

## **Glossary for ABI proposal “Automating Avian Point Counts to Improve Population and Community Monitoring”**

(with partial credit to Wikipedia)

**Amplitude** – This is the volume or signal intensity of an acoustic signal at a specific time.

**Attenuation** – Reducing the amplitude or frequency range of a signal.

**Bandwidth** – The range of frequencies in a signal. Can be measured from arbitrary points such as the highest and lowest frequency that are at least 10% of the maximum amplitude.

**Band Pass Filter** - A filter that eliminates part of an acoustic signal outside of a specified frequency range. A low pass filter attenuates high frequencies (reduces their amplitude), and a high pass filter attenuates low frequencies.

**Bayesian Classifier** - A Bayesian classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a Bayesian classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.

**Center Frequency** – The mid-point or average frequency of an acoustic signal.

**Confusion Matrix** – A species-by-species matrix reporting the proportion of correctly identified song or call clips and the proportions of clips misidentified as other species. The rows and the columns of the confusion matrix are both name lists of species and they are in the same order. When a given call of a species at row X is correctly identified, it contributes a percentage of correctness at the diagonal position (the Xth row and the Xth column). If all calls of such a species are correctly identified, this diagonal entry will be 1, or 100%. Likewise, off-diagonal entries represent misidentified percentages. If a call of a species at row X is identified as the species at column Y, it contributes a percentage to the entry at the row X and the column Y. Ideally, one would like to have the confusion matrix to be all 1's on the diagonal, and all 0's over all off-diagonal entries.

**Conspecific** – A member of the same species.

**Cramer-Rao Lower Bound (CLRB)** – The CLRB is the theoretical lower bound on the variance of an estimator.

**Cross-correlation** – A function that estimates the correlation coefficients between two vectors or matrices at all possible lags. For example, the cross-correlations between two temporal vectors of sound amplitude from two spatially separated microphones could be calculated. The lag number of the highest correlation would be an estimate of the length of the temporal delay in the arrival of the sound at the more distant microphone.

Dawn Chorus – The period of most active singing by birds in the few hours just after sunrise.

Decibels (dB) – A logarithmic measure of sound intensity or power.

De-noise – A process of removing background information from a signal using filters, Wavelets, EMD, or other approaches.

Detection Probability – The probability of detecting an individual when it is present.

Discriminant Analysis – A multivariate technique for testing for differences between a priori groups. Requires multivariate normality of the data. Discriminant functions can serve as the basis for classification functions which can be used to predict group membership for novel data. So, classification functions developed from a library of known songs could then be used to identify and classify unknown songs.

Dynamic Time Warping (DTW) – An algorithm for measuring the similarity/dissimilarity of two sequences that may vary in their speed or tempo. For example, two birds that sing at slightly different rates or have slightly different lengths of pauses between syllables.

Empirical Mode Decomposition (EMD) – A method to decompose a signal into a finite and often small number of Intrinsic Mode Functions (IMF) that works well on non-stationary and nonlinear data such as bird calls. Often provides a clear decomposition of signals into separate time-frequency components (the IMF's), and thereby gives rise to sharper representation of the signals in the time-frequency-energy domain.

Envelope – An envelope of a family of curves is a curve that is tangent at some point to each member of the family of curves.

Feature – Any measure attempting to characterize the time-frequency content of a bird song or syllable.

Filter – A system that performs mathematical operations on a signal to remove or enhance parts of the signal. The transfer function of the filter determines how the signal will be modified or enhanced. For example, a high pass filter attenuates low frequencies (reduces their amplitude) and allows high frequencies to pass unmodified.

Frequency – The number of occurrences of a repeating event over time. Higher pitched sounds are higher frequency (more wave peaks per second). For example, the upper end of a piano keyboard is in the 4000 Hz or 4 kHz range where Hz is the number of cycles of the sound wave to impinge on the microphone sensor per second.

Frequency-Amplitude Spectra - A Frequency-Amplitude Spectra is a profile of the signal energy pooled among time bins over all frequencies in the time-frequency decomposition of the signal. It is computed as the sum of Gabor Coefficients among all time bins for each

frequency bin. It can be depicted graphically as a plot of the sum of Gabor Coefficient on the y-axis versus frequency on the x-axis.

**Gabor Coefficients** – The results of the Gabor Transform (see the next entry). The absolute value of the matrix of Gabor coefficients represents the sound energy in bins defined by time-frequency combinations for the sound with  $N/2$  ( $T < N < L$ ) frequency bins and  $L/T$  time bins (where  $L$  is the signal length).

**Gabor Transform** – Dennis Gabor, Nobel Prize winner for Holography, developed an alternative decomposition of a signal known as the *Gabor Transform* that is much less redundant than the spectrogram and provides an equivalent time by frequency decomposition of a signal and whose inverse recovers the initial waveform exactly. The absolute value of the matrix of Gabor coefficients represents the sound energy in bins defined by time-frequency combinations for the sound with  $N/2$  ( $T < N < L$ ) frequency bins and  $L/T$  time bins (where  $L$  is the signal length and  $T$  is the width of a time bin in samples). The resolution of a *Gabor Transform* depends on the parameters  $N$  and  $T$ , so more highly resolved decompositions use higher values of  $N$  and lower values of  $T$ .

**Heaping Bias** – The tendency of human observers to perceive sound sources as closer than they actually are. This results in overestimates of the number of sound sources within a fixed radius study plot, or “heaping” of sounds that are actually out of the study plot into the study plot.

**Hidden Markov Models (HMM)** – A statistical model which is assumed to be a Markov Process with an unobserved state. In a Hidden Markov model, the state is not directly visible, but output dependent on the state is visible. Each state has a probability distribution over the possible outputs. HMMs are used widely for temporal pattern and speech recognition.

**High Pass Filter** – See Filter.

**Intrinsic Mode Function (IMF)** – Components of a signal that are produced via Empirical Mode Decomposition. In a mixed signal comprised of different bird songs, each species should be associated with a unique set of Intrinsic Mode Functions.

**Low Pass Filter** – See Filter.

**Mel Frequency Cepstral Coefficients (MFCC)** – The amplitudes of a linear cosine transformation of the log power spectrum of a signal plotted on a mel scale. A mel scale is a perceptual scale of pitches judged by listeners to be equal distance apart. The mel frequency cepstrum is thought to more closely represent the human auditory system’s response.

**Microphone Array** - Two or more microphones arranged to detect and record the same signal.

**Neural Network** – A technique used in classification (among other applications). A non-linear statistical data modeling tool that can be used to model complex relationships between inputs and outputs or to find patterns in data. Networks are developed in a learning phase

based on known inputs (bird songs or syllables) and known outputs, and later applied to unknown inputs (bird songs or syllables) to classify their identity.

Noise – Extraneous sounds not produced by the sound emitter of interest.

White Noise – Noise where the amplitudes are independent of frequency. White noise is often added to signals in virtual sound studies to examine the effects of added noise on signal recognition.

Pink Noise – Noise where the amplitudes are inversely proportional to frequency. In natural environments where birds will sing the background noise is most likely pink noise. The sound of wind rustling leaves, water flowing in streams, rain fall, air planes, and motor vehicles are low frequency sounds so the sum of their amplitudes is relatively high compared to high frequency noise (insect noises, mammal alarm class, etc.) generating a pattern higher amplitudes of low frequency noise.

Omni-directional microphone - A microphone that is equally sensitive in all directions. A pick-up pattern that is equally sensitive in all direction.

Parabolic microphone – A type of directional microphone that includes a parabolic dish to focus sound energy from a wider area and hence increase the received signal intensity.

Passive Acoustic Location (PAL) – Passive Acoustic Location involves using the time-delay-of-arrival (TDOA) of acoustic signals received at each pair of sensors in a sensor array (in our case microphones) to estimate the location of the sound source. There are a variety of approaches to obtaining location estimates using PAL. Most commonly two-stage estimation methods have been employed in wildlife research, but one stage methods are available. Two-stage methods involve estimating first the cross-correlation function between the signals received by each pair of sensors to estimate the difference in the time of arrival of the sound at each sensor (TDOA), and then calculating the locations. A more recent one stage approach that may yield lower estimation errors is based on the total signal energy (TSE) and estimates the location of a sound as the point at which the sum of the signal strength is maximized among all sensors.

Point Count – A means of estimating the abundance or individual bird species, the number of species present, and species composition of a bird community.

Raw Counts – Counts of the number of individuals of each bird species detected during a point count.

Sampling Rate – The number of times/second that a signal is sampled in a digital recording. Normal CD quality recordings sample at 44,100 samples/second. The Nyquist Frequency is the highest frequency that can be detected and is equal to half the sampling rate. Therefore, CD recordings can include sounds as high in frequency as 22,050 Hz (cycles per second). Since most bird song occurs in the 2,000-8,000 Hz range, sampling at 44,100 samples/second is over sampling.

Segmentation – Segmentation is the process of dividing a digital recording into clips that represent vocalizations produced by a single individual bird.

Short Time Fourier Transform (STFT) – A Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.

Shotgun microphone – A microphone with a pickup pattern that limits sound detection to a relatively narrow region hence eliminating unwanted noise.

Signal to Noise Ratio (SNR) – A measure of the ratio of strength of the meaningful information to the background (noise) in a signal. It is measured in decibels as

$$SNR_{dB} = 10 \log_{10} \left( \frac{A_{signal}}{A_{noise}} \right),$$

where A is the root-mean-square of the amplitude of signal or noise.

Speaker Diarization - Speaker diarization is a problem analogous to detecting and identifying birds from their vocalizations that involves trying to automatically transcribe a board meeting or a broadcast news program. The goal is to separate the utterances of several human speakers that may or may not be fixed in location who have been recorded by an array of microphones.

Spectral Coefficients – Measure of sound energy in a time by frequency decomposition of a signal.

Spectrogram – A representation of the sound energy or spectral density of a signal in the time-frequency domain based on the short-time Fourier transform.

Spherical Spread – The loss of sound intensity in still air as a function of distance from the sound source. For each doubling of distance a signal loses an additional 6 dB in volume or intensity. Additional loss of signal intensity can occur from attenuation by vegetation and other causes.

Stochastic Sequence Modeling – See Hidden Markov Models.

Support Vector Machines – A set of related supervised learning methods used for classification and in regression. An approach to classification based on machine learning.

Syrinx – The vocal organ of birds. It is located at the base of the trachea where it divides into the bronchi.

Temporal-Removal Model – A model used in abundance estimation that assumes that detection of new individuals will decline over the course of the sample period. Application of temporal removal models requires that the time an individual is detected be recorded. The model assumes that the population of birds within the sample radius is closed (no

movement in or out, or births or deaths during the sample period), so as birds are detected they are “removed” from the population and fewer individuals remain to be detected in subsequent time intervals.

**Time-Amplitude Spectra** – A Time-Amplitude Spectra is a profile of the signal energy over the time course of a vocalization. It is computed as the sum of Gabor Coefficients among all frequency bins for each time bin. It can be depicted graphically as a plot of the sum of Gabor Coefficient on the y-axis versus time on the x-axis.

**Time-Delay-of-Arrival (TDOA)** – Difference in the time a signal arrives at a pair of microphones. TDOAs are used in Passive Acoustic Location (PAL).

**Total Signal Energy (TSE)** - A more recent one stage approach to Passive Acoustic Location (PAL) that may yield lower estimation errors is based on the total signal energy (TSE) and estimates the location of a sound as the point at which the sum of the signal strength is maximized among all sensors.

**Wavelet Transform** – A wavelet system is a mathematical *basis* or *frame* (of a function/vector space  $H$ ) formed by translations and dilations of one (or a set of several) waveform(s) in  $H$ . A *frame* is a basis-like set of functions/vectors in  $H$  that provides often-times non-unique representations of any function in  $H$ .

The Wavelet Transform of a function  $f$  stands for a “projection” of  $f$  onto a wavelet system. It produces wavelet coefficients (as fast as the Fast Fourier Transform) that are double indexed in translation and dilation parameters. The dilation parameter represents resolution levels/scales of a function. Consequently, a wavelet transform produces components of a function/signal at different resolution levels/scales.

The Wavelet Transform is known for its superb capacity in noise reduction and data compression. Its component at the lowest resolution level may also provide a more reliable “envelope” information of a signal that is useful in signal segmentation and localization.

**Waveform** – The shape of the signal. The waveform is a plot of amplitude on the y-axis versus sample number or time on the x-axis.