A Correlation-Maximization Denoising Filter Used as An Enhancement Frontend for Noise Robust Bird Call Classification

Wei Chu and Abeer Alwan

Speech Processing and Auditory Perception Laboratory
Department of Electrical Engineering
University of California, Los Angeles

Supported in part by the NSF

Outline

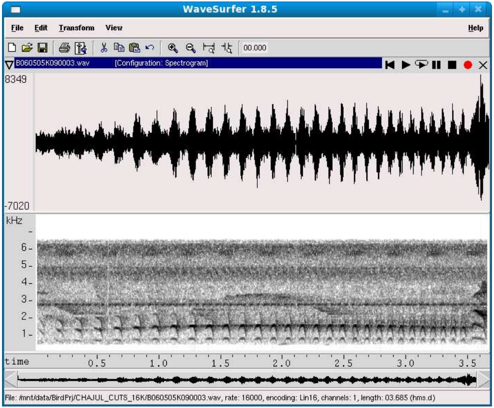
- Motivation
- Bird Call Analysis
- Bird Call Classifier Design
- Denoising Filter Design
- Experiments

Motivation of noise robust bird call classification

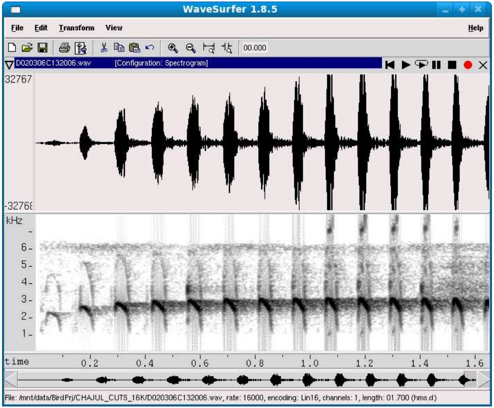
- Songs are important in the communication between birds of specific species.
- Behavioral and ecological studies could benefit from automatically detecting and identifying species from acoustic recordings.
- It is a challenge to correctly classify the bird calls under noisy conditions.
- In this work, we analyze 5 types of Antbirds.

Now let us listen to several examples of Antbird calls:

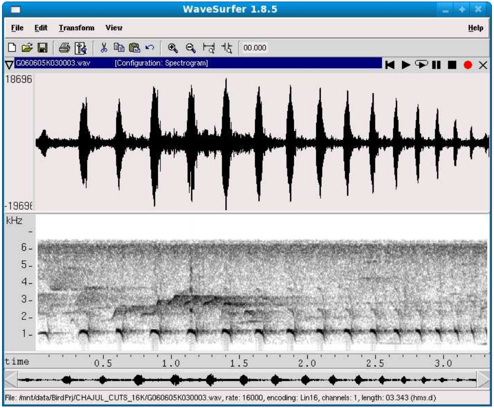
Waveform and spectrogram of a Barred Antshrike (BAS) call



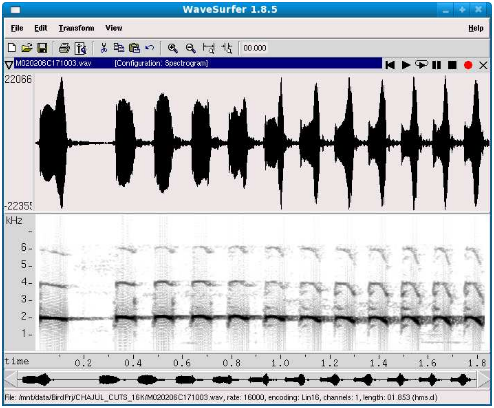
Waveform and spectrogram of a Dusky Antbird (DAB) call



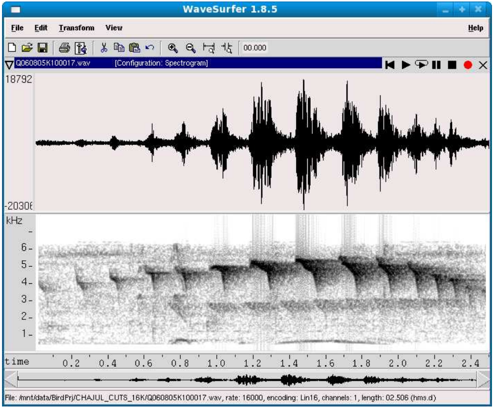
Waveform and spectrogram of a Great Antshrike (GAS) call



Waveform and spectrogram of a Mexican Antthrush (MAT) call

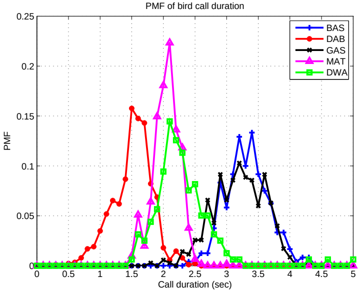


Waveform and spectrogram of a Dot-winged Antwren (DWA) call



Antbird Call Properties

- A bird call consists of a sequence of chirps.
- The interval between chirps and the chirp intensity gradually decrease over time.



A histogram of bird call duration of 2246 samples from 5 bird species. The duration ranges from 0.5 to 5 seconds.

Automatic bird call classification involves several aspects:

- Waveform denoising: the focus of this paper
- Feature extraction: Mel-Frequency Cepstral Coefficients (MFCCs)
- Acoustic modelling: Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM)
- Learning model parameters from observations
- Decoding observations

Why denoising is needed?

Different kinds of background noise can be observed in the recordings:

- Other bird chirps
- Insect sounds
- Sounds of other animals

We propose a Correlation-Maximization based filter to suppress background noise existed in the bird calls.

A prevailing denoising approach: Wiener filtering

Clean $X(f)$ is corrupted by an additive noise $\Rightarrow$ noisy $Y(f)$.
$\widehat{\mathrm{SNR}}(f)$: an estimation of $\mathrm{SNR}(f)$:

$$\widehat{\mathrm{SNR}}(f) = \frac{|\hat{X}(f)|^2}{|\hat{N}(f)|^2} \quad (1)$$

The estimated clean spectrum is :

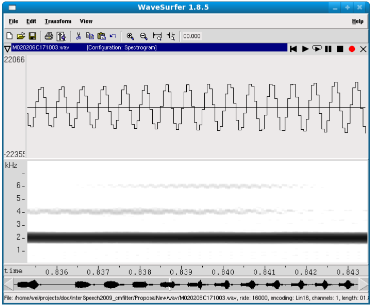
$$|\hat{X}(f)|^2 = H(f)|Y(f)|^2 = \frac{\widehat{\mathrm{SNR}}(f)}{1+\widehat{\mathrm{SNR}}(f)}|Y(f)|^2 \quad (2)$$

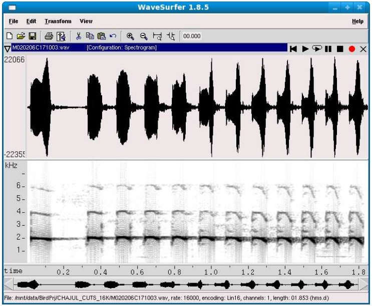
The noncausal Wiener filter converts the denoising problem into an SNR estimation problem [1].

Futher Analysis of the Bird Call

Two Levels of Bird Call Periodicity

1. Short phonation period (Left): ranges from 0.2 - 1.0 ms
2. Interval between chirps (Right): ranges from 0.06 - 0.3 sec, slowly decreases with time. ⇒ instruct the denoising!





Correlation-Maximization Filter

Suppose an FIR filter with L taps:

$$\mathbf{h} = [h[1], h[2], \cdots, h[L]]^T \tag{3}$$

is used for denoising the noisy bird call $y[n]$.
The output of the filter is the estimated clean signal $\hat{x}[n]$:

$$\hat{x}[n] = \sum_{k=1}^{L} h[k]y[n-k] \tag{4}$$

$y[n]$ and $\hat{x}[n]$ is then segmented into frames.

Correlation-Maximization Filter (cont.)

Two Assumptions

1. $y[n]$ and $\hat{x}[n]$ are wide sense stationary: The bird chirps are repeating periodically.
2. A single $\mathbf{h}$ for each bird call: The spectral distributions of different frames in a bird call are similar.

The cross correlation function of $\hat{x}[n]$ at lag k of frame m:

$$\phi_{\hat{x}}^{m}[0,k] = \mathbf{h}^T \Phi_y^m[0,k]\mathbf{h} \tag{5}$$

$\mathbf{h} = [h[0], h[1], \cdots, h[L]]^T$: coefficients of the FIR filter.
$\Phi_y^m[0,k]$: cross correlation function of $y[n]$ (independent of $\mathbf{h}$)

Use Dynamic Programming (DP) to Search the Chirp Interval

Searching the chirp interval in each frame over $\hat{x}[n]$.

DP: minimizing the distortion induced by background noise

- Local cost at lag k of frame m: $-\bar{\phi}_{\hat{x}}^{m}[0, k]$
- Transition cost of from lag k_i at to k_j:

$$d(k_i, k_j) = e^{\alpha|k_i - \delta - k_j|} - 1 \tag{6}$$

Purpose: prevent chirp intervals from greatly varying in two consecutive frames.

A trellis structure of $K \times M$ for dynamic programming is built.

Correlation-Maximization Filter (cont.)

The effect of an optimal filter **h**

Removing the additive noise in the corrupted signal so that the minimum accumulative cost is achieved in chirp interval searching:

$$\mathbf{h}^* = \arg\min_{\mathbf{h}} \mathcal{F}(\mathbf{h}, \mathbf{s}) \quad (7)$$

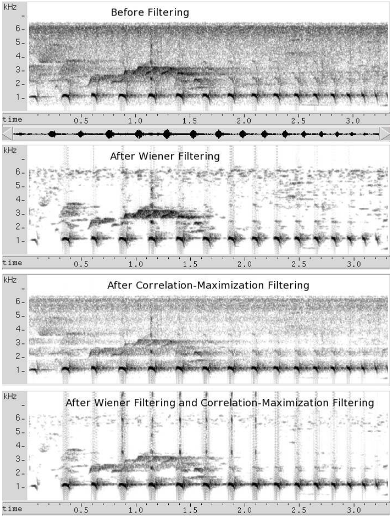
$\mathbf{s}$: an valid path in the trellis: $\mathbf{s} = s_1, s_2, \cdots, s_M$,
$\mathbf{h}^*$: the optimal denoising filter.
the accumulative cost $\mathcal{F}(\mathbf{h}, \mathbf{s}) = \Psi(\mathbf{h}, \mathbf{s}) + \Theta(\mathbf{h}, \mathbf{s})$.
$\Psi(\mathbf{h}, \mathbf{s})$: accumulative local cost;
$\Theta(\mathbf{h}, \mathbf{s})$: accumulative transition cost.

Speed Up: From Brute Force to N-Best

- There are K^M possible paths in a $K \times M$ trellis. Suppose the average iteration times of the gradient search is $\bar{I}$, this brute-force approach needs $K^M \times \bar{I}$ iterations which is computationally unacceptable.
- We can assume that $\mathbf{s}^*$ is within a path subset denoted by $\mathcal{S}(\mathbf{h})$ in each iteration. The subset is composed of the top N-best paths from the dynamic programming using the trellis.
- That means the gradient descent search is only needed to be applied to the N-best paths, not all the paths at each iteration.
- Let J denotes the size of N-best search, the total gradient search iterations is reduced to $J^2 \times \bar{I}$.
- Typically, for Antbird calls, $K = 49$, $1 \leq M \leq 50$, $J = 20$.

The spectrograms of a GAS call before and after filtering

- (a) other non-target bird chirps: 0.6 - 1.6 seconds
- (b) both target and non-target bird chirps are enhanced after Wiener filtering
- (c) Correlation-Maximization filter suppressed the non-target chirps while enhancing the target chirps
- (d) non-target chirps and background noise are suppressed when cascading two filters



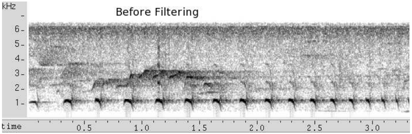
Correlation-Maximization Filter

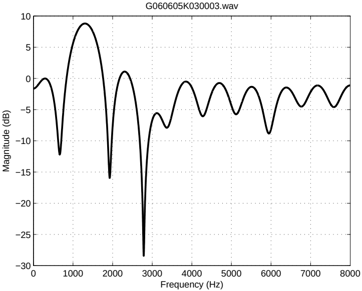
The frequency response of the CM filter for a GAS call

- enhanced the the target bird call;
- minimized the interference introduced by background noise and other bird.

filter h's characteristic

- pass-band: 800 - 1750 Hz
- stop-band: 2600 - 8000 Hz
- dip: around 2800 Hz





Data Set

Researchers from UCLA Ecology and Evolutionary Biology department collected 2 hours of bird calls (3366 calls) from 5 species. We split the corpus into a training and testing set with a ratio of 2:1.

Table: 2.1 The number of bird calls in the training and test sets. BAS: Barred Antshrike; DAB: Dusky Antbird; GAS: Great Antshrike; MAT: Mexican Antthrush; DWA: Dot-winged Antwren.

	BAS	DAB	GAS	MAT	DWA	Total
Training	240	888	350	609	159	2246
Testing	120	444	175	304	77	1120

The training set has 85 minutes of recordings; the testing set is 42 minutes long.

Setting

- A band-pass filter with cutoff frequencies at 360 Hz and 6500 Hz is used to remove the irrelevant frequency components.
- Downsamped from 44.1 kHz to 16 kHz.
- The taps of the filter $L = 20$.
- The frame length $N = 600ms = 9600samples$.
- The dimensions of MFCC features is 39.
- GMM: 256 Gaussians; HMM: 6 states, 256 Gaussians / state.

Classification Results Analysis

Table: 2.2 The classification error rate using the bird call test set. **W+/CM+**: feature extraction using the output of the Wiener/Correlation-Maximization based denoising filter

	GMM	HMM
MFCC	8.7%	5.4%
W+MFCC	5.9%	4.9%
CM+MFCC	5.3%	4.6%
CM+W+MFCC	**4.7%**	**4.1%**

- HMM based classifier is better than the GMM classifier when using the same features.
- Correlation-Maximization based denoising filter is effective before extracting MFCC features.
- Cascading the CM filter and Wiener filter is most effective.

Conclusions and Future Work

The Correlation-Maximization based denoising filter is effective in reducing the classification errors of the bird call which has a quasi-periodic structure in the time domain and an invariant power spectral density across frames.

Future work

- Extract better features for classification, such as long-term features and the modulation frequency features;
- Detect the bird call in an audio stream.
- Use Dynamic Bayesian Network to represent the probabilistic relationships between the observed bird calls and the bird species.

Thank you!

Q & A?

S. Boll,
“Suppression of acoustic noise in speech using spectral subtraction,”
IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 27, no. 2, pp. 113–120, 1979.