame tokens to mark time with no signal such as from threshold-triggered recordings.

Support for Anabat file formats

Added support for Anabat file formats. Converts zero-crossing timings to samples internally so they can be viewed and manipulated by Song Scope.

Support for additional sampling rates

Added several new sampling rates including 9.6, 19.2, 120, 192, 240, 384, 480, and 768kHz

Add support for scanning without recognizers

You can now perform a scan without loading or selecting any recognizers. In this mode, all detected events are logged and no meaning is given to quality or score values.

Version 2.4

New feature: Flexible annotation file handling

Added "Save annotations as..." and "Load annotations" options to write

annotation files to directories other than the default SongScopeNotes directory under the original source recording files. Additionally, loading annotations merges new annotations with existing ones.

New feature: support for floating seat licenses

License manager software now allows floating seat licenses on a network for greater flexibility. Contact Wildlife Acoustics for more information.

Bug fix: batch scan with .ssn and other files

Batch scanning with a mix of .ssn files and non-.ssn files caused Song Scope to crash.

Version 2.3

Enhancement: Finding .ssn files after moving .ssr files

If you move a recognizer .ssr file and the relative paths to referenced annotation .ssn files changes, you will now be prompted to find the new relative location of the new .ssn file. If the rest of the .ssn file hierarchy is intact, then Song Scope will automatically locate all the remaining .ssn files. This also fixes a bug under Windows in which Song Scope might hang when trying to open a recognizer file after changing the relative position of annotation files.

Enhancement: Playback scrolling

With prior versions, the cursor remained fixed around the center of the screen while the spectrogram scrolled during playback. This has been changed such that the cursor will move up to the 90% mark on the spectrogram and then start scrolling. We believe that this will make viewing spectrograms during playback easier for most users.

Enhancement: Choosing winners when scanning with multiple recognizers

Improved how the winning match is chosen when scanning recordings with multiple recognizers in parallel. The quality value is no longer a factor as long as the minimums are reached, and the scores are adjusted to take into account the Gaussian variance differences between models.

Version 2.2

New feature: Spectrogram Slice View

Added a new spectrogram slice plot that can be optionally displayed below the spectrogram plot (with or without the waveform plot). The spectrogram slice plot

graphs the power spectrum of a particular windowed FFT slice. The graph shows the relative power levels across the frequency spectrum represented by the selected FFT slice. The slice plot also shows the estimated power spectrum using Song Scope's recognizer algorithms to help visualize the effects of the "Max Resolution" parameter. To select the FFT slice, hold the shift button while moving the mouse in either the spectrogram or waveform plot.

Version 2.1B

Bug fix: Recognizer unable to open .ssn files

There was a bug in which adding new annotations to an existing recognizer could result in creating a .ssr file that can no longer be edited because the file hierarchy relationship between the .ssr file and dependent .ssn files is no longer correct.

Version 2.1A

Bug fix: Audio playback on Microsoft Vista

Another bug causing hangs was found in the Vista implementation of a Microsoft system library call (in particular waveOutGetPosition()). We have worked around the problem by avoiding the problematic call.

Version 2.1

Bug fix: Audio playback on Microsoft Vista

There was a bug in which audio playback did not work under Vista causing Song Scope to hang. This has now been corrected.

wxWidgets upgrade

Upgraded to wxWidgets 2.8.8 library from 2.8.7. This should fix a number of bugs and improve performance of the graphical user interface.

Version 2.0: Changes from Release 1.12

Compatibility

Version 2.0 can read *.ssr* recognizer files created from older versions of Song Scope. However, recognizer files created in version 2.0 can not be used by older versions.

Algorithm Improvements

In version 2.0, there have been several improvements made to the classification

algorithms. By default, recognizers built in version 2.0 will use the new 2.0 algorithms. However, if you prefer to build recognizers using the old 1.0 version, you can choose 1.0 in the new Algorithm control on the Detector control panel.

To upgrade *.ssr* recognizer files from older versions, open the file, change the Algorithm control on the detector control panel to 2.0, then regenerate and save the new recognizer. You may want experiment a little for best results. In particular, the new algorithms seem to work best with a smaller Max Resolution value in the 4-6 range. You may also be able to increase the Dynamic Range a little for better results.

Quality score improvements

In version 2.0, quality values are now in the range 0-100. A quality value of 100 indicates that the duration (time), density (symbols), and complexity (states) of a vocalization are average compared to the training data. Lower quality values indicate a greater distance from the mean compared to the training data. A minimum quality value of 50 should allow approximately half of the training data through, and a minimum quality value of 10 should allow approximately 90% of the training data through. Quality values below 10 are outliers that are less likely to be a positive match.

History

Version 2.0

Bug fix: Windows audio playback

On some Windows platforms, playback sound was choppy. This has been fixed with improved buffering.

Viewing Spectrograms

Overview

At the core of Song Scope is a spectrogram viewer that lets you see a visual representation of the audio data. For a quick tutorial, see Signal Processing Basics.

Song Sleuth can show three different kinds of plots:

First, the Waveform Plot displays a visual representation of the audio signal in what is known as the "time domain", meaning that it shows the amplitude of the sound pressure wave as it vibrates over time across the microphone.

Second, the Spectrogram Plot displays a visual representation of the audio signal in what is known as the "frequency domain", meaning that it shows the relative power levels of the different frequency components of the sound wave over time.

Third, the Spectrogram Slice Plot displays a visual representation of the power spectrum for a particular time slice of the spectrogram.

Signal Processing Basics

Song Scope is designed to be easy to use for field biologists and others to automate the process of finding specific kinds of vocalizations in long field recordings. While Song Scope implements sophisticated algorithms for efficient signal detection and pattern recognition, you don't need to have a strong background in mathematics, statistics, and digital signal processing to use Song Scope. That being said, there are a few fundamentals you should understand.

The Nature of Sound

Sound is a wave of pressure vibrations created by vibrating objects that propagates through air, water, and other materials. The speed of sound through air at sea level is approximately 340.29 meters (1116.4 feet) per second. In open air, sound propagates away from the vibrating source in all directions creating an expanding wave of air pressure vibrations in the shape of the surface of a sphere.

As the wave expands away from the source, the amplitude of the vibrating pressure decreases with the square of the distance. In other words, a sound from 10 feet away will have 1/100th the amplitude of the sound from only 1 foot away, and a sound from 100 feet away will have 1/10000th the amplitude of the sound from only 1 foot away.

One of the remarkable qualities of our ears and brain is that our perception of volume is "logarithmic" with changes in pressure. We perceive a sound from 10 feet away as having 1/10th the volume as a sound from 1 foot away, even though the pressure amplitude is only 1/100th as much. This enables us to hear quiet/distant sounds as well as loud/close sounds. Or to put it another way, the human ear has a very large dynamic range.

A microphone measures changes in air pressure and produces an electric voltage in proportion to the pressure levels. Unlike the human ear, microphones have a limited dynamic range. Very sensitive microphones can pick up small changes in pressure (e.g. form distant or quiet sound sources), but would be overwhelmed by a close/loud sound.

Frequency

Frequency represents the "pitch" of a sound and is measured as the rate at which the air pressure vibrates. Faster vibrations are higher pitched. For example, middle C on a musical scale has a vibration frequency of around 262 cycles per second (or Hertz, Hz). An octave above middle C will have twice the frequency (524 Hz), and an octave below middle C will have half the frequency (131Hz). Some very high-pitched birds sing at frequencies of 8,000 to 10,000Hz.

Digitizing Sound

To digitize sound, the signal pressure level of a sound wave is measured (sampled) thousands of times per second resulting in a sequence of numerical values corresponding to the pressure level as it changes through time. The rate at which sound is sampled is called the sample rate, usually measured in samples per second, or Hertz (Hz).

In order to detect high frequencies, a faster sample rate must be used. In fact, the sample rate must be at least twice as fast as the highest frequency. In other words, a sample rate of at least 524Hz would be needed (and in reality, just a little more) in order to detect middle C (262Hz), and a sample rate of at least 20,000Hz would be needed to detect some very high-pitched birds singing at 10,000Hz. Music CDs are sampled at 44,100Hz.

Song Scope uses 16-bits to represent each sample of sound as an integer between -32,768 and +32767. These values are proportional to the changes in sound pressure levels measured by a microphone.

Time Domain

The Waveform Plot is used to display sound waves in the "time domain" and shows the actual sample values corresponding to changing sound pressure levels through time.

Frequency Domain

The Spectrogram Plot is used to display sound waves in the "frequency domain" and shows the individual frequency components that make up sound as the sound changes through time. For example, a sustained middle C would show up as a horizontal line through time at 262Hz.

This simple visualization is calculated from the time domain sound data using an algorithm called the Fast Fourier Transform, or FFT. The FFT is a fundamental concept in digital signal processing.

Fast Fourier Transform (FFT)

The FFT transforms a time domain signal into the frequency domain. The input to the FFT is a "window" of N time domain samples, where N is a power of two. The output of the FFT is used to compute the power spectrum of the window represented as N/2 "frequency bins". The bins are evenly spaced and represent frequencies between zero and half the sample rate.

For example, consider a time domain signal sampled at 44,100 samples per second and an FFT window size N of 256 (2 to the power of 8). The output of the FFT can be used to calculate 128 frequency bins between 0Hz and 22,050Hz (with a resolution of 22,050/128 = 172.26Hz between each bin). The relative power level in each bin is typically represented on a log scale measured in decibels. Notice that a single FFT window of 256 samples divided by 44,100 samples per second represents a slice of time equal to 256/44,100 = 5.80 milliseconds.

A spectrogram can be generated by computing a series of FFTs. The output of each FFT represents the frequency power levels over a narrow slice of time. The series of slices can be used to show how the frequency components of a signal change through time. It is also common practice for these FFT windows to be overlapping and averaged together to smooth out edge transitions. Using the example above, if there is a 50% overlap in FFT windows, then there will be twice as many slices in the spectrogram with a new slice every 2.90 milliseconds. If there is a 75% overlap in FFT windows, then there will be four times as many slices in the spectrogram with a new slice every 1.45 milliseconds.

Notice that there is a trade-off between the resolution of frequency and time. Larger FFT windows can resolve more frequencies (more frequency bins), but are wider in time (more samples), and thus have lower resolution in time. Shorter FFT windows are shorter in time (fewer samples) and can therefore observe faster changes in time, but can't resolve as many frequencies.

Beam Forming

An array of multiple microphones can be used to focus sounds from a particular direction. Depending on direction, a sound wave will propagate across each of the microphones at different times. The greater the distance between microphones, the longer it takes for the sound to travel from the nearest microphone to the furthest microphone.

If a sound wave arrives at two microphones at exactly the same time, and these waves are added together, the combination of waves is said to be "constructive" because the sound waves are in phase with each other. The amplitude of such a combined pair of waves will be twice the amplitude of any of the individual waves.

On the other hand, if sound waves arrive at two microphones at different times and added together, the combination of these out-of-phase waves results in "destructive interference". If the waves are off by just the right amount (half a wavelength), they cancel each other out.

Beam forming is a technique in which the signals received from two or more microphones are delayed by different amounts and then added together such that

the audio signals from a particular direction are all in phase and combine constructively resulting in a stronger signal, while sounds from other directions experience destructive interference and are reduced in amplitude.

Song Scope's beam-forming algorithms will automatically adjust the delays of multiple audio channels to maximize a selected signal's amplitude.

Opening and Viewing Audio Files

Supported Audio Formats

At this time, Song Scope only supports 16-bit PCM *.wav* files and Audio IFF (*.aif/.aiff*) files as well as Wildlife Acoustics Audio Compression format (*.wac*) files at a variety of sampling rates. In addition, Song Scope can open Anabat *.??#* files which are converted from zero-crossing information to 16-bit samples internally. You will need to have recordings in these formats to use Song Scope.

Opening Audio Files

From the **"File"** menu, select **"Open..."**. This will pop up a file "chooser" window from which you can select one or more *.wav*, *.aif*, *.??#* or *.wac* files to open. Each selected file will then be opened in its own Song Scope window.

Linux users can also invoke Song Scope from the command line with a list of files as arguments. A window will be opened for each file specified.

In addition to opening the *.wav*, *.aif*, *.??#* or *.wac* file, Song Scope will look for a Song Scope annotation (*.ssn*) file in a *SongScopeNotes* subdirectory and, if present, will load the annotations.

In addition to *.wav* files, Song Scope can also open *.ssn* annotation files directly.

Navigating Between Windows

If you have many windows open, you can use the **"Go"** menu to quickly navigate to a specific window and bring it to the top. MacOS users can use the standard "Window" menu instead.

Viewing Audio Files

Each audio file window consists of 3 major parts, from top to bottom:

- The Waveform Plot displays a "time domain" representation of the audio signal.
- The Spectrogram Plot displays a "frequency domain" representation of the audio signal.
- The control panel contains several folder tabs, each with a collection of controls including Display Controls, Mixer Controls, Spectrogram Controls, Detector Controls, and Recognizer Controls.

- There is also an optional Spectrogram Slice Plot that may be inserted between the Spectrogram Plot and the control panel. This plot displays the power spectrum of a particular time slice of the Spectrogram Plot.

View Controls

The following controls can be found just below the waveform and/or spectrogram plots and above the control panel, from left to right, as follows:

 **Toggle Control Panel**

Toggle between showing and hiding the control panel.

 **Split View**

Show both the waveform and spectrogram plots.

 **Spectrogram View**

Show just the spectrogram plot and hide the waveform plot

 **Waveform View**

Show just the waveform plot and hide the spectrogram plot

 **Split Spectrogram/Slice View**

Shows just the spectrogram plot with the spectrogram slice plot

 **Split Waveform/Spectrogram/Slice View**

Shows the waveform, spectrogram and spectrogram slice plots

 **Play Audio**

Begin audio playback through the computer's sound hardware starting at the current display position and continuing until the end of the recording, or stopped (see below). You can also simply press the space bar to begin playback.

 **Stop Audio**

Stop audio playback (see above). You can also simply press the space bar to stop playback.

 **Zoom Out**

Zoom out on the time (horizontal) axis. Each time this button is pressed, the displayed portion is reduced by a factor of two.

Zoom Slider

The zoom slider adjusts the time (horizontal) axis magnification. Moving the slider to the left zooms out and moving the slider to the right zooms in.

 **Zoom In**

Zoom in on the time (horizontal) axis. Each time this button is pressed, the display portion is magnified by a factor of two.

 **Go Back**

Go to the previous annotation.

Scroll Bar

A horizontal scroll bar can be found between the plots and the control panel that lets you scroll through time to view different parts of the audio file.

 **Go Forward**

Go to the next annotation.

Waveform Plot

The waveform plot displays a time-domain representation of the audio signal. The horizontal axis represents time while the vertical axis represents the relative sound pressure level.

The horizontal ruler indicates time as measured from the beginning of the recording in seconds. Longer recordings will show minutes and seconds in MM:SS format, and still longer recordings will show hours, minutes and seconds in HH:MM:SS format.

Song Scope can display the waveform plot in one of three different ways as follows:

 **Linear Scale**

The linear scale is displayed when a sound file is first opened and shows the relative sound pressure levels as represented by the 16-bit audio stream. The values are between -32768 and +32767.

 **Logarithmic Scale**

The log scale shows the relative sound pressure levels in decibels which is $20 \log_{10}(|x|)$.

 **Logarithmic Scale with Signal Detection**

The log scale with signal detection shows the same relative sound pressure level in decibels as above, but also color codes the results of Song Scope's signal detection algorithm. Different colors are used to show syllables, inter-syllable gaps, and silent intervals between vocalizations. For more information, see Viewing Song Scope Signal Detection

You can move the mouse pointer over the waveform plot to view the corresponding time and relative sound pressure levels as shown in the status bar at the bottom of the Song Scope window.

You can also hold down the left mouse button while moving the mouse to view the difference in time and relative sound pressure levels in the status bar. When the left mouse button is released, the time portion of the signal will be selected.

Spectrogram Plot

The spectrogram plot displays a frequency-domain representation of the audio signal. The horizontal axis represents time while the vertical axis represents frequency (measured in Hz, or cycles per second).

The horizontal ruler indicates time as measured from the beginning of the recording in seconds. Longer recordings will show minutes and seconds in MM:SS format, and still longer recordings will show hours, minutes and seconds in HH:MM:SS format.

Different colors are used to represent the relative signal power levels (on a logarithmic scale) for different frequency components. The color bar at the bottom of the spectrogram plot indicates which colors are used to represent the relative power levels as measured in decibels. The choice of colors can be changed by using the Display Controls.

Song Scope can display the frequency plot in one of three different ways as follows:

 **Linear Scale**

The linear scale is displayed when a sound file is first opened and shows the frequency in Hz.

 **Logarithmic Scale**

The logarithmic scale shows the same frequencies as in the linear scale above, but stretches out the range on a log scale such that low frequencies are spread out and high frequencies are closer together. The log scale is more representative of how sound is heard and is used internally by Song Scope when comparing sounds. The log scale can be changed by using the Spectrogram Controls.

 **Logarithmic Scale with Signal Normalization**

The log scale with normalization shows the same frequency scale as above, but the power levels are "normalized" to illustrate how Song Scope will interpret the signal for pattern matching. Song Scope's signal detection algorithms are used to determine if a given time slice contains signal or not. Power levels are only shown for detected signal.

You can move the mouse pointer over the spectrogram plot to view the corresponding time, frequency and relative power levels as shown in the status bar at the bottom of the Song Scope window.

You can also hold down the left mouse button while moving the mouse to view the difference in time and frequency in the status bar. When the left mouse button is released, the time and frequency portion of the signal will be selected.

Spectrogram Slice Plot

The spectrogram slice plot displays the power spectrum of one moment in time from the spectrogram plot. The horizontal axis represents frequency (measured in Hz, or cycles per second) while the vertical axis represents relative power (measured in decibels). The power spectrum is normalized such that the strongest frequency present (within the selected range) is set to 0dB.

You can select which time slice is displayed by moving the mouse pointer along the time axis of either the Spectrogram Plot or the Waveform Plot while depressing the "Shift" key.

Different colors are used as follows:

Outside Selected Frequency Range

Portions of the spectrum that are outside of the frequency range as selected by the Frequency Min and Frequence Range parameters are shown in colors used to represent relatively weak signals in the spectrogram (a dark blue by default).

Inside Selected Frequency Range

Portions of the spectrum that are within the frequency range as selected by the Frequency Min and Frequency Range parameters are shown in colors used to represent medium signals on the spectrogram (a green by default).

Spectrum Estimation

A power spectrum estimation line is drawn on the power spectrum graph in a color used to represent strong signals on the spectrogram (a red by default). Song Scope recognizers employ feature reduction techniques to simplify the power spectrum. The power spectrum estimate is driven largely by the Max Resolution recognizer parameter. The effect of changing this parameter can be visualized by adjusting the slider located in the lower left corner of the spectrogram slice plot. Larger values will result in an estimation that is closer to the actual power spectrum. Too much detail can result in poor recognizer performance because it is trying to match to specific details of a particular recording or individual. Too little detail results in very little distinction and can result in a higher false positive rate with other noise sources in similar frequency ranges.

Selecting Audio Samples

You can select a portion of audio samples in the time and frequency dimensions by using the mouse pointer in either the waveform or spectrogram plots. First, move the mouse to the corner of a rectangle you wish to select. The status bar at the bottom of the Song Scope window will display the precise position of the cursor. Next, press and hold the left mouse button and move the mouse to the opposite corner. The status bar will display the precise range currently selected. When the mouse button is released, the selection will be made as shown by a solid box in both the waveform and spectrogram plots. This selection remains in effect until a new selection is made. Note that a double-click of the left mouse button will clear the selection.

Selections made in the waveform plot will select all displayed frequencies in the spectrogram plot (bounded by the frequency range specified in the spectrogram controls. Selections made in the spectrogram plot will select only those specified frequencies.

After a selection is made, you can right click (MacOS users can control-click instead) inside the selection to display a pop-up menu with the following options:

Play Selection

Listen to the selected time and frequency portion of the audio recording through the computer's sound hardware. Note that you can also simply press the space bar to play the current selection.

Focus Beam

This option is only available when working with multi-channel audio recordings. Automatically enables multi-channel mixing and adjusts per-channel delay controls to "focus" the microphone array using beam forming.

Adjust Levels

Automatically adjust the brightness and contrast display controls such that the strongest frequency component in the selection is shown with the "brightest" color and the weakest frequency component in the selection is shown with the "darkest" color.

Zoom To Fit

Automatically adjust the horizontal (time axis) zoom such that the selection completely fills the waveform and/or spectrogram plots.

Annotate...

Pops up a dialog window so that you can annotate the selection.

Annotate as *class*:*subclass*:*id*

Annotate the selection repeating the most recent annotation with the class, subclass and id indicated.

Save as...

Save the selected audio portion as a *.wav* file.

An alternative method for selecting vocalizations is to use the Logarithmic Scale with Signal Detection mode of the Waveform Plot. In this mode, you can simply right click on a detected vocalization. The vocalization will be automatically selected and the above pop-up menu will appear.

Display Controls

The display tab of the control panel contains the following controls used to adjust the appearance of the spectrogram and waveform plots as follows:

Brightness

The brightness control adjusts the signal levels shown in the spectrogram plot by shifting all the power levels up or down the color scale.

Contrast

The contrast control adjusts the visible dynamic range from the lowest to highest power levels as represented by the color bar. When a signal is below the range, it is displayed with the "darkest" color. Increasing the contrast decreases the dynamic range so that only stronger signals will be visible, while decreasing the contrast increases the dynamic range so that weaker signals will become visible.

Hue

Colors are represented in a circular "wheel" that rotates through magenta, blue, cyan, green, yellow, red, and back to magenta. By default, the color bar shows the dynamic range (as determined by the contrast control) with violet representing the weakest signals and red representing the strongest signals. The hue control can be used to rotate the colors of the color bar to display a different range of colors to represent the relative strength of different frequency components in the spectrogram plot. The waveform plot always shows a background color corresponding to the weakest signal and a foreground color corresponding to the strongest signal.

Saturation

The saturation control adjusts the display from a gray-scale (from black to white) to full color.

Luminosity

In addition to color hues (with full saturation), power levels can also be represented from dark (black) to full color (red, as adjusted by the hue control). The luminosity control adjusts the extent to which "darkness" is used to represent power levels. With the lowest luminosity, the weakest signals appear black. With the highest luminosity, the weakest signals appear violet, as adjusted by the hue control. Note that if a gray scale is used, the luminosity control should be set to the lowest setting and duplicates the functionality of the brightness control.

Inverse Video

The inverse video control reverses light and dark from normal settings.

Mixer Controls

The mixer tab of the control panel contains controls for selecting and mixing individual channels of a multi-channel audio file (e.g. a stereo *.wav, .aif* or *.wac* file has two channels) into a combined "mixed" signal. The combined signal is the output of the mixer which is then used to display waveforms and spectrograms, signal detection, and recognition.

For simple applications with only one microphone sensor (single channel), you only need to worry about the sample rate control described below.

For multi-channel applications (more than one microphone channel), beam forming is used to enhance the signal of a particular vocalization by digitally focusing the microphone array in a specific direction. The remaining controls are used for this purpose.

Sample Rate

The sample rate control adjusts the sample rate from the audio source file to a sample rate used by Song Scope to display spectrograms and compare vocalizations.

Playback Speed

The playback speed control adjusts the playback speed from the audio source file when played through your computer's sound system or saved to a *.wav* file. The default value is "Normal" speed. You can choose to playback audio at 1/2, 1/4, 1/8, and 1/16th speed.

Max Sample Delay

The maximum sample delay control specifies the number of source audio sample times it takes for a sound wave to propagate across the microphone array. If the source recording is made at sample rate R, and the greatest distance between any two microphones is D, then this value should be set to at least R * D / S where S is the speed of sound. Note that S is equal to approximately 343.6 meters per second or 1127 feet per second, depending on temperature, barometric pressure, and other variables. The maximum sample delay control is used with Song Scope's built-in beam forming algorithm.

Channel Enable

The per-channel enable check boxes are used to enable and disable the multi-channel mixing. Either only one channel is selected or all channels are selected. If one channel is selected, then this channel is the "reference" channel, and only that

channel is used (other channels are effectively disabled). If all channels are selected, then the multiple channels are combined according to the gain and delay controls described below. The last selected reference channel is used by Song Scope's built-in beam forming algorithm.

Channel Gain

The per-channel gain control adjusts the relative gain of the channels before they are combined by the mixer.

Channel Delay

The per-channel delay control adjusts the sample delay (limited by the maximum sample delay described above). Each channel is delayed by the number of samples specified before being combined by the mixer.

Spectrogram Controls

The spectrogram tab of the control panel contains the following controls used to adjust the spectrograms produced by Song Scope.

For more information, please see Signal Processing Basics.

FFT Size

The FFT size control adjusts the window size (in samples) of the Fast Fourier Transform algorithm used to produce spectrograms.

FFT Overlap

The FFT overlap control adjusts the amount of overlap between FFT windows.

Frequency Minimum

The frequency minimum control adjusts the lowest frequency displayed on the spectrogram and used in comparing vocalizations. The value indicates the lowest FFT bin number.

Frequency Range

The frequency range control adjusts the range of frequencies (as measured from the minimum frequency specified above) displayed on the spectrogram and used in comparing vocalizations. The value indicates the number of FFT bins (up to the maximum).

Amplitude Gain

The amplitude gain control amplifies the input signal on both the spectrogram and waveform plot. This can also be used to adjust the volume during playback through the computer's sound hardware.

Background Filter

The background filter reduces background noise. The algorithm requires an estimate of the background power spectrum. The background filter control can be used to disable the filter, or to specify the number of seconds over which to average background noise levels to be used for the background power spectrum estimate.

Making and Using Annotations

Annotations are manually entered notes about individual vocalizations found in recordings and contain a hierarchy of classification information in addition to arbitrary comments.

Classification Hierarchy

Song Scope defines a hierarchy of classification information included in annotations as follows:

Class

A class is a label used to identify a vocalization at a high level. Typically, this could be the name of a particular species under investigation.

Subclass

A subclass is a label used to divide vocalizations in a particular class down into meaningful groups for classification. This is intended to divide vocalizations into types that can be compared to each other. For example, many species of birds have a number of calls and songs among their repertoire of vocalizations. The subclass could be used to divide vocalizations into flight calls, contact calls, and spring songs, for example.

Id

The id is a label used to divide vocalizations in a particular class down into different variations. For example, this could indicate different individuals, or it could indicate typical regional variations. Song Scope uses the id to measure the quality of different recognizer models as follows: By building a recognizer model with all the vocalizations excluding a particular id, the ability of the recognizer to detect the excluded vocalization is an indication of how well the model performs. For more information see Pattern Recognition Basics. If you do not specify an id, Song Scope will automatically assign a unique id associated with the recording. In other words, by default, Song Scope will assume that vocalizations found in different recordings are likely to be from different individuals while vocalizations on the same recording with the same class are likely to be from the same individual.

Creating a new Annotation

We strongly recommend that you pay attention to the Signal Detection settings and view when making annotations. Annotations should include some of the

"quiet" portions immediately preceding and following the vocalization (in other words, better to crop vocalizations too wide rather than too narrow).

To create a new annotation, select the vocalization to be annotated. Then use the pop-up menu to select the Annotate... command. A dialog will appear in which you can type in the class, subclass, id, and any additional comments you like. The dialog also contains a shortcut menu tree listing all the previously assigned classes, subclasses and ids. You can click on the name of a class, subclass, or id to have fields filled in automatically instead of typing them again each time. Once created, the annotation will appear in the waveform and spectrogram plots as a solid box with the class, subclass, id, and comments displayed near the box.

For multi-channel audio recordings, the mixer per-channel gain and delay settings are saved with the annotation. In other words, if beam forming is used to focus a microphone array in the direction of the vocalization, the mixing parameters are stored with the annotation. The settings for a particular annotation can be restored by left-clicking on an annotation or by selecting the annotation using the forward and backward annotation navigation controls.

Selecting Annotations

To select an annotation, simply right click (MacOS users can use control click instead) in the annotation on either the spectrogram or waveform plots to display a pop-up menu with the following options:

Play Selection

Listen to the annotation through the computer's sound hardware. Notice that you can also simply press the space bar to play the current selection.

Focus Beam

This option is only available when working with multi-channel audio recordings. Automatically enables multi-channel mixing and adjusts per-channel delay controls to "focus" the microphone array using beam forming. If you want to update the mixing parameters of an annotation, first set the desired mixing parameters (e.g. with this "Focus Beam" choice), and then edit the annotation.

Adjust Levels

Automatically adjust the brightness and contrast display controls such that the strongest frequency component in the selection is shown with the "brightest" color and the weakest frequency component in the selection is shown with the "darkest" color.

Zoom To Fit

Automatically adjust the horizontal (time axis) zoom such that the annotation completely fills the waveform and/or spectrogram plots.

Edit...

Pops up a dialog window so that you can edit the annotation. The current mixing parameters will be applied.

Delete

Delete the selected annotation

Save

Save the selected annotation to a *.wav* file

Saving Annotations

Song Scope annotation files are designated with the *.ssn* suffix. By default, these files are stored in a subdirectory called *SongScopeNotes* located in the directory containing the original *.wav*, *.aif*, *.??#* or *.wac* audio file. The *.ssn* file is given the same base filename as the corresponding audio file. Song Scope annotation files contain both the annotation information as well as a copy of the actual audio samples that make up the corresponding vocalization (with any mixing parameters applied as described above). This lets you retain the portions of audio that were interesting enough to annotate without the need to retain the much larger original field recordings.

Note that annotation files only store the mixed audio samples and the class, subclass, id and comments labels. None of the display, spectrogram, or detector parameters are saved in the annotation file (they are instead saved to and loaded from recognizer files).

When a *.wav*, *.aif*, *.??#* or *.wac* recording file is opened, Song Scope looks for a corresponding *.ssn* file in the *SongScopeNotes* subdirectory and automatically loads the previously saved annotations.

From the **"File"** menu, select **"Save Annotations"** to save annotations and their corresponding audio samples to a Song Scope *.ssn* file.

If annotations were created or edited without being saved and the Song Scope window is closed, Song Scope will ask if the annotations should be saved first.

Annotations can also be saved to a different location by using the **"Save Annotations As..."** menu. However, Song Scope will not automatically load annotations when opening the source recording file unless they can be found in

the default location as described above.

Loading and Merging Annotations

Song Scope can load or merge annotations from *.ssn* files by choosing **"Load Annotations..."** from the **"File"** menu. This is helpful to load annotations that were saved somewhere other than the default SongScopeNotes subdirectory. Loaded annotations are merged together.

Opening Annotation Files

Song Scope can also open a *.ssn* file directly to review and playback saved annotations. The annotations are displayed in their original time positions on the spectrogram and waveform plots, with all the empty space between notations filled with background spectrum information.

Creating Recognizers to Detect Vocalizations

Song Scope Recognizers are used to compare vocalizations found in long field recordings against a specific vocalization of interest using patented and sophisticated digital signal processing algorithms developed by Wildlife Acoustics.

Recognizers are built from training data, a collection of recorded vocalizations representative of the vocalization of interest. In other words, you need to start with confirmed recordings of the species you are interested in finding to build a recognizer. First, you annotate these recordings to indicate which vocalizations belong to the class and/or subclass corresponding to the vocalization of interest. Once built, you can then use the recognizer to quickly search long field recordings for matching vocalizations.

From the **"File"** menu, select **"New Recognizer"** to start a new recognizer. Alternatively, you can open a previously generated recognizer *.ssr* file.

There are a number of issues to consider when choosing the parameter settings and training data to use in building a recognizer discussed in the next sections, Pattern Recognition Basics and Signal Detection.

You can then Import and Review Training Data from your annotated recordings, and Generate and Save your recognizer.

Finally, you can use your recognizers to Search Recordings for Vocalizations.

Pattern Recognition Basics

Challenges

Digital pattern recognition by machine is extremely difficult and inefficient when compared with the human brain (and brains from several other species, for that matter). It is easy for us to take for granted our ability to instantaneously recognize sights, sounds, and smells with extremely high accuracy, even in difficult conditions (e.g. in the presence of noise and low light). Digital computers have a very long way to go, despite decades of research and investment in the field of speech recognition and computer vision.

Computers are good with 1's and 0's, and can easily compare bit-by-bit two data sets to tell you if they are an exact match or not. But pattern matching is never an exact match in the real world. The animals we study will never produce exactly the same sound twice (at least, not when comparing bit-by-bit). In addition, the signals we must analyze are corrupted by random noise (from the wind blowing, other animals vocalizing, sound waves bending around trees, etc). The problem of pattern recognition is not a simple binary 1 or 0 problem. Rather, it is "fuzzy".

When considering the analysis of animal vocalizations, another significant factor is the degree to which vocalizations vary. There can be considerable individual and regional variations within a particular species. And even a single individual may be capable of producing a wide range of vocalizations.

Consider a machine that does a bit-by-bit comparison of two vocalizations. Only an exact duplicate would be considered a match, which will never happen given real world variation and random processes. So instead, imagine a machine that "blurs" the data. Now two blurry patterns that are similar could be a more exact match than a bit-by-bit comparison. This is one of the techniques used by most successful digital pattern recognition systems and is known as "feature reduction". The idea here is that some information in the signal is important for identification, and the rest is not. If we could reduce the signal by removing all the elements that did not contribute to identification, then it would be easy to compare the remaining features and determine if they matched. This is easier said than done, however, because the set of features that might help identify one vocalization may be different than for another.

Now lets consider the degree to which a given vocalization may vary. If a vocalization has wide variation among individuals, then pattern matching requires elimination of more features and broader acceptance of patterns. But if we go too far, then many incorrect vocalizations may be falsely identified resulting in undesirable "false positives".

In the face of real world noise, the problem becomes even more difficult. Once

again, the human brain effortlessly separates simultaneous sounds received in our ears from all directions. But to the digital machine, the competing sounds merge to form an ambiguous jumble that is much more difficult to separate.

Finally, consider mimics. There are species of birds, like the Northern Mockingbird, that have so much song-to-song variation that it is impossible to rely on any direct pattern matching to detect their vocalizations (though humans can easily detect the variation and repetition of syllables to quickly identify the mockingbird).

For all the reasons cited above, it is impossible to build a perfect classifier capable of identifying each and every occurrence of a particular vocalization with 100% accuracy. That being said, in the face of these challenges, the Wildlife Acoustics algorithms perform remarkably well, especially with the careful selection of training data and parameter settings.

Feature Reduction in Song Scope

As described above, feature reduction is an important aspect of digital pattern matching. Song Scope incorporates a number of feature reduction techniques that can be adjusted for particular vocalizations as follows:

Background Noise Reduction

Song Scope makes use of a background noise filter used to reduce background noise levels. In addition to reducing background noise, the vocalization spectrogram is also sharpened by reducing smearing effect of echos. Since background information has nothing to do with the signal of interest, reduction or elimination of background noise is a good place to start in feature reduction.

Frequency Band Limiting

Most vocalizations can be described as occurring within a finite range of frequencies. (At faster sampling rates, many harmonics of the fundamental frequency may be detected, but these harmonics generally only add redundant information to the underlying signal and can therefore be eliminated as part of feature reduction). The frequency range can be specified using the Frequency Minimum and Frequency Range controls in the Spectrogram Control panel. The signal detector also makes use of frequency band limiting for finding candidate vocalizations.

Sample Rate

Related to the frequency band limiting above, a sampling rate of twice the maximum frequency is all that is needed to resolve the band-limited frequency range. Higher sampling rates will only add additional processing overhead. The

sample rate can be set by using the Sample Rate Control on the Mixer Control panel.

Log Frequency Bins

Song Scope uses a log frequency scale in recognizers because the log frequency scale compresses high frequency information. It is common for the high frequencies to contain less or redundant information (such as harmonics of lower frequencies) compared with lower frequencies. The effect of using a log frequency scale can be observed visually by viewing the spectrogram plot on a log frequency scale.

Power Normalization

Song Scope normalizes the power spectrum in each FFT slice to reduce spectral features to a small dynamic range. This has the added benefit of eliminating noise and competing audio signals that are not within the dynamic range. The dynamic range can be configured by the Dynamic Range control on the Detector Control Panel. The effect of normalizing power can be observed visually by viewing the spectrogram plot on a long frequency scale with power normalization.

Dimension Reduction

Song Scope reduces the frequency bins of each FFT slice further to a "feature vector" consisting of a small number of dimensions. The size of the feature vector can be configured by the Maximum Resolution control in the recognizer control panel.

Training Data

One of the most important factors in building an accurate recognizer is selecting the training data. For vocalizations that are particularly consistent with little variation, less training data will do the job. But for vocalizations with significant variation, training data will be needed that covers a range of these variations. Occasionally, a new variation may be encountered that doesn't fit the model and will fail to be recognized. The good news is that Song Scope makes it easy to then incorporate this new variation into the model so that it will be more easily recognized the next time a similar variation is encountered.

In pattern recognition, there is a phenomenon known as "over training". When there is very little training data available, statistical models will tend to retain too much detail such that the small amount of training data fits perfectly, but even small variations are rejected. To prevent over training, Song Scope builds and tests several different models and uses the annotation id to identify the different variations available in the training data. For each annotation id, Song Scope will build a model with all the training data excluding those specific vocalizations

marked with the id, and then tests the model against the excluded vocalizations. In this way, Song Scope attempts to maximize the generalization of the model to cope with previously unobserved data. For this to work, it is important to have at least a few examples of different variations and mark the annotation id field appropriately.

When different vocalization subclasses are distinct and unlike each other at all (e.g. the difference between nasal call notes and whistled songs in many species), it would be better to divide these into different recognizers (i.e. one recognizer per subclass) in order to avoid overly complicating a model with two or more completely unrelated sets of data. On the other hand, if different vocalization subclasses share many common elements with each other, it may be better to combine them into a single recognizer model in order to have more training data available to capture the variation accurately. For a given vocalization of a given species, different combinations may produce better results than others. Song Scope makes it easy to build different recognizers with different permutations of training data so you can easily and quickly try different combinations until satisfactory results are achieved.

Signal Detection

Signal detection is used to find the candidate vocalizations within a recording. These candidate vocalizations can then be compared against vocalizations of interest using recognizers. The goal of signal detection is to locate candidate vocalizations in potentially noisy recordings and determine where the vocalization begins and ends.

A typical vocalization consists of a series of "syllables" tightly grouped together into a "song" spanning only a few seconds in total duration or less. Syllables are usually only a fraction of a second in duration, and the inter-syllable gaps between syllables are also only a fraction of a second in duration. Sometimes the inter-syllable gap is so small that multiple syllables may appear to merge together.

Viewing Song Scope Signal Detection

To see a visual representation of Song Scope's signal detection at work, you can display the waveform plot using Logarithmic Scale with Signal Detection. This view shows the total power levels within the selected frequency range as set by the Frequency Minimum and Frequency Range spectrogram controls. Different colors are used to indicate the results of Song Scope's signal detection algorithm:

No Signal

Portions of the recording that are not considered to be part of a candidate vocalization are shown in colors used to represent relatively weak signals on the spectrogram (a dark blue by default).

Syllable

Portions of the recording considered to be part of a syllable within a candidate vocalization are shown in colors used to represent the strongest signals on the spectrogram (red by default)

Inter-syllable Gap

Portions of the recording considered to be part of an obvious inter-syllable gap are shown in colors used to represent medium signals on the spectrogram (a green by default).

Soft Gap

Portions of the recording considered to represent a likely inter-syllable gap where two syllables merge together are represented with a color used to represent signals between medium and strong (yellow by default).

Note that recordings with weak signals or low noise may not register any signal above 0dB for color-coding. You can increase the gain by adjusting the Gain Control to bring these signals into view.

Also note that you can adjust the display colors using the Display Controls

Signal Detection Controls

The following controls can be used to tune the signal detection algorithm for optimum results when looking for a particular candidate vocalization:

Frequency Band

The signal detection algorithm looks at total power in the visible frequency band as determined by the Frequency Minimum and Frequency Range controls in the Spectrogram Control Panel. In a typical audio recording, there is significant noise in lower frequencies that make it difficult to accurately detect signals unless they are filtered out. You should therefore adjust the minimum frequency as high as possible without clipping the lowest frequency component of the vocalization of interest. By limiting the upper frequencies, you can avoid triggering the signal detection when vocalizations of a higher frequency animal are heard.

Max Syllable

The Max Syllable control on the detector control panel is used to specify the largest syllable likely to be encountered in the vocalization.

Max Syllable Gap

The Max Syllable Gap control on the detector control panel is used to specify the largest inter-syllable gap likely to be encountered in the vocalization. If a "quiet" interval exceeds this value, then the Song Scope detector will mark the end of the vocalization. If this value is too small, the detector may not group all the syllables of a song together. If this value is too large, it is possible that a second vocalization from a different and unrelated individual may be incorrectly joined with the vocalization of interest.

Max Song

The Max Song control on the detector control panel is used to specify the largest vocalization likely to be encountered in the vocalization.

Dynamic Range

The Dynamic Range control on the detector control panel is used to limit the dynamic range of the relative power levels in the vocalization as part of Feature

Reduction. In addition, dynamic range plays a role in signal detection in cutting off weaker vocalizations in favor of selecting stronger ones for candidates. If this value is too low, there will not be enough information to detect important elements of the vocalization for accurate recognition. If this value is too high, the signal detector and recognizers will be more susceptible to background noise. The optimum value for the dynamic range control is the expected signal to noise ratio of the field recordings. That is, the difference in decibels between the typical background noise and candidate vocalizations within the specified frequency band.

Algorithm

The Algorithm control on the detector control panel is used to select between the new and improved version 2.0 classification algorithms and the older version 1.0 algorithms (for backward compatibility with older recognizers or in case the older algorithms perform better for some applications).

Importing and Reviewing Training Data

To start a new recognizer, use the **"File"** menu and select **"New Recognizer"**. Alternatively, you can also open a previously generated recognizer for editing by opening the Song Scope Recognizer file (*.ssr*) directly by selecting **"Open..."** from the **"File"** menu.

The recognizer window looks just like the windows used to view audio files, except that the waveform and spectrogram plots are pushed over to the right side of the window to make room for a recognizer control panel on the left side.

Use the **"File"** menu and select **"Import Notations"** to import files containing training data. These can be the *.wav* audio files that you have already annotated, or you can open the corresponding Song Scope Notation files (*.ssn*) in the *SongScopeNotes* subdirectory directly.

The recognizer control panel shows each imported vocalization sorted by class and subclass in a tree structure. By default, each new loaded vocalization is selected as indicated by a check box next to the vocalization line. Each vocalization indicates the source recording file from which it originated, the time index and duration of the vocalization, and the Id as specified when making the annotation, or automatically assigned uniquely to each recording by Song Scope.

You can click on each individual vocalization line to view the vocalization in the spectrogram and/or waveform plots. Note that by default, the recognizer displays the waveform plot using Logarithmic Scale with Signal Detection, and the spectrogram plot using Logarithmic Scale with Signal Normalization. These views are important because they reflect how Song Scope will "see" the visualizations for building models.

You can double click a vocalization line to toggle between selecting and unselecting it for inclusion in generating a recognizer. You can also double click a class or subclass (or "All Classes") to select or unselect all of the vocalizations contained within.

You can right click (or Command-C for Mac OS X) on the tree to copy the annotation list to the clipboard, and then paste it into a spreadsheet or text file. This feature is handy for building reports to list the training data used.

You should adjust the settings described in the sections on Feature Reduction and Signal Detection Controls for optimum results and review each of the included vocalizations to make sure they are representative of the vocalization you are interested in and are not corrupted by noise that could contaminate the recognizer.

Generating and Saving Recognizer

Additional Controls

There are two more controls that determine how the recognizer will be constructed:

Maximum Complexity

The Maximum Complexity control limits the size of the recognizer to the specified number of "states". If the training data is highly varied with vocalizations consisting of many syllable types, more complexity (and more training data) may be required to accurately model the vocalization. For readers with more experience in pattern recognition techniques, Song Scope makes use of Hidden Markov Models, and this control limits the number of states used to generate a model for the vocalization.

Maximum Resolution

The Maximum Resolution control limits the size of spectral "feature vectors" as described in Dimension Reduction. Many bird vocalizations are "narrow band", meaning they have tight spectral components representing "whistle-like" sounds. These vocalizations are not particularly complex, and a feature vector of only 6 or so dimensions often provides sufficient spectral resolution. On the other hand, vocalizations rich in spectral complexity may require more dimensions to represent them accurately. You should also be aware that low quality (e.g. open microphone) recordings may require a lower resolution to match the "fuzzier" spectral resolution, while a higher resolution may be more suitable for higher quality (e.g. parabolic or otherwise very high signal-to-noise ratio) recordings.

The Spectrogram Slice Plot displays the power spectrum estimate for a given time slice. This can be used to visualize the effect of Max Resolution on spectrum estimation (there is a slider in the plot that can be used to try different values for the spectrum estimation).

Generation

Once you are satisfied with the selection of training data and parameter settings, just press the **"Generate Recognizer"** button. Song Scope will then begin building several permutations of models (based on trying different numbers of syllable types from simple to more complex models). For each model, and for each annotation id, Song Scope will build the model excluding vocalizations marked with the specific id, and then test the performance of the model against the excluded vocalizations. This process can take quite a bit of time if you have a

lot of training data and need to build very complex models. On even fairly fast machines, it may take 30 minutes to an hour to build some models. Fortunately, this is not something you will need to do very often.

Results

When the recognizer completes, the "Recognizer Information" section of the recognizer control panel will display information about the generated recognizer as shown below. The most important is the cross training result as a measure of how well the model is expected to perform. Some of the results are related to details of the algorithms and should not be of any consequence to most users.

Cross Training:

Cross training shows the average and standard deviation of the "fit" of excluded annotation ids when building the model. A low score may indicate that the generated model may not accurately represent the vocalization. If this is the case, a more complex model may be required (by adjusting the maximum complexity setting), more training data may be needed, or the vocalization may need to be split into subclasses.

Total Training:

Total training shows the average and standard deviation of the "fit" of all the training data in the final model which includes all of the training data. It will typically show a slightly higher score and slightly smaller standard deviation than the cross training result described above.

Model States:

Indicates the size of the model as a number of states.

Feature Vector:

Indicates the number of dimensions in the spectral feature vectors, the same as the Maximum Resolution control setting.

Syllable Types:

Syllable types indicates the number of different syllable classes that were used to construct the final model. Song Scope tries different values up to 1/4th of the maximum model complexity and chooses the value that scored highest during cross training.

State Usage:

Indicates the average and standard deviation in the number of different states traversed by each vocalization

Mean Symbols:

Indicates the average and standard deviation of the number of symbols contained within each vocalization

Mean Duration:

Indicates the average and standard deviation of the duration of each vocalization.

Saving the Model

From the **"File"** menu, select **"Save..."** to save the Song Scope Recognizer to a *.ssr* file. The filename will be used as the name of the recognizer.

How to Build Good Recognizers

Choosing an appropriate sample rate

See the Sample Rate control on the mixer control panel.

You should first decide what sample rate is most appropriate for your application. While many recordings are made at CD-audio quality 44,100 samples per second, this is often **not** the best choice.

The sample rate should be at least twice the frequency of the highest dominant frequency in a vocalization. However, we recommend that you choose a sample rate that is not much higher than this. While faster sampling rates can resolve high frequency harmonics of a vocalization, these harmonics usually offer only redundant information and are not particularly helpful in identification. In addition, the limited dynamic range available in noisy field environments render higher frequency harmonics undetectable compared to high-quality recordings made by directional microphones under ideal conditions. In addition, given the limited frequency resolution of a given FFT size, higher sampling rates result in less frequency detail in the lower frequencies that may in fact be more important for recognition.

There are computational advantages to choosing a sample rate that is an integer factor of the source sample rate. In other words, a recording made at 44,100 samples per second can be more efficiently reduced to 22,050 (divided by 2) or 11,025 (divided by 4) rather than being converted to 16,000 samples per second.

For amphibian monitoring, most frogs vocalize under 4,000Hz, so sampling rates over 8,000Hz are recommended. If the source recording is sampled at 44,100 samples per second, we would recommend using a sample rate of 11,025 samples per second.

For birds, most species vocalize well below 10,000Hz, so a sample rate of 22,050 samples per second is sufficient. That being said, many birds such as owls and doves have vocalizations under 1,000Hz, so sampling rates of only 2,000Hz would be acceptable for these species.

Choosing an appropriate FFT size

See the FFT Size and FFT Overlap controls on the spectrogram control panel.

After adjusting the sample rate as described above, you should next choose the optimum FFT parameters. The best way to do this is by viewing the spectrogram of a specific vocalization and see how changing the FFT size affects the

spectrogram plot. Larger FFT sizes will show more frequency resolution at the expense of detail on the time axis while smaller FFT sizes will show more detail on the time axis at the expense of frequency resolution. For example, for vocalizations with a rapid pulsing "trill", smaller FFT sizes may be better to resolve the individual pulses.

Choosing an appropriate frequency band

See the Frequency Minimum and Frequency Range controls on the spectrogram control panel.

After adjusting the FFT parameters as described above, you should next choose the optimum minimum frequency. We recommend that you use the Logarithmic Scale view of the spectrogram plot because the minimum frequency plays a very important role in determining the log frequency scale used by the recognizer.

It is also important to understand that background noise is generally stronger in lower frequencies and will corrupt a signal making it difficult to recognize accurately.

We recommend that you adjust the minimum frequency as high as possible and just below the lowest frequency component of the vocalization of interest. It is best to do this while observing vocalizations in the spectrogram plot using the logarithmic scale view.

After setting the minimum frequency, you can then adjust the frequency range to just above the highest frequency component of the vocalization. The combination of these two controls sets the range of frequencies that the recognizer will consider effectively eliminating background noise sources in lower frequencies or competing signals (from other species) in higher frequencies.

Choosing an appropriate background filter

See the Background Filter control on the spectrogram control panel.

We recommend that you **always** enable the background filter. A setting of one second is best for most applications.

Choosing appropriate signal detection parameters

The most important parameter in signal detection is the Dynamic Range control on the detector control panel. The dynamic range control sets a limit on how much of the signal energy (in decibels measured relative to the peak signal) will be used in comparing waveforms.

The dynamic range should be matched with the expected signal-to-noise ratio of the field recordings to be analyzed. If the dynamic range is set too high (e.g. much higher than the signal-to-noise ratio), then Song Scope will be looking for spectral details that are lost in the noise in actual field recordings resulting in poor recognition performance. On the other hand, if the dynamic range is set too low, spectral details important to accurate classification may not be considered.

The dynamic range setting is used in conjunction with the other signal detection controls to classify portions of a signal into syllables and inter-syllable gaps in a song. Using the Logarithmic Scale with Signal Detection view of the waveform plot, you can visually see the effects of changing these controls while viewing a specific vocalization. We recommend that you use this view and adjust the settings appropriately while keeping in mind that the dynamic range should also be related to recordings made under actual field conditions.

The dynamic range setting determines how much of the signal's frequency components are considered. The effects can be seen by using the Logarithmic Scale with Signal Normalization view of the spectrogram plot.

Training Data

We recommend that you use several different recordings for training data representative of the different variations common in a particular vocalization.

Compensating for different quality recordings

It is common to have access to very high quality recordings for training data, such as those made with parabolic reflectors. However, open microphone field recordings, such as those made by unattended field recorders like Song Meter, will sound different. In addition to the dynamic range paramter discussed above, you may also want to reduce the model resolution by lowering the Maximum Resolution value. In our experience, a value of 8 is good in these conditions. In addition, you may find that increasing the Maximum Syllable and Maximum Syllable Gap will help. We recommend setting the detection parameters to work best for the recordings you wish to scan, rather than for the higher quality training data. A little trial and error goes a long way to find the best parameters for the job.

How to tell if the generated model is any good

When the model is built, pay attention to the "Cross training" percentage and standard deviation displayed in the "Recognizer Information" window. A low score (e.g. < 50%) or a large standard deviation (e.g. > 15%) may indicate that the generated model is not expected to perfrom well. In this case, you may wish to try higher values for the Maximum Complexity or different values (larger and smaller) for the Maximum Resolution parameters. You may also try breaking up

the training data into smaller subclasses.

Beware that a high-scoring model is not necessarily a good discriminating model as it might simply be matching things too easily and could result in high scores (and false positives) for incorrect vocalizations as well.

Searching Recordings for Vocalizations

First, generate a recognizer *.ssr* file representing the specific vocalization type you are searching for.

When viewing an audio file, you can load one or more recognizers. From the **"File"** menu, select **"Load Recognizers..."**. Select one or more recognizers you would like to load.

The loaded recognizers will appear on the "Recognizers" tab of the control panel with their cross training results shown. Note that recognizers will be grouped together into "Recognizer Groups" based on the control settings that were used to generate the recognizer. This is because these same settings must be used in the current recording in order for the recognizers to make "apples-to-apples" comparisons. Recognizers in the same group share the following parameter values in common:

- Sample Rate
- FFT Size
- FFT Overlap
- Frequency Minimum
- Frequency Range
- Background Filter
- Max Syllable
- Max Syllable Gap
- Max Song
- Dynamic Range
- Algorithm Version

You can only use one recognizer group at a time when searching a recording. Within a group, you can select which recognizers you would like to include. During the search, Song Scope will automatically set the parameters to match those defined by the selected recognizer group.

The following additional controls are available in the recognizer control panel:

Minimum Quality

The Minimum Quality control adjusts a sensitivity filter that decides which recognizer results are good enough to be displayed during the search. The value specifies the minimum "quality" value required. The quality value is on a scale from 0.00 to 9.99 and represents a statistical distribution of parameters from the training data used to build the recognizer. A quality value of 5.00 indicates the

statistical average values, with lower numbers indicating less confidence and higher numbers indicating greater confidence. The default minimum quality setting of 2 will discard the bottom 20% of training data as of questionable quality. A minimum quality setting of 0 will disregard quality values and allow all candidate vocalizations through. Song Scope considers a number of the statistical factors listed in the Recognizer Generator Results.

Minimum Score

The Minimum Score control, like the Minimum Quality control described above, also decides which recognizer results are good enough to be displayed during the search. The value specifies the minimum "score" value required. The score value is on a scale from 0.00% to 100.0% and represents the statistical fit of the candidate vocalization to the recognizer's model. A candidate vocalization must achieve both the minimum quality and the minimum score to be counted. Note that the average score (and standard deviation) among training data for a given model is listed in the list of loaded recognizers and should be considered when deciding what value to use for minimum score.

Show Top Matches

The Show Top Matches Control limits the number of displayed results when there are more than one recognizer in a group being used in a search at the same time. By default, only the top (1) result will be displayed.

To start the scan, all you have to do is press the **"Start Scan"** button. Song Scope starts scanning from the currently displayed position in the recording file, and continues to the end of the recording or until you cancel. A result window is displayed with one row for each candidate vocalization in the recording and the following columns:

File name

The name of the scanned audio file. Note that this is used with batch processing to indicate which of possibly several files is the source of the candidate vocalization.

Time Offset

The Time Offset indicates in HH:MM:SS format the offset into the recording where the candidate vocalization begins.

Duration

The Duration indicates the length of the candidate vocalization in seconds.

Level

The peak signal level of the vocalization in decibels

Quality

The Quality indicates a signal quality confidence factor on a scale from 0.00 to 9.99. The quality is used by the sensitivity filter to decide if a signal is a suitable candidate for the recognizer. A value of 5.00 indicates that the signal characteristics are average to the characteristics found in the training data. A lower quality value indicates that there is less confidence in the signal being a good match, and higher values indicate greater confidence.

Score

The Score indicates how well the candidate matched the recognizer model. A successful match should fall in the same range as the recognizer's cross training results. The sensitivity control will also filter candidates based on their recognition scores.

Recognizer

The Recognizer column indicates the name of the recognizer corresponding to the result. Note that if the Show Top Matches control is greater than one, there may be multiple results for each candidate vocalization.

Comments

The Comments column can be used to record comments for vocalizations before saving the results. You can click on this field and type in arbitrary comments.

By default, results are sorted by time offset. You can re-sort the results by clicking on the column heading.

Clicking on a row in the result window will cause the spectrogram and waveform plots to center on the corresponding candidate and the vocalization to be automatically selected. You can press the space bar to play the selected audio segment.

Right clicking the mouse (or control-click for Mac OS X users) displays a pop-up window enabling you to "select all" and "copy" so that you can copy the results to the clipboard and paste it to an external spreadsheet or text editor program. Alternatively, Mac OS X users may also use command-a for select all and command-c to copy.

You can also save the results in a text file (tab delimited) by selecting **"Save..."** from the **"File"** menu. You can also re-load previously saved results from a batch window (even if the results were not created by a batch scan). This feature lets

you browse through old scan results without re-scanning.

Note that you can also scan without selecting any recognizers, in which case Song Scope will only use the signal detection parameters, and not any HMM-based classifiers, to build a spreadsheet with all detected events. In this case, the quality and score values are meaningless and set to zero.

Batch Processing

Batch processing is used to automatically search many different recordings for vocalizations.

From the **"File"** menu, select **"Batch..."**. This will open a new "Batch Processing" window. The control panel for this window contains a special "Batch" tab not found in other windows.

The "Batch" tab displays the currently selected search directory and contains the following controls:

Change Directory

Push the "Change Directory" button to select a different search directory.

Include Subdirectories

Select "Include subdirectories" to recursively search all subdirectories for recordings to scan

.WAV Files

Select ".WAV Files" to scan all *.wav* files found in the search directory. If "Include subdirectories" is also checked, then all subdirectories will also be searched for *.wav* files.

.WAC Files

Select ".WAC Files" to scan all *.wac* files found in the search directory. If "Include subdirectories" is also checked, then all subdirectories will also be searched for *.wac* files.

.AIF Files

Select ".AIF Files" to scan all *.aif/.aiff* files found in the search directory. If "Include subdirectories" is also checked, then all subdirectories will also be searched for *.aif/.aiff* files.

.??# Files

Select ".??# Files" to scan all *.??#* files found in the search directory. If "Include subdirectories" is also checked, then all subdirectories will also be searched for *.??#* files.

.SSN Files

Select ".SSN Files" to scan all Song Scope *.ssn* notation files found in the search directory. If "Include subdirectories" is also checked, then all subdirectories will also be searched for Song Scope *.ssn* notation files.

After configuring the settings in the "Batch" tab of the control panel, you can use the "Recognizers" tab to load and select recognizers as described in Searching Recordings for Vocalizations.

When you press the "Start Scan" button in the "Recognizers" tab, the selected recognizers will be used to scan all the recordings in the search directory as specified in the "Batch" tab. The results for all files will be tabulated in a single Batch Processing Results window. Clicking on a given result will cause the corresponding candidate vocalization to be displayed in the batch window.

