BIRD PHRASE SEGMENTATION BY ENTROPY-DRIVEN CHANGE POINT DETECTION

Ni-Chun Wang[⋆]*,Ralph E. Hudson*[⋆]*, Lee Ngee Tan*[⋆]*, Charles E. Taylor*[†]*, Abeer Alwan*[⋆]*, and Kung Yao*[⋆]

[⋆] Electrical Engineering Department, [†] Department of Ecology and Evolutionary Biology,
University of California-Los Angeles, Los Angeles, CA 90095, USA

ABSTRACT

A bird phrase segmentation method using entropy-based change point detection is proposed. Spectrograms of bird calls are usually sparse while the background noise is relatively white. Therefore, considering the entropy of a sliding time-frequency block on the spectrogram, the entropy dips when detecting a signal and rises when the signal ends. Rather than applying a hard threshold on the entropy to determine the beginning and ending of a signal, a Bayesian change point detection is used to detect the statistical changes in the entropy sequence. Tests on a database of Cassin's Vireo (*Vireo cassinii*), our proposed segmentation method with spectral subtraction or a novel spectral whitening method as the front-end generates more accurate time labels, lower the false alarm rate than the conventional time-domain energy detection method and achieves high phrase classification rate.

***Index Terms*—** Bird phrase segmentation, Entropy, Change point detection, Spectrogram, Bird phrase classification

1. INTRODUCTION

An automated system capable of reliably segmenting bird songs and identifying species is an indispensable tool for analyzing an audio database, used for studying behavior of vocalizing species, and a more refined understanding of bird communication [1]. Several species identification methods have been shown to be useful in many aspects; however the automated segmentation of the bird songs has received less attention. Manually segmented bird songs were used for bird species identification in [2–5]. Time-domain energy detection is used in [6–8] for segmentation. In [9], the authors used Kullback-Leibler (KL) divergence between the audio power spectrum and the uniform distribution for segmentation. A time-frequency segmentation by machine learning methods is proposed in [10].

Our goal is to find accurate and consistent phrase labels such that the segmentation results could be passed to a phrase classifier for reliable classification. Phrases are typically the basic units of understanding the bird communication. A phrase usually consists of several syllables with short silence in between, which makes phrase segmentation non-trivial. We propose a time/frequency segmentation method by entropy-driven change point detection. Entropy is calculated from the sliding time-frequency blocks in the spectrogram. Since the spectrogram of single bird songs is generally sparse in the sense that high power signals occupy a small fraction of the time and frequency bins. An example is shown in Fig. 1(a). This is because each phrase usually consists only a single frequency at any given instant. Harmonics may be present but even so the instantaneous spectrum is sparse. The phrase could be slowly modulated in amplitude or frequency [11] but still the spectrogram is sparse. In contrast, when there is no phrase present, the spectrogram displays a random response whose statistics do not vary with time or frequency. Therefore, the entropy drops when the sliding block is moving from an time interval without any bird songs ("quiet period" will be used hereafter) to the start of a song. The entropy stays low in the time interval of a bird phrase ("call period" will be used hereafter) and rises up as the block is leaving the call period. A polynomial-based spectral whitening method is also proposed to serve as the front-end of the system. The purpose is to enlarge the difference between the entropy levels of a call period and a quiet period.

When applying the segmentation method to the long field recordings, the entropy of bird phrases is not always at the same level and is sometimes even higher than the one at certain quiet periods, depending on the interference level. Further, the energy received from a song may depend on which direction the bird is temporarily facing [12]. Applying a hard threshold as considered in other time-domain segmentation methods in the literature [6–9] would easily miss those phrases. Instead, we propose using a change point detection method to detect the abrupt changes in the statistics of the data. Therefore, it can distinguish call periods from quiet periods as long as there are changes in the entropy level. Change point detection (CPD) detects change in the generative parameters of a time series. CPD has been shown to be a key aspect of many real world applications [13–17]. In this paper, we use an online Bayesian CPD proposed in [18]. This approach is based the assumption that the data models across segments are i.i.d. and non-overlapping, which is a reasonable assumption for bird phrase segmentation.

2. ENTROPY AND FRONT-END PROCESSING

In the following, we will describe how we characterize each time-frequency block in the spectrogram by its entropy in an efficient way. Additionally, two types of front-end processing are introduced. Spectral whitening is applied to the spectrogram before calculating entropy in order to further distinguish the entropy level between a call period and a quiet period. Spectral subtraction can also be used as a front-end processing to mitigate the interference and background noise.

2.1. Entropy Calculation

A time-frequency block of time length w and containing F frequency bins from f_1 to f_F is sliding horizontally from the beginning of the spectrogram, as shown in Fig. 1(a). The selection of the block length and the frequency range depends on the targeted bird species. The frequency limits f_1 and f_F should be properly selected to cover only the frequency band of interest. The block length w should be

This work is partially supported by NSF grant IIS-1125423. The work of N.-C. Wang is partially supported by National Science Council, Taiwan. (TMS-094-2-A-002). The authors would like to thank Dr. Martin L. Cody for providing the field recordings and George Kossan for annotating the phrases.

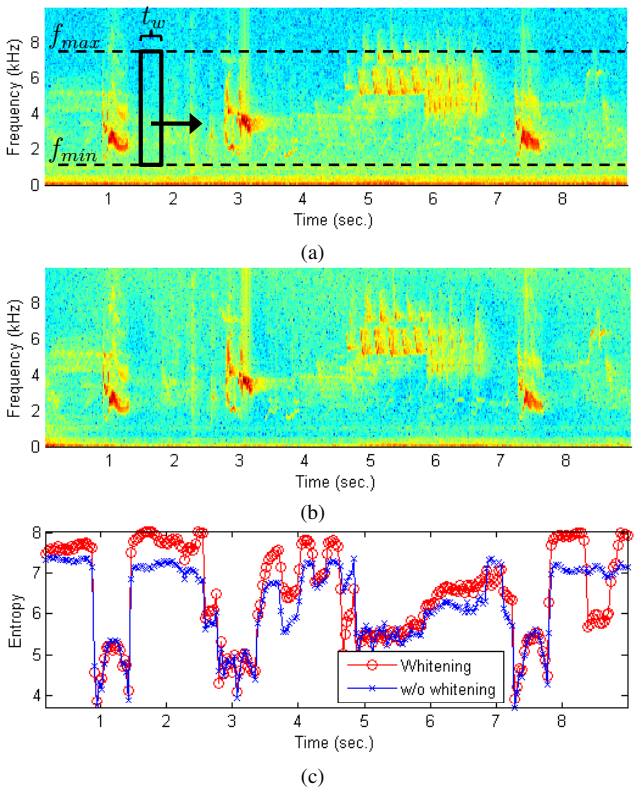

Fig. 1. (a) Spectrogram of a sampled bird vocalization recording and a time-frequency sliding block for calculating the entropy. (b) The whitened spectrogram (c) The entropy sequences with and without whitening calculated from the sliding time-frequency block.

no longer than the length of the shortest quiet period between two phrases.

Entropy is calculated from each time-frequency block. Denote τ as the block shift and $p(n,f)$ as the power spectrum at time n and frequency bin f calculated from the short-term Fourier transform (STFT). The entropy h_k is obtained by

$$h_k = -\sum_{n=k\tau+1}^{k\tau+w}\sum_{f=f_1}^{f_F} z(n,f)\ln z(n,f), \tag{1}$$

where $z(n,f) = p(n,f)/\sum_{n=k\tau+1}^{k\tau+w}\sum_{f=f_1}^{f_F} p(n,f)$ is the normalized power spectrum within the block. There are two main reasons why we calculate the entropy from a time-frequency block instead of at every time instant: 1) The entropy sequence is smoothed to prevent the "border effect" [6,19,20], which usually causes segmentation errors at the beginning and toward the end of a call. 2) The entropy given by a block is more representative and suffers less from bursty background noise.

However, when the block rate is high, calculating entropy purely by (1) is expensive in terms of memory consumption and computational load. To this end, an alternative expression of h_k is developed;

$$h_k = -\frac{B_k}{A_k} + \ln A_k, \tag{2}$$

where $A_k = \sum_n \sum_f p(n,f)$, $B_k = \sum_n \sum_f p(n,f)\ln p(n,f)$. The lower and upper limits of the summations are the same as in (1) but they are omitted here for simplicity. The calculation of these two terms can be significantly reduced by saving two partial sums over frequencies of interest, $a_n = \sum_f p(n,f)$ and $b_n = \sum_f p(n,f)\ln p(n,f)$, at the completion of each STFT. Then A_k and B_k can be obtained by accumulating the partial sums over the current block time to get inputs to the entropy calculation; i.e.

$$A_{k+1} = -\sum_{n=k\tau+1}^{(k+1)\tau} a_n + A_k + \sum_{n=k\tau+w+1}^{(k+1)\tau+w} a_n. \tag{3}$$

B_k can be updated in the similar fashion.

2.2. Spectral Cleaning and Whitening

The entropy defined in (1) is maximized if $z(n,f)$ in a block is uniformly distributed. Therefore, the entropy of a block with white background noise is higher than the one with color background noise. In contrast, the entropy is low when there are only few strong power components that dominate others within the block. It implies that we can get a lower entropy if we are able to standardize the noise within a block. From these observations, one can enlarge the difference of the entropy levels between a quiet period a call period by applying either a spectral cleaning or a spectral whitening method. The spectral cleaning method we use is spectral subtraction [21–23] which is a well-known noise reduction technique in speech processing. As for the spectral whitening, we proposed a computationally efficient polynomial-based method.

2.2.1. Polynomial-based Whitening Filter

The basic principle is to multiply the power spectrum at time n, $\mathbf{p}_n = \begin{bmatrix} p(n,f_1) & \cdots & p(n,f_F)\end{bmatrix}^T$, by a c^{th} degree polynomial over the frequency bins of interest. The coefficients of this polynomial have to be adaptively adjusted over time. Let the whitening filter at time t and over frequency bins f_1 to f_F be written as $\mathbf{Q}\cdot\mathbf{g}_t$, where $\mathbf{g}_t$ is a $(c+1)\times 1$ vector representing the polynomial coefficients varying with time, and $\mathbf{Q}$ is a $F\times(c+1)$ matrix with orthonormal columns. The degree of the approximation polynomial c need not to be large because the background noise is usually not rapidly changing. A quadratic polynomial to capture the dynamic of the spectrum is used in this work. For quadratic polynomial the three basis columns of $\mathbf{Q}$ are, a constant vector, a vector linear in frequency, and a vector quadratic in frequency. With proper shift and scale in the frequency, these vectors can easily be made orthonormal.

It is not desired to whiten the bird call power spectrum along with the background noise power spectrum. To reduce the sensitivity to the sparse and high energy bird calls when present, the whitening polynomial is set to capture the variation of the time-averaged log power spectrum. Namely, the polynomial should satisfy

$$\mathbf{Q}\mathbf{g}_n + \boldsymbol{l}_n = \mathbf{0}, \tag{4}$$

where $\mathbf{0}$ is a zero vector, $\boldsymbol{l}_n = \frac{1}{M}\sum_{i=n-M+1}^{n}\ln\mathbf{p}_i$ and M should be much larger than the number of $\mathbf{p}_i$'s in a single bird phrase. From (4), it can be derived that the polynomial coefficients $\mathbf{g}_n$ should be updated in a recursive manner as the new STFT output $\mathbf{p}_{n+1}$ is available,

$$\mathbf{g}_{n+1} = \mathbf{g}_n + \frac{-\mathbf{Q}^T\ln\mathbf{p}_{n+1} - \mathbf{g}_n}{M}. \tag{5}$$

In Fig. 1(b) and 1(c), we show the whitened spectrogram and compare the entropy sequences of the same recording before and after performing the proposed whitening method. It is clear that the entropy level of the quiet period becomes higher while the entropy of the bird call period is about the same level, which shows the effectiveness of the proposed method.

3. BAYESIAN CHANGE POINT DETECTION AND POST-PROCESSING

Given the entropy sequence computed by the method described in Section 2, we need to distinguish the bird calls from the quiet periods by its level. A Bayesian change point detection method is used to judge the starting and end points of a bird phrase.

3.1. Online Bayesian Change Point Detection

The change point detection technique used here is based on the work in [18]. In this section, we will briefly introduce this method with some modification made in order to fit our application.

Denote $\mathbf{h}_{1:t}$ as the data sequence, $h_1, h_2, \cdots, h_t$, observed from time 1 to t. Assume all the data are independently sampled from the same class of probability distribution, but the parameter set of the distribution could be changing over time. Therefore, the sequence is divided into mutually exclusive segments, and the data within each segment are independently sampled from the distribution of the same parameter set. A change point occurs when there is a change in the parameter set, which is at the beginning of each segment. Define the run length r_t as the time length since the last change point observed at time t, so $r_t = 0$ indicates a change point at time t. The objective is to estimate the run length by

$$\hat{r}_t = \max_{r_t=0,1,\cdots,t} P(r_t|\mathbf{h}_{1:t}). \tag{6}$$

The posterior probability in (6) is obtained by computing the joint probability $P(r_t \cap \mathbf{h}_{1:t})$ recursively [18];

$$\begin{aligned} &P(r_t \cap \mathbf{h}_{1:t}) \\ &= \sum_{r_{t-1}} P(r_t|r_{t-1})P(h_t|\mathbf{h}^{r_{t-1}})P(r_{t-1} \cap \mathbf{h}_{1:t-1}), \end{aligned} \tag{7}$$

where $\mathbf{h}^{r_{t-1}}$ denotes the data set associated with the run length r_{t-1}.

The modeling of the probability $P(r_t|r_{t-1})$ will be discussed in Section 3.2. The predictive probability $P(h_t|\mathbf{h}^{r_{t-1}})$ in (7) is associated with the data probability model through

$$P(h_t|\mathbf{h}^{r_{t-1}}) = \int P(h_t|\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathbf{h}^{r_{t-1}})d\boldsymbol{\theta}, \tag{8}$$

where $P(h_t|\boldsymbol{\theta})$ is the postulated data model with parameter set $\boldsymbol{\theta}$ and is always the same class of distribution as previously mentioned. In Bayesian approach, the parameter set $\boldsymbol{\theta}$ is assumed to be random so $P(\boldsymbol{\theta}|\mathbf{h}^{r_{t-1}})$ can be viewed as the prior distribution of $\boldsymbol{\theta}$ at time t. Consequently, the posterior probability $P(\boldsymbol{\theta}|\mathbf{h}^{r_t})$ resulting from the integrand in (8) will be used as the prior at time $t+1$. By using the prior that is *conjugate* to the data model, the resulting posterior probability is still in the same class of distribution as the prior [24]. Consequently, the integrand in (8) will always take a fixed form if the conjugate prior is applied, greatly reducing the computational complexity of evaluating (8).

3.2. Bird Phrase Segmentation by CPD

In the following, we will discuss how to apply the theory in [18] to our application. For bird phrase segmentation, the input data to the change point detection is the entropy sequence h_t from (2), which is assumed to be Gaussian. Hence, $P(h_t|\boldsymbol{\theta})$ is now a Gaussian pdf and $\boldsymbol{\theta}$ consists of the mean μ and the variance σ^2. Gaussian pdf has a natural conjugate prior since it belongs to the exponential family [24]. The conjugate prior of a Gaussian likelihood with mean μ and variance σ^2 is a Gaussian-inverse-gamma distribution,

$$\begin{aligned} &\mathcal{NIG}\left(\mu, \sigma^2 \,|\boldsymbol{\eta} = \{m, \tau, \alpha, \beta\}\right) \\ &= \frac{\left(\sigma^2\right)^{-\alpha-\frac{3}{2}}}{\Gamma(\alpha)} \frac{\beta^\alpha}{\sqrt{2\pi\tau}} \exp\left\{-\frac{(\mu-m)^2 + 2\tau\beta}{2\tau\sigma^2}\right\}. \end{aligned} \tag{9}$$

Namely, $\boldsymbol{\theta}$ is now parametrized by the *hyperparameter set* $\boldsymbol{\eta}$. As a result, using this conjugate prior we can obtain the posterior probability $P(\mu, \sigma^2|\mathbf{h}^{r_t})$ also in the form of $\mathcal{NIG}(\mu, \sigma^2|\boldsymbol{\eta}^{r_t})$, where $\boldsymbol{\eta}^{r_t}$ denotes the hyperparameter set that is updated by the observations $\mathbf{h}^{r_t}$ and thus can be viewed as the sufficient statistic of $\mathbf{h}^{r_t}$. Therefore, the predictive probability in (8) is the integral of the product of a Gaussian pdf and a Gaussian-inverse-Gamma pdf, which by a straightforward derivation can be shown to be a Student's t-pdf,

$$P(h_t|\mathbf{h}^{r_{t-1}}) = \mathcal{T}\left(h_t \,\middle|\, 2\alpha^{r_{t-1}}, \frac{\alpha^{r_{t-1}}}{\beta^{r_{t-1}}(\tau^{r_{t-1}}+1)}, m^{r_{t-1}}\right). \tag{10}$$

The Student's t-pdf is a function of three parameters,

$$\mathcal{T}(x|\nu, \lambda, m) = \sqrt{\frac{\lambda}{\pi\nu}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\lambda(x-m)^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{11}$$

By comparing the conjugate prior and the resulting posterior, the updating equations for the hyperparameter set $\boldsymbol{\eta}^{r_t}$ when observing the new data h_t can be easily derived as,

$$\begin{aligned} m^{r_t} &= \frac{\tau^{r_{t-1}} h_t + m^{r_{t-1}}}{\tau^{r_{t-1}} + 1},\ \tau^{r_t} = \frac{\tau^{r_{t-1}}}{\tau^{r_{t-1}} + 1}, \\ \alpha^{r_t} &= \alpha^{r_{t-1}} + \frac{1}{2},\ \beta^{r_t} = \beta^{r_{t-1}} + \frac{(h_t - m^{r_{t-1}})^2}{2(\tau^{r_{t-1}} + 1)}. \end{aligned} \tag{12}$$

Using the newly observed data h_t and the hyperparameters, we can obtain the predictive probability through (10) easily without performing numerical integration at every time instant.

The run length transition probability $P(r_t|r_{t-1})$ in (7) can be modeled if there is a prior knowledge of the distribution of the time between change points. Let T be a random variable defined as the time between contiguous change points, which is the length of a call period or a quiet period between calls. Therefore, the cumulative distribution function (CDF) $F_T(t)$ is bird species-dependent. The probability that a change point occurs at time t given the previous run length can be obtained in terms of $F_T(t)$

$$p(r_t = 0|r_{t-1}) = \frac{F_T(t) - F_T(t-1)}{1 - F_T(t-1)}. \tag{13}$$

Given r_{t-1}, r_t could only be 0 or $r_{t-1}+1$. Hence, $P(r_t = r_{t-1} + 1|r_{t-1})$ is simply $1 - p(r_t = 0|r_{t-1})$ and the run length transition probability is fully defined by $F_T(t)$. The empirical $F_T(t)$ can be estimated from the database of the target species.

Using (6), (7) and the probability models discussed above, once $\hat{r}_t$ is found to be 0, a change point at t is declared. After marking the change points, we need to determine if the segment between change points is a call period or a quiet period. The idea is to compare the time-averaged entropy of the segment with a threshold γ_h. If it is lower than γ_h, the segment is determined to be a call period; otherwise, it is a quiet period. The threshold γ_h should be adaptively adjusted over time, in the way similar to the one used in the energy detection [6]. The main difference between the proposed method with the thresholding method in energy detection is that our threshold is less sensitive to the short-term variation in the entropy due to the bursty noise, since it is updated based on the time-averaged entropy of predefined segments thanks to the change point detection.

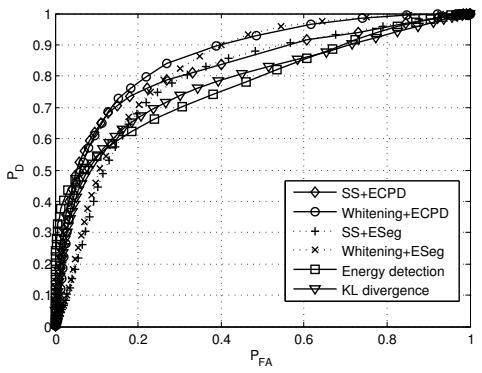

Fig. 2. ROC curves of the entropy-based CPD segmentation, the entropy-based segmentation with hard thresholding, energy detection and KL divergence segmentation.

4. EVALUATION

We evaluate the proposed segmentation method on recordings of Cassin's Vireo (*Vireo cassinii*) from a single territory. Thirteen separate recordings were obtained between 23 April and 8 June, 2010 in Amador county, California (38°29'0"N, 120°38'04"W) in a mixed conifer-oak forest at approximately 800 meters elevation. The length of each recording varies from 72 seconds to 551 seconds, and the total length is over 50 minutes. Manual annotation was performed to note the phrase class, and the start and end time of each phrase in the song. The phrases were categorized into one of the 63 phrase classes based on both visual examination of their spectrograms and auditory recognition. There were 852 phrases of Cassin's Vireo annotated in the recordings that are not severely overlapped with other species' calls. For the notable vocalizations by other species, the start and end time were labeled and are all classified as "others".

The spectrogram of the recordings were obtained by STFT with a Hamming window applied. FFT size is 512, and the frame hop size is 20% of the FFT frame size. The time length w of the time-frequency block for calculating entropy was set to 138.8ms. The frequency range of the block was set from $f_1 = 1.5$kHz to $f_F =$ 7kHz. The block rate is 144Hz.

First, we evaluate the proposed method by the effectiveness of capturing bird songs. Let $L(\cdot)$ be the length of a given time interval, and denote I_m and I_a as the bird phrase intervals (including the class "other") labeled by human and by the proposed method, respectively. Also, let I_m^C be those time intervals without any human labels. Define the detection rate P_D and the false alarm rate P_{FA} as

$$P_D = \frac{L(I_m \cap I_a)}{L(I_m)} \quad \text{and} \quad P_{FA} = \frac{L(I_m^C \cap I_a)}{L(I_m^C)}. \qquad (14)$$

The intersection here means the overlap between two intervals. Based on (14), the receiver operating characteristic (ROC curve) of the proposed method is plotted in Fig. 2. The "SS+ECPD" and "Whitening+ECPD" represent the entropy-based CPD segmentation with spectral subtraction and spectral whitening filter as the front-end processing, respectively. Using whitening filter as the front-end detects songs better than using spectral subtraction. This is because spectral subtraction is a spectral cleaning technique and it tends to remove those relatively low-energy background noise. Some weak bird phrases may also being removed from the spectrogram. In order to compare use of CPD over using a hard threshold to detect the changes of the entropy sequence, the results of using hard thresholding to replace CPD are also shown as "SS+ESeg" and "Whitening+ESeg." It is clear that in the low P_{FA} region, the detection rate of using CPD is significantly higher than using a hard threshold. The difference is insignificant at the high P_{FA} and high P_D region, which is generally not the desired operating region. This shows that by using CPD the system is able to detect more bird calls. The time-domain energy detection and the KL divergence method [9] are also shown for comparison. Both entropy-based CPD segmentation results outperform these two methods in every region of the ROC curves.

Table 1. Phrase Classification Rates of Sparse Representation-based (SR) and Support Vector Machine (SVM) Classifiers Training by ECPD and Human-annotated phrases (HA)

	Trained by ECPD		Trained by HA	
Testing set	SR	SVM	SR	SVM
SS+ECPD	81.97%	79.00%	79.55%	79.93%
Whitening+ECPD	81.04%	76.4%	77.33%	76.21%

The segmented phrases of Cassin's Vireo by the proposed method were also tested on the bird phrase classifier. The sparse representation-based (SR) classifier [25] and the support vector machine (SVM) classifier [26] were considered in the experiment. In SR classifier, 7 training tokens per phrase were used. Since not every phrase class has enough tokens for training and testing, 30 classes were considered in the classification experiment. The 7 training tokens for each phrase class were randomly chosen, while all the remaining tokens were used for testing. The dimension of the feature vector was set to 128. The multi-class SVM classifier was implemented using the LibSVM [27]. The classifier for each training set was trained using a five-fold cross-validation to search for an optimal pair of regularization factor.

In each classifier, two different training scenarios were considered. The first training set was chosen from the phrases segmented by the proposed method, and the other training set was chosen from the human-labeled phrases. The testing set was always selected from the phrases generated by ECPD. The classification results are listed in Table 1. As expected, the classification rates of the experiment using ECPD training set are mostly higher than the ones using human-annotated training set, since there are less mismatches between and training and testing data. However, the differences between the results of these two scenarios are all less than 4%, which implies that the ECPD phrases stay fairly close to the human annotated phrases. The classification rate is up to 81.97% which shows that combining the proposed method with phrase classifiers is promising in providing a reliable automated system for analyzing bird recordings.

5. CONCLUSION

We proposed a bird phrase segmentation method by entropy-based change point detection. To enlarge the difference between the entropy of a call period and the one of a quiet period, a polynomial-based whitening filter is proposed as the front-end of the segmentation to whiten the spectrogram of the background noise. Instead of using hard threshold, a Bayesian change point detection is used to monitor the statistical changes in the entropy sequence. Experimental results show that the proposed method is very effective on capturing bird calls. It is also shown to be practical to combine the proposed segmentation method with phrase classifiers. This automated system would facilitate the analysis of long field recordings.

6. REFERENCES

[1] D. J. Mennill, "Individual distinctiveness in avian vocalizations and the spatial monitoring of behaviour," *Ibis*, vol. 153, no. 2, pp. 235–238, 2011.

[2] V. M. Trifa, A. N. G. Kirschel, C. E. Taylor, and E. E. Vallejo, "Automated species recognition of antbirds in a Mexican rainforest using hidden markov models," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 2424–2431, 2008.

[3] A. Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, vol. 5, Apr. 2003, pp. V545–V548.

[4] C.-H. Lee, C.-C. Han, and C.-C. Chuang, "Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1541–1550, Nov. 2008.

[5] P. Somervuo and A. Harma, "Bird song recognition based on syllable pair histograms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, vol. 5, May 2004, pp. V825–V828.

[6] S. Fagerlund, "Bird species recognition using support vector machines," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, Jan. 2007.

[7] A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in recognition of bird sounds," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, Jan. 2007.

[8] P. Somervuo, A. Harma, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 6, pp. 2252–2263, Nov. 2006.

[9] B. Lakshminarayanan, R. Raich, and X. Fern, "A syllable-level probabilistic framework for bird species identification," in *International Conference on Machine Learning and Applications, 2009*, Dec. 2009, pp. 53–59.

[10] L. Neal, F. Briggs, R. Raich, and X. Fern, "Time-frequency segmentation of bird song in noisy acoustic environments," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2012–2015.

[11] C. Elemans, K. Heeck, and M. Muller, "Spectrogram analysis of animal sound production," *Bioacoustics : the International Journal of Animal Sound and its Recording*, vol. 18, no. 2, pp. 183–212, 2008.

[12] G. L. Patricelli, M. S. Dantzker, and J. W. Bradbury, "Acoustic directionality of red-winged blackbird (agelaius phoeniceus) song relates to amplitude and singing behaviours," *Animal Behaviour*, vol. 76, no. 4, pp. 1389–1401, 2008.

[13] A. Tartakovsky, B. Rozovskii, R. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," vol. 54, no. 9, pp. 3372–3382, 2006.

[14] Y. Chen, K. Hwang, and W.-S. Ku, "Collaborative detection of DDoS attacks over multiple network domains," vol. 18, no. 12, pp. 1649–1662, 2007.

[15] M. A. Osborne, R. Garnett, and S. J. Roberts, "Active data selection for sensor networks with faults and changepoints," in *International Conference on Advanced Information Networking and Applications*. Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 533–540.

[16] V. Spokoiny, "Multiscale local change point detection with applications to value-at-risk," *The Annals of Statistics*, vol. 37, no. 3, pp. 1405–1436, Jun. 2009.

[17] T. Roth, P. Sprau, M. Naguib, and V. Amrhein, "Sexually selected signaling in birds: A case for bayesian change-point analysis of behavioral routines," *The Auk*, vol. 129, no. 4, pp. 660–669, Oct. 2012.

[18] R. P. Adams and D. J. C. MacKay, "Bayesian online changepoint detection," *Technical Report, University of Cambridge*, Oct. 2007. [Online]. Available: http://arxiv.org/abs/0710.3742

[19] D. Li, I. K. Sethi, N. Dimitrova, and T. McGee, "Classification of general audio data for content-based retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 533–544, Apr. 2001.

[20] M. Spina and V. Zue, "Automatic transcription of general audio data: preliminary analyses," in , *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings*, vol. 2, Oct. 1996, pp. 594 –597 vol.2.

[21] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," vol. 27, no. 2, pp. 113–120, Apr. 1979.

[22] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," vol. 7, no. 2, pp. 126–137, Mar. 1999.

[23] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, May 2002, pp. IV–4164–IV–4164.

[24] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC, Jul. 2003.

[25] L. N. Tan, K. Kaewtip, M. L. Cody, C. E. Taylor, and A. Alwan, "Evaluation of a sparse representation-based classifier for bird phrase classification under limited data conditions," in *Proc. of 13th Annual Conference of the Iternational Speech Communication Association (IterSpeech 2012)*, 2012.

[26] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics*, vol. 4, no. 4, pp. 206–214, Sep. 2009.

[27] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.