Segmentation of birdsong

Author Name[1], *Co-author Name*[2]

[1]Author Affiliation
[2]Co-author Affiliation
author@university.edu, coauthor@company.com

Abstract

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

2. Entropy-based segmentation of birdsong

The entropy of spectrograms of bird song recordings can be effectively used for distinguishing between background and bird vocalizations [**?**]. Spectrogram of single bird song is generally sparse i.e. high power components acquire only a small portion of time-frequency bins and the background noise of spectrogram is relatively white. Hence the entropy of sliding time frequency block over spectrogram is low when block contains a signal and is high when only background is covered.

2.1. Entropy Calculation

The time-frequency block of time length ***w*** and having F frequency bins ranging from ***f1*** to ***fn*** is slided horizontally from beginning to the end of spectrogram. Entropy is calculated for each time-frequency block. ***p(n,f)*** is power spectrum at time ***n*** and frequency ***f***. The entropy is calculated using following equation:

$$h_k = \sum_{n=kT+1}^{kT+w} \sum_{f=f1}^{fn} z(n,f) \ln z(n,f)$$

Here ***T*** is time-frequency block shift and ***z(n,f)*** is normalized power spectrum.

$$z(n,f) = \frac{p(n,f)}{\sum_{n=kT+1}^{kT+w} \sum_{f=f1}^{fn} p(n,f)}$$

The entropy calculated from block is less susceptible to the bursts of background noise in comparison to the entropy calculated at each time instance.

2.2. Thresholding to detect change points

To detect the change points, thresholding is used. First of all, the entropy is smoothed by using moving average. The local minimums and local maximums are calculated on this smoothed entropy. Then the difference between consecutive local minimums and local maximums is calculated. If this difference is greater than pre-defined threshold, then corresponding local minimum is the change point. Two contiguous change points correspond to the start and end of a bird vocalization. These change points can be tracked back to get the start and end time of the vocalization in sound recording. Smoothing entropy helps in decreasing missed detection rate to some extent.

Table 1: Table showing Correct (%) Missed Detection(%) and False alarm (%) for particular thresholds and moving average windows

File	Span	Threshold	Correct (%)	False Alarm (%)	Misse
1.wav	9	1.2	77.7	15.6	
2.wav	3	1.45	73	45	
3.wav	3	1.5	71	22	
4.wav	9	1.15	75.1	21	
5.wav	7	1.5	76.5	18.5	
6.wav	9	1.25	82.1	15.07	
7.wav	7	1.25	79.6	17.8	
8.wav	9	1.15	78.8	19.3	
9.wav	5	1.5	86.06	13.6	
10.wav	9	1.6	79.6	18.3	
11.wav	3	1.45	91.1	32.7	
12.wav	7	1.4	85.1	11.3	

2.3. Experimentation and Performance Analysis

For experimentation, the labeled recordings of Cassins Vireo (*Vireo cassinii*) are used [**?**]. The total duration of recordings is about 45 minutes. Out of 45 minutes, about 5 minutes of recordings correspond to the phrases of Cassin's Vireo. To calculate spectrogram, frame length of 20 ms and increment of 5 ms is used. The time-frequency window of 138.8 ms is used along with increment of 15 ms to calculate entropy.

For performance analysis, three metrics i.e. correct detection rate, missed detection rate and false alarm rate are used. These metrics are calculated using following equations:

$$\text{Correct (\%)} = \frac{\text{Correctly classified frames}}{\text{Total frames}} \times 100$$

$$\text{Missed Detection (\%)} = \frac{\text{Call frames classified as background}}{\text{Total call activity frames}} \times 100$$

$$\text{False Alarms (\%)} = \frac{\text{Background frames classified as calls}}{\text{Total background frames}} \times 100$$

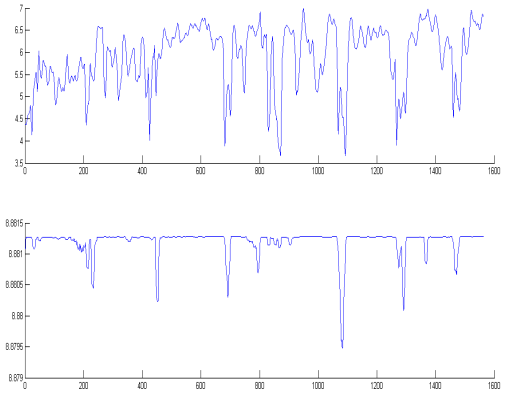
Table 1 shows results generated by applying entropy based segmentation with thresholding on different sound recordings.

2.4. The whitening Spectrogram using PCA before Entropy Calculation

To whiten the spectrogram (*PS*), the overall mean is calculated. This mean is subtracted from all values of spectrogram matrix. The covariance matrix of mean subtracted spectrogram is calculated. The Eigen values matrix (*S*) and Eigen vectors matrix (*U*) of this covariance matrix are calculated. The spectrogram matrix is whitened using the following equation:

$$WhitePS = diag(\frac{1}{\sqrt{diag(S)+\epsilon}}) * U' * PS$$

The missed detection rate is decreased if whitened spectrogram is used for entropy calculation. This entropy remains almost constant for background but dips enough to mark the presence of bird vocalizations. Even low energy bird vocalizations can detected using entropy calculated from whitened spectrogram. Following figure depicts the difference between entropy calculated from normal spectrogram and whitened spectrogram:



3. Conclusions

CHECK

4. Acknowledgements

The ISCA Board would like to thank the organizing committees of the past INTERSPEECH conferences for their help and for kindly providing the template files.