ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Animal Bioacoustics
Session 2aAB: Conditioning, Segmentation, and Feature Extraction in Bioacoustic Signals

2aAB5. Conditioning for marine bioacoustic signal detection and classification

David Mellinger*

*Corresponding author’s address: Coop. Inst. for Marine Resources Studies, Oregon State University, 2030 SE Marine Science Dr., Newport, OR 97365, David.Mellinger@oregonstate.edu

Marine acoustic signals are characterized by certain types of noise and interference. Conditioning methods applied to spectrograms can be used to reduce or even remove these sounds, making bioacoustic signals more evident and simplifying the tasks of detection and classification. One family of methods is for making a long-term estimate of noise at each frequency and subtracting this estimate from the spectrogram; this has the beneficial effects of whitening the noise spectrum and removing relatively stationary noise sources such as vessel sound, but has the detrimental effect that relative spectrum levels - important in echolocation click classification - are altered. Another method estimates the spectrum in narrow bands at each time step and subtracts this estimate from the corresponding spectrogram frame; this method is useful for tonal sound detection and classification in that it removes short-duration clicks from snapping shrimp and echolocating animals. Other methods for removing other, more rare types of noise are presented as well. Examples and performance characterization of these methods are presented. [Funding from ONR and N45.]

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Most marine acoustic signals have one or more types of noise or interference. Conditioning processes applied to signal time series or to spectrograms can be used to reduce or even remove these sounds, making bioacoustic signals more evident and simplifying the tasks of detection, classification, and localization (DCL).

Conditioning is done for two primary reasons. One reason is to remove noise – unwanted sound – from a signal or some representation of it. This improves the signal-to-noise ratio of the signal and usually improves the performance of DCL tasks. A second reason is to put a signal into a canonical form so that later processing steps can operate on a signal that has less variability due to recording equipment, noise sources, etc.

This paper reviews some conditioning methods useful for marine bioacoustic signals, and identifies characteristics of the signal or of the noise field that make the methods useful. The emphasis is principally on methods that can be applied continuously in time, rather than on methods that rely on a fixed-duration time window. Methods are illustrated by application to marine bioacoustic recordings.

TIME-DOMAIN METHODS

An example of a canonicalization method is an automatic gain control (AGC) process that operates on the time-series representation of a signal. The result of AGC processing is a signal with an overall amplitude level that is relatively constant compared to the wide variations in amplitude among possible input signals. This limits the dynamic range of the signal, making it easier to record the signal within amplitude limits determined by hardware (e.g., a fixed number of bits per sample) at the expense of losing absolute amplitude information. Automatic gain control can be implemented using these equations or something similar (adapted from Morgan 1975):

\[
\hat{x}(t) = \frac{rx(t)}{\tilde{x}(t)}
\]

\[
\tilde{x}(t) = (ax^p(t - 1) + (1 - a)|x(t)|^p)^{1/p}
\]

where \(x(t)\) is the discrete-time input signal, \(\hat{x}(t)\) is the conditioned (gain-controlled) signal, \(\tilde{x}(t)\) is the time-decaying average value of the signal using the \(L^p\) norm, and \(r\) is a constant that controls the absolute average value of \(\tilde{x}(t)\). The exponent \(p\) is often 1, which averages the absolute signal value, or 2, which averages the signal power. Other averaging functions \(\tilde{x}(t)\) are possible too, such as one that simply averages absolute values of the previous \(n\) samples:

\[
\tilde{x}(t) = \left(\frac{1}{n}\sum_{i=1}^{n}|x(t - i)|^p\right)^{1/p}
\]

Another conditioning method used in marine bioacoustic recording systems (e.g., Fox et al. 2001, Wiggins and Hildebrand 2007) is filtering for spectrum equalization. The idea is that since ambient ocean noise in most places is not flat-spectrum – usually being louder at lower frequencies (Wenz 1962) – one can employ a filter that compensates for this spectral tilt and thus makes background noise nearly flat. This provides greater dynamic range for recording the signal without distortion. (Spectral distortion can be corrected: Since the filter has a known frequency response, its effects can be reversed if desired to recover the original intensity of each frequency over time.)

SPECTROGRAM-DOMAIN METHODS

Both marine and terrestrial bioacousticians use spectrograms widely, and a number of conditioning methods have been developed for application to spectrograms.

One widely-used method is to estimate the background noise level in the spectrogram and subtract it from the spectrogram. If noise is estimated across the entire spectrum in a log-scaled spectrogram (which most are), this is equivalent to the time-domain AGC of Eq. (1); because the spectrogram is log-scaled, subtraction in the spectrogram domain is the same as division in the time domain. More usefully, noise can be estimated in some time/frequency region near a call of interest. One method (Mellinger et al. 2004) that has been found to be widely useful is to estimate and subtract noise separately in each spectral bin:
\[ \hat{S}(t, f) = S(t, f) - M(t, f) \]
\[ M(t, f) = \alpha S(t, f) + (1 - \alpha) M(t - 1, f) \] (3)

where \( S(t, f) \) represents the spectrogram, \( M(t, f) \) represents the time-averaged value of the spectrogram in frequency bin \( f \), and \( \alpha \) is a time constant that represents how quickly the time-averaging process adapts to new sounds. Some means is needed for the initial values \( M(0, f) \) at the start of processing (time \( t=0 \)) and across all frequencies \( f \); these may be obtained by averaging the first few spectrogram values for each frequency bin. This method is illustrated in Figs. 1 and 2, which show an unconditioned spectrogram and the same spectrogram after applying this method, respectively.

This method is useful when there are relatively stationary noise sources, such as ship sounds, wind/wave noise, or electrical noise, that occupy fixed frequency bands for long durations. An additional advantage is that it can be used to “whiten” the spectrum, that is, to make the background noise flat-spectrum. This form of canonicalization is helpful for making the background noise level approximately 0 at each frequency, greatly assisting many DCL algorithms, as well as enhancing faint calls to make them more apparent to DCL algorithms.

**FIGURE 1.** Spectrogram of a harbor seal (*Phoca vitulina*) roar vocalization (at 17-21 s, 0-1500 Hz) as originally recorded, i.e., without any conditioning. Noise sources include a steady hum near 60 Hz, perhaps due to electrical noise, apparent as a horizontal band near the bottom, and numerous click sounds from snapping shrimp, apparent as vertical streaks. Spectrogram parameters: Frame and FFT length 0.043 s, overlap 50%, Hamming window, for a filter bandwidth of 95 Hz.
Another method of estimating and removing background noise relies on a time-frequency (T-F) region around a call of interest. Because the call typically occupies less than 50% of the spectrogram cells in the box, the background noise can be estimated as the median value of all spectrogram cells in the box times a constant $c$. Indeed, sorting the spectrogram cell values reveals a noise amplitude curve that can be quite useful for estimating noise and then removing it from the spectrogram:

$$\hat{S}(t,f) = \max (0, S(t,f) - cN_m)$$

(4)

where $S(t,f)$ is again the spectrogram, $\hat{S}(t,f)$ is the conditioned spectrogram, $c$ is a constant, and $N_m$ is the $m^{th}$-percentile spectrogram value from all spectrogram cells in the box surrounding the call. Here the estimated noise level in the spectrogram is simply $cN_m$, which is further elucidated below.

An example is shown in Figs. 3, 4, and 5, which respectively show a ‘boing’ call of a minke whale (*Balaenoptera acutorostrata*) with a T-F box surrounding the call, the sorted spectrogram values in this box, and the spectrogram after subtraction of the estimated noise level. The curve in Fig. 4 shows all spectrogram values in the box sorted into ascending order. It would be possible to choose a noise level as simply a value at the 98th and 99th percentile, but this value would be heavily dependent on the exact percentile chosen as well as the size of the containing box. A better way to choose the noise level is to select the value where the sorted-values curve is flattest, namely the region near the 50th percentile (median), and multiply it by a constant $c$. Because the curve is flat, the value chosen is not highly sensitive to the exact percentile chosen, nor is it very sensitive to the size of the T-F box containing the call. Values for $c$ near 1.3-1.5 have been found to work well for marine bioacoustic sounds.

This method is useful when noise is relatively flat-spectrum, rather than having tonal components such as ship propeller sounds. It is often used when extracting acoustic characteristics (feature vectors) from calls for use in detection and classification.
FIGURE 3. Spectrogram of a North Pacific minke whale ‘boing’ call with a time-frequency box containing the call as well as some background noise. Spectrogram parameters: Frame and FFT length 0.085 s, 50% overlap, Hamming window, for a filter bandwidth of 48 Hz.

FIGURE 4. Sorted spectrogram values for the spectrogram cells in the time/frequency box of Fig. 3. The spectrogram is log-scaled. The values corresponding to the minke call are all at the far right; background noise values occupy positions up to at least the 90th percentile. Because of the relative flatness of the curve, the precise percentile chosen as the noise level has a relatively small impact on the estimated noise level. The median (50th percentile) is used for Fig. 5.
FIGURE 5. The spectrogram of Fig. 3 after conditioning by subtraction of the noise estimate \( c \times \text{median value } N_{50} \) derived from Fig. 4. Most of the noise has been eliminated and most of the call retained, though a narrow band of noise near 900 Hz is still apparent.

Some other techniques for conditioning have the aim of removing specific types of noise present in marine bioacoustic signals. One such noise type is a click sound; clicks are produced by many marine animals such as odontocete cetaceans (the suborder Odontoceti, comprising dolphins, porpoises, and toothed whales) and snapping shrimp (shrimp of the family Alpheidae). Click sounds can be removed by making an estimate of the average or median spectrum level over a broad range of frequencies at each spectrogram time step and subtracting this from the spectrogram. This can be done using a process similar to that of Eq. (1), but operating across frequency (vertically) in each time-slice of the spectrogram instead of across time (horizontally) in each frequency bin:

\[
\begin{align*}
S(t, f) &= S(t, f) - M(t, f) \\
M(t, f) &= \alpha S(t, f) + (1 - \alpha) M(t, f + \Delta f)
\end{align*}
\]

where the variables are as for Eq. (1), and \( \Delta f \) representing the difference between the frequencies of adjacent frequency bins in the spectrogram. In this case \( \alpha \) is a frequency constant controlling how quickly the process responds as it operates across frequencies. The plus sign in the \( f + \Delta f \) term indicates that this form of the equation operates from high frequencies to low. Analogously to Eq. (1), some means is needed for initializing \( M(t, f_{\text{max}}) \) when the process begins at the highest available frequency \( f_{\text{max}} \); this is of somewhat more consequence than for Eq. (1), since the initialization must occur at every time step in the spectrogram rather than just once at the start of processing. Again, this can be done by averaging the first (highest) several spectrogram values in each time step.

For click noise removal, a noise estimate can also be made in a manner similar to Eq. (4) and Fig. 4 by calculating a noise estimate \( cN_{\text{m}} \) for each time-slice of the spectrogram. Again, this is the product of a constant \( c \) and a certain percentile value \( N_{\text{m}} \) of all spectrogram cells between some chosen frequency bounds \( f_0 \) and \( f_1 \). After calculating this noise estimate \( cN_{\text{m}} \) for each time-slice in the spectrogram, the noise is subtracted according to Eq. (4).

Fig. 6 illustrates removal, by the process of Eq. (5), of snapping shrimp clicks from the spectrogram of Fig. 2. Nearly all of the click sounds are removed, though there are a few residual clicks that survive the process.
FIGURE 6. The spectrogram from Fig. 2 after application of conditioning for click removal. Note the removal of nearly all the snapping shrimp click sounds but some distortion, appearing similar to amplitude modulation, in the harbor seal roar. The frequency constant for click removal was 1 kHz.

Another type of conditioning, median-replacement conditioning, is useful for removing residual click sounds, such as seen in Fig. 6, and noise ‘speckles’ that appear in spectrograms. The idea is simply to replace each spectrogram cell by the median of a small area surrounding and including the cell:

$$S(t, f) = \text{median}_{t - k_t < t < t + k_t, f - k_f < f < f + k_f} [S(t, f)] \quad (6)$$

where $k_t$ indicates the extent in time over which the median is taken, and $k_f$ indicates the extent in frequency. If $k_t = 0$, then the median is calculated only across nearby frequencies (i.e., along a vertical line centered on the cell), and if $k_f = 0$, it is only across nearby time steps (i.e., along a horizontal line centered on the cell). Of course, the value selected is not required to be the median (the 50th percentile); other percentile values work as well.

An illustration of median filtering used to remove residual parts of click sounds is shown in Fig. 7. In this figure, the median filter used operated only across time, i.e., $k_f = 0$, since the primary target was click sounds extending across multiple frequencies. It may be seen that the median filter removes nearly all of the remaining click sound present in Fig. 6 as well as the background noise ‘speckle’ seen in most spectrograms.
FIGURE 7. The spectrogram from Fig. 6 after application of a time (horizontal) median filter. Note the disappearance of the remaining snapping shrimp click sounds and great reduction in the amplitude modulation distortion, at the expense of some sharpness in the image. Important frequency banding information, perhaps from formants in the harbor seal roar, is still present, as are the amplitude changes in the roar. The length of the median filter was 0.12 s and the 30th-percentile value was used rather than the median (50th percentile).

CONCLUSION

A variety of methods were described for conditioning time series and spectrograms, and the application of many of these methods were illustrated with examples containing marine mammal sounds. The methods appeared to be effective at removing several types of noise from recordings.

These methods have been applied heuristically, with method choice dependent on the nature of the noise present. Although all of the methods have been used successfully for conditioning sounds for use in DCL algorithms, their performance has not been quantified in any meaningful way. Performance evaluation for conditioning algorithms is an open area of research, and the author looks forward to any ideas on the subject.

ACKNOWLEDGMENTS

This work was funded by Office of Naval Research grants N00014-11-IP20086_FDOC_24295771, N00014-12-IP20052, and N00014-10-1-0387 as well as by Naval Postgraduate School grants N00244-10-1-0047 and N00244-11-1-0026 (funded from Navy N45 and Army AIST programs). Thanks also to Steve Martin and the Pacific Missile Range Facility for the recording of the minke boing sound. The harbor seal sound was recorded under a grant from the David and Lucile Packard Foundation.

REFERENCES


