

An Intelligent Nocturnal Animal Vocalization Recognition System

Heng-Ming Chen, Chenn-Jung Huang, You-Jia Chen, Chao-Yi Chen, and Sheng-Yuan Chien

Abstract—An intelligent nocturnal animal identification system is proposed in this work to assist in the recognition of different kinds of nocturnal animals. Firstly, each sound sample underwent the steps of the noise reduction and the syllable segmentation. Next, the segmentation was converted into Mel Frequencies Cepstral Coefficients (MFCCs) as the main identification feature of the proposed work. To increase the recognition accuracy and lower the computation overhead, the system can guide a user to input environmental information as the parameters of the decision tree classification module. It then proceeds with the clustering, and the results are classified with a composite classifier. A series of simulations were conducted to demonstrate the feasibility of the proposed algorithm. Drawing from 18 species, 408 sound samples, and 4477 feature segmentations were utilized to verify the accuracy rate of our work. Our simulations showed that the proposed approach can effectively increase the recognition rate and decrease the training time. In particular, the collected sound sample of the *Apus affinis* is only one that results in a lower accuracy rate. We expect to expand the database of sound samples, and thus we expect to be able to dramatically improve on the above-mentioned problem, thereby increasing the accuracy rate for each species.

Index Terms—Composite classifier, decision trees, feature extraction, machine learning, nocturnal animal vocalizations, noise reduction, syllable segmentation, support vector machines.

I. INTRODUCTION

In general, features of animal vocalizations include time domain and frequency domain. Time domain features are calculated directly from the sound waveform, and include such identifying characteristics as zero crossing rate and signal energy. A time domain signal is first transformed to the frequency domain using a Fourier transform and new features are thereby derived from the transformed frequency signal.

In this work, an automatic identification system for mixed nocturnal animal vocalizations is proposed to recognize various nocturnal animal species based on recorded audio signals that were sampled from recordings of nocturnal animal sounds in an outdoor environment. The sampled signals were first converted into frequency signals. Subsequently, syllable segmentation and feature selection methods were employed to separate the original nocturnal animal calls into syllables and to derive the input feature sets for the classifiers. Experimental results reveal that the

proposed identification system is effective in identifying mixed nocturnal animal vocalizations in real-time. To the best of our knowledge, this work is the first research that addresses the problem of real-time identification of nocturnal animal vocalizations in the literature.

The remainder of this article is organized as follows. In Section II, we describe related work in audio recognition. The architecture of the intelligent identification system for nocturnal animal vocalizations is presented in Section III. Experimental results and analysis are given in Section IV. Finally, Section V concludes the work.

II. RELATED WORK

Features used in sound recognition applications are usually chosen such that they represent certain meaningful characteristics. The selection of the actual features used in recognition is a critical element of the recognition system. Nocturnal animal sounds can be taken as an organized sequence of brief sounds from a species-specific vocabulary. Those brief sounds are usually called syllables. Through the use of pre-processing, we can extract those useful syllables and separate the nocturnal animal calls from sounds which are not made by non-nocturnal animals.

Recently, most of the research on sound recognition for animal calls has focused on animal species identification, such as that of bird species [1], [2]. Identification of different species of animals based on recorded calls has benefitted people's understanding of animal calls. Vilches *et al.* [3] used data mining techniques for classification, and analyses were performed on a pulse-by-pulse basis, in contrast to traditional syllable-based systems. Somervuo *et al.* [2] studied different parametric representations of birdsong syllables. Notably, all of the above-mentioned research work focused on bird calls. As it turns out, however, the collection of bird call samples is not inherently difficult because most of the birds can be seen and their calls heard during the daytime. Moreover, voice recognition technology (VRT) has become accurate and fast enough to be useful in a variety of writing scenarios. The variation of the vocalization characteristics of a particular animal may be classified in two ways. One concerns the variation among different individual animals, while the other is concerned with the variation within the same animal. The former can be utilized for the recognition of individuality, and the latter for monitoring of the animal's condition [4]. In birds, vocal cues appear to be a major component of recognition. Curiously, vocal recognition seems absent or limited in some highly social species such as the Black-legged Kittiwake and Rissa Tridactyla [5]. The ability to identify individuals reliably is often a key prerequisite for animal behaviour studies in the wild. In primates, recognition

Manuscript received June 13, 2014; revised August 25, 2014. This work was financially supported in part by the National Science Council of Taiwan under contract numbers NSC 102-2511-S-259-010-MY3 and 99A058, respectively.

The authors are with the National Dong Hwa University, Taiwan (cjhuang@mail.ndhu.edu.tw).

of other group members can be based on individual differences in the voice, but these cues are typically too subtle for human observers to perceive [6]. Colonial living imposes strong selection pressures on the communication systems of species, with many animals communicating on the same sensory channels simultaneously. Colonial species often exhibit complex individual vocal signatures that encode a particular caller's identity in their vocalizations [7].

In sound recognition applications, the call boundary detection is an essential problem to be resolved. To a large extent, the performance of sound recognition depends on whether a call boundary detection algorithm can accurately detect the end points of the syllables. This construct of segmentation before the recognition process assists the system in finding meaningful and complete calls. Proper segmentation of nocturnal animal calls can shed light on the characteristic of the calls, thereby leading to higher accuracy in nocturnal animal recognition.

Many researchers [8], [9] have proposed call boundary detection algorithms in the literature in recent years. These algorithms work effectively in noise-free conditions. Most of these algorithms, including short-time energy [8], [10], short-time amplitude [11], and short-time zero-crossing rate [12], adopt short-time speech signals as the basis of discriminating between voiced and unvoiced speech signals, and thus can detect the end points easily from input signals. Nevertheless, the above-mentioned end point detection algorithms fail to work effectively in noisy environments.

Some investigations focused on recognition of different animal calls [13], [14]. Mitrovic *et al.* [13] used machine learning technology to recognize different animal calls, including those of birds, cats, cows and dogs. Guodong *et al.* [14] employed support vector machines to identify 16 different classes of animal sounds. In this work, the nocturnal animal vocalization detection process is different from the above approaches. We first separate nocturnal animal vocalizations into different groups by means of a decision tree technique [15]. After the scope of the identified targets is narrowed, we then apply a composite classifier approach to identify the precise species of the nocturnal animal in question. This kind of two-level identification of different animal calls, to the best of our knowledge, has never been investigated in the literature.

III. ARCHITECTURE OF THE IDENTIFICATION SYSTEM FOR NOCTURNAL ANIMAL VOCALIZATIONS

The architecture of the identification system for nocturnal animal vocalizations, as shown in Fig. 1, can be divided in five main modules, namely: signal preprocessing, syllable segmentation, feature extraction, decision tree clustering, and a composite classifier. Undesirable information is first removed from the raw input signals in order to preserve the desired characteristics of nocturnal animal calls during the signal preprocessing stage. The resulting signal is then segmented via the syllable segmentation method and the segmented syllables are further processed with the Mel Frequencies Cepstral Coefficients (MFCC) method in order to extract the features that are suitable for the classifier used in the present research. Notably, decision trees are used to separate the nocturnal animals into different clusters, to ease

the burden of the classifier. A composite classifier is chosen in this work because it is a powerful tool for classification and prediction by combining different classifiers, such as support vector machine (SVM), k -th nearest neighbors, and back propagation neural network into the voting scheme.

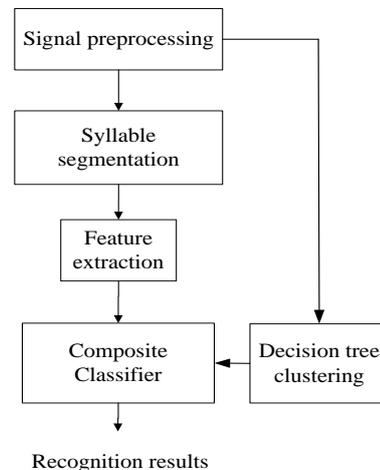


Fig. 1. Architecture of the identification system for nocturnal animal vocalizations.

A. Signal Preprocessing

The recorded sound signal is resampled at 44.1k Hz frequency and saved in 16-bit mono format. The amplitude of each sound signal is normalized within the range $[-1, 1]$ for ease of further processing. A de-noise filter [16] is first employed to remove noise during the signal analysis. Then, Fourier analysis and spectral noise gating [17] are applied to retrieve the de-noise parameters and “purify” the data in the noisy environment.

Time frequency block thresholding method was proposed by Yu *et al.* [18] to dynamically adjust spectrogram filter parameters via using the Stein risk estimator, which gives an indication of the estimator's accuracy. It was thus used as a pre-processing filter for the Speech Tutor corpus. Here, we applied the block thresholding method of Yu *et al.* to deal with the signal preprocessing step. The time-frequency de-noising procedures calculate a short-term Fourier transform (STFT) of the noisy signals, and operate the resulting coefficients to weaken the noise signals.

B. Syllable Segmentation

A syllable is basically a sound that a nocturnal animal produces with a single blow of air from the lungs. The rate of events in nocturnal animal vocalization is so high that the separations of individual syllables are difficult in a natural environment due to reverberation. Once the syllables have been properly segmented, a set of features can be collected to represent each syllable. In this work, an end-point detection method is used to produce syllable segments, since a cleaner waveform can be obtained by accurately measuring the positions of the two endpoints of the waveforms. In conjunction with energy measurement, zero-crossing rate is used for detecting the start and end positions of each syllable segment in this work.

1) Energy measurement (EM)

Energy measurement (EM) is used to detect the energy change during a short-term period in order to distinguish

voiced and unvoiced speech. There are three ways to calculate the EM, including the sum of squares, the sum of mean squares, and the sum of absolute values, and EM is the most appropriate method in the real-time system is adopted in this work in order to decrease the overall computational time.

The zero-crossing rate (ZCR) is equal to the number of zero-crossings of the waveform within a given frame, that is, the rate where the signal changes from positive to negative or back. This feature has been used widely in both speech recognition and music information retrieval, and is regarded as a significant feature by which to classify percussive sounds.

2) R-S Endpoint detection

At first, the energy measurement curve $E(n)$ is given, and then we can draw a horizontal line by setting the energy threshold Thd_E , recording the EM curve as the points P1 and P2, as shown in Fig. 2. We can use these two points as the start and end points of the segment for recording signals.

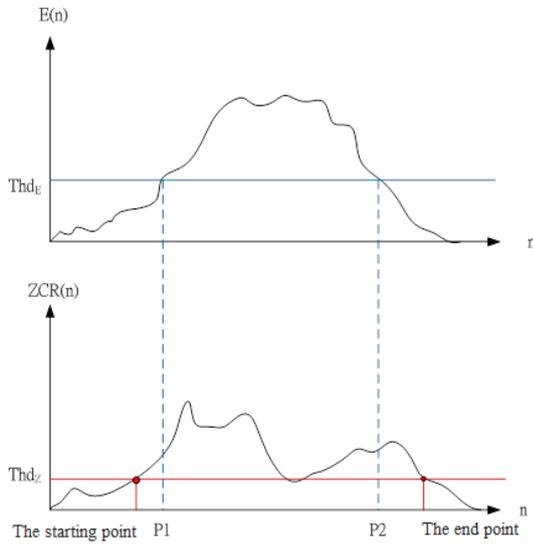


Fig. 2. The diagram of the energy measurement and the threshold.

Notably, the reason why we only set up P1 and P2 as the initial result of endpoint detection is that the signals between these points are probably a part of the voiced sounds for recording signals that cannot effectively be used to discover the unvoiced or fricative sounds. To obtain accurate starting and end points of a sound segment, we have to utilize the ZCR curve and a given threshold value Thd_Z to form a horizontal line.

As shown in Fig. 2, the accurate coordinates of end points for the sound signals are extended to both sides from P1 and P2.

Next, we collate the P1 in the upper part of Fig. 2 with the ZCR curve in the lower part of Fig. 2, and we can decide the starting point according to the following criteria.

- 1) If the ZCR is lower than Thd_Z in this frame, P1 is decided as the starting point for this segment.
- 2) Otherwise, if the ZCR is higher than Thd_Z , we will search towards the left side, and compare the ZCR in the previous frame to check if there is another ZCR higher than Thd_Z . When the first ZCR that is lower than Thd_Z is found, the corresponding time in this frame is the starting point of this voiced sound.
- 3) The method to search for the end point of the voiced

sounds is similar with II; the difference is that the search proceeds toward the other side, that is, it is extended to the right side from P2.

C. Feature Extraction

For this section, an algorithm called Mel-Frequency Cepstral Coefficients (MFCC) is used to extract features. The aim of this transformation process is to obtain a new voice representation that is more compact, less redundant, and more suitable for statistical modeling. The MFCC is based on the known variation of the human ear's critical bandwidths with regard to frequency, which means it filters the space linearly at low frequencies and logarithmically at high frequencies. MFCC is used in order to capture the phonetically important characteristics of the voice. There are several steps in order to implement the MFCC, which are as follows:

Step 1: Pre-emphasis

A pre-emphasis filter is utilized here to flatten the speech spectrum before spectral analysis. Its purpose is to compensate for the high-frequency part of the speech signals being suppressed during the human sound production mechanism.

The most commonly used filter is a high-pass Finite Impulse Response (FIR) filter, described as follows:

$$H(z) = 1 - a \times z^{-1}, \quad (1)$$

where a is between 0.9 and 1.0.

Step 2: Frame Blocking

The framing process is first applied to the voice signal of the producer. This signal is partitioned or blocked into N segments (frames) which mean that each frame can be analyzed independently and represented by a single feature vector.

We allow overlap between neighboring frames to reduce discontinuity between them. For example, in order to accomplish the frame blocking, we usually use a 20-25 ms window applied at 10 ms intervals (frame rate of 100 frames/s and overlap between adjacent windows of about 50%).

Step 3: Windowing

The third stage of the processing is to window each of the individual frame such as to minimize the signal discontinuities of the speech signal at the edges of each frame. The most commonly used window is a Hamming window.

Step 4: Fast Fourier Transform

The next stage in the process is the Fast Fourier Transform (FFT), a computationally efficient algorithm of the Discrete Fourier Transform (DFT), where each frame of N samples is converted from time domain to frequency domain.

Step 5: Triangular Bandpass Filters

We compute the energy of each triangular filter band as follows:

$$mel(f) = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right), \quad (2)$$

Step 6: Discrete Cosine Transform

We apply DCT on the 20 log energy E_k obtained from the triangular bandpass filters to get L Mel-scale cepstral coefficients.

Since we have already performed FFT, DCT transforms the frequency domain into a time-like domain called quefrency domain. The obtained features are similar to the cepstrum, and thus are referred to as the Mel-scale cepstral coefficients, or MFCC. MFCC alone can be used as the feature for speech recognition. For better performance, we can add the log energy and perform a delta operation.

D. Decision Tree Clustering

The purpose of decision tree clustering is to narrow the scope of the identification targets, lower the complexity of computation, and enhance the recognition accuracy. A decision tree clustering, as illustrated in the example in Fig. 3, can be constructed in this work based on the parameters extracted from the collected samples, including atmospheric temperature, water temperature, altitude, humidity, month, county location, and habitat. Each leaf node enumerates the possible occurrence of nocturnal animal vocalizations for the combination of the collected parameters. The motivation for using a decision tree here is that it is a powerful and widely used tool for classification and prediction. Moreover, being a rule induction approach, it allows for the represented rules to be readily be expressed and understood [15].

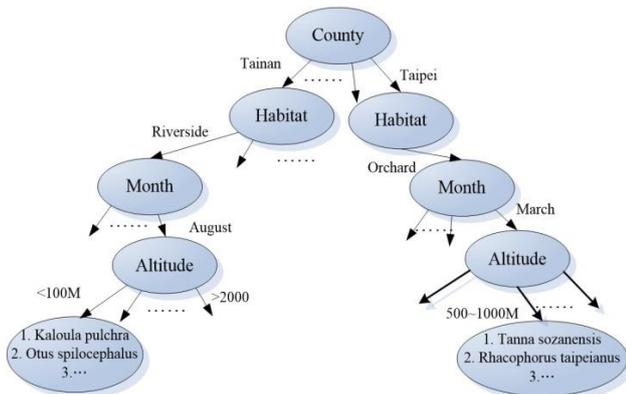


Fig. 3. An example of the decision tree clustering constructed in this work.

E. Classification Using Voting Machine Technique

Voting is a well-known aggregation procedure that combines different opinions of voters into consensus. In the simplest form of voting method, each data item is assigned a number of votes. Since there are advantages and disadvantages for different classification algorithms, we thus try to combine support vector machine (SVM), k -th nearest neighbors, and back propagation neural network into the voting scheme in this paper in order to classify the nocturnal animal calls.

1) Support vector machines

In a decade, support vector machines (SVMs) [19], [20] have attracted much attention as a new classification technique with good generalization ability. The basic idea of SVMs is to map input vectors into a high-dimensional feature space and linearly separate the feature vectors with an optimal hyper-plane in terms of margins, i.e. distances of given examples from a separating hyper-plane.

2) K -th nearest neighboring

The k -th nearest neighbor (KNN) algorithm [21], [22] is a classification algorithm based on closest training example feature space. The training phase of the algorithm consists of

storing the feature vectors and class labels of the training samples. In the actual classification phase, the same features are computed as before for the test sample whose class is not known. Distances from the new vector to all stored vectors are computed and k closest samples are selected. The new point is predicted to belong to the most numerous class within the set.

The KNN method is a simple yet effective method for classification in the areas of pattern recognition, machine learning, data mining, and information retrieval. It has been successfully used in a variety of real-world applications. The performance of KNN can be comparable with the state-of-the-art classification methods with simpler computation.

3) Back propagation neural network

A neural network is able to recognize patterns and generalize from them. An essential feature of this technology is that it improves its performance on a particular task by gradually learning a mapping between inputs and outputs. Generalization is used to predict the possible outcome for a particular task. This process involves two phases known as the training phase and the testing phase. Back propagation (BP) neural network [23], [24] is one of neural networks most in common use. In back-propagation, the learning procedure basically follows that of a traditional feed-forward neural network. However, there are two main differences. The first difference is the use of the activation function of the hidden unit y_j , and the second is that the gradient of the activation function is contained.

A back-propagation neural network consists of several layers of nodes including an input layer, one or more hidden layers and an output layer. Each node in a layer receives its input from the output of the previous layer nodes. The connections between nodes are associated to synaptic weights that are iteratively adjusted during the training process. Each hidden and output node is associated to an activation function. Several functions can be used as activation functions, but the most common choice is the sigmoid function.

Provided that the activation function of the hidden layer nodes is non-linear, a back-propagation neural network with an adequate number of hidden nodes is able to approximate every non-linear function. The adjustment of the synaptic weights in an error back-propagation algorithm consists of four steps:

- 1) The network is initialized by assigning random values to synaptic weights.
- 2) A training pattern is fed and propagated forward through the network to compute an output value for each output node.
- 3) Actual outputs are compared with the expected outputs.
- 4) A backward pass through the network is performed, changing the synaptic weights on the basis of the observed output errors.

Steps i through iv are iterated for each pattern in a training set until convergence.

F. Web Interface Design of Nocturnal Animal Identification System

A screen capture snapshot of the homepage for the on-line nocturnal animal identification system is shown in Fig. 4. The

button highlighted in the upper left hand corner of the webpage is used to pop out the recognition window. The main portion of the screen provides the list of nocturnal animal species. The second window can be popped out by double-clicking any item on the list to get more information for that corresponding species. The first part is the sample file uploading area, which is located in the upper portion of the window. In the recognition window, as shown in the upper half of Fig. 5, the users can press the “Start recognition” button to activate the classification process when the sample file is completely uploaded and the parameters are provided. The lower half area of the window displays the recognition result, including nocturnal animal species and their characteristics, e.g., photos, which can be adopted by the users to collect the information for the nocturnal animals.



Fig. 4. The website containing the on-line nocturnal animal vocalization identification system.



Fig. 5. Screen capture snapshot of the recognition result web page.

IV. EXPERIMENTAL RESULTS

A series of experiments were conducted to verify the effectiveness of the proposed online identification system for a mixture of nocturnal animal vocalizations. The experiments were implemented in Java (Version 7 Update 25). The test platform was an Intel Pentium(R) Dual-Core E5300 2.60GHZ CPU, 4-G RAM with Windows 7 Professional operating system. The audio signals collected include the 339 samples of 18 nocturnal animals, from which 3546 syllable segments were extracted in this work. Table I. lists the numbers of the nocturnal animal samples and syllable

segments used in our experiments. We applied the leave one out cross validation (LOOCV) technique [25] to remove a single sample and train on the rest, and then tested the decision tree clustering and composite classifier techniques on the removed single sample in each trial.

TABLE I: NUMBERS OF THE NOCTURNAL ANIMAL SAMPLES AND SYLLABLE SEGMENTS

Species	Number of samples	Number of syllable segments
Hyla chinensis	4	98
Rhacophorus taipeianus	31	182
Microhyla fissipes	67	876
Buergeria japonica	34	235
Fejervarya limnocharis	32	331
Rhacophorus prasinatus	21	298
Kaloula pulchra	10	98
Hoplobatrachus rugulosus	4	52
Babina okinavana	11	387
Hylarana guentheri	21	132
Kurixalus idiotocus	94	701
Caprimulgus affinis	1	11
Apus affinis	1	20
Tanna sozanensis	1	4
Cryptotympana takasagona	2	42
Otus spilocephalus	3	40
Gryllus bimaculatus	1	29
Gorsachius melanolophus	1	10
Total	339	3546

We first examined the performance of the decision tree clustering. The total recognition rate was 62.45% before decision tree clustering was applied, and the total recognition rate was significantly improved to 92.08 % after decision tree clustering had been applied. Notably, the recognition rate used in this work is defined as the sum of the number of true positives and true negatives divided by total number of samples.

We next evaluated the effectiveness of the composite classifier. The recognition results are listed in Table II. Notably, the performance metrics used in this work include true positive rate (TPR), false positive rate (FPR), precision, recall, and F-measure. TPR is also known as recall, and FPR is one minus the specificity or true negative rate. Precision for a class is the number of true positives divided by the sum of true positives and false positives. Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class. As for F-measure, it is a measure that combines precision and recall and is the harmonic mean of precision and recall,

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{1}$$

It can be seen from Table II that for 17 out of 18 species, the system achieved satisfactory performance. It can be seen

that the true positive rate and recall for *Apus affinis* is only 66.7%. We observed that only one sample for *Apus affinis* was collected. Although this sample was separated into 20 syllable segments, the overlap of signals among syllable segments was significant. This is the reason that the recognition rate for this species was significantly lower than for other species. We believe that the situation could be improved if more samples and syllable segments were collected for this species in the future. In addition, although only one sample was collected for several other species, including *Caprimulgus affinis*, *Tanna sozanensis*, *Gryllus bimaculatus*, and *Gorsachius melanolophus*, the recognition rates for these species were much better because there was no overlap among different syllable segments, and accordingly, the syllable segments for these species could be more easily separated and recognized.

TABLE II: RECOGNITION RESULTS OF NOCTURNAL ANIMAL VOCALIZATIONS

Species	TPR	FPR	Precision	Recall	F-Measure
<i>Hyla chinensis</i>	1	0.001	0.952	1	0.976
<i>Rhacophorus taipeianus</i>	0.882	0.003	0.928	0.886	0.901
<i>Microhyla fissipes</i>	0.876	0.018	0.872	0.878	0.869
<i>Buergeria japonica</i>	0.798	0.017	0.835	0.798	0.812
<i>Fejervarya limnocharis</i>	0.919	0.011	0.922	0.918	0.928
<i>Rhacophorus prasيناتus</i>	0.861	0.009	0.875	0.859	0.862
<i>Kaloula pulchra</i>	0.864	0.001	0.912	0.861	0.886
<i>Hoplobatrachus rugulosus</i>	1	0.001	0.92	1	0.958
<i>Babina okinavana</i>	0.781	0.023	0.819	0.762	0.799
<i>Hylarana guentheri</i>	0.922	0.008	0.832	0.924	0.875
<i>Kurixalus idiotoocus</i>	0.942	0.023	0.892	0.941	0.925
<i>Caprimulgus affinis</i>	1	0	1	1	1
<i>Apus affinis</i>	0.667	0	1	0.667	0.8
<i>Tanna sozanensis</i>	1	0	1	1	1
<i>Cryptotympana takasagona</i>	1	0	1	1	1
<i>Otus spilocephalus</i>	0.842	0	1	0.842	0.914
<i>Gryllus bimaculatus</i>	1	0	1	1	1
<i>Gorsachius melanolophus</i>	1	0	1	1	1

V. CONCLUSIONS

An intelligent identification system for nocturnal animal vocalizations is proposed in this work, with the aim of effectively recognizing nocturnal animal vocalizations in the outdoor environment. With this system, the public can upload recordings of unidentified nocturnal animal vocalizations to the nocturnal animal vocalization identification website to request the discrimination of the nocturnal animal species based on the uploaded samples. A series of experiments were performed to verify the effectiveness of the proposed work. Experimental results verified the effectiveness of decision tree clustering and the composite classifier in the application of identification of nocturnal animal vocalizations.

Since the recorded nocturnal animal vocalizations in rural fields are usually chorus audio samples with unbalanced volume, tone and rhythm, the quality of the collected samples is usually poor. In addition, surrounding noise in the outdoor

environment further increases the inherent difficulty in collecting samples of nocturnal animal vocalizations. In future work, we intend to use multiple microphones or adopt some automatic recording method at fixed locations to collect abundant samples with good quality for use in the training of classifiers, and thus expect to be able to achieve a higher accuracy rate in recognizing the mixture of nocturnal animal vocalizations.

REFERENCES

- [1] H. Tyagi, R. M. Hegde, H. A. Murthy, and A. Prabhakar, "Automatic identification of bird calls using spectral ensemble average voiceprints," in *Proc. the 13th European Signal Processing Conference*, Florence, Italy, 2006.
- [2] P. Somervuo, A. Hämäläinen, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 6, pp. 2252–2263, 2006.
- [3] E. Vilches, I. A. Escobar, E. E. Vallejo, and C. E. Taylor, "Data mining applied to acoustic bird species recognition," in *Proc. the 18th International Conference on Pattern Recognition*, vol. 3, pp. 400–403, Hong Kong, 2006.
- [4] Y. Ikeda, Y. Ishii, "Recognition of two psychological conditions of a single cow by her voice," *Computers and Electronics in Agriculture*, vol. 62, pp. 67–72, 2008.
- [5] H. Mulard, T. Aubin, J. L. F. White, S. A. Hatch, and E. T. Danchin, "Experimental evidence of vocal recognition in young and adult black-legged kittiwakes," *Animal Behavior*, vol. 76, pp. 1855–1861, 2008.
- [6] A. Mielke and K. Zuberbühler, "A method for automated individual, species and call type recognition in free-ranging animals," *Animal Behavior*, pp. 1–8, 2013.
- [7] B. J. Pitcher, R. G. Harcourt, and I. Charrier, "Individual identity encoding and environmental constraints in vocal recognition of pups by Australian sea lion mothers," *Animal Behavior*, vol. 83, pp. 681–690, 2012.
- [8] J.-C. Janqua, "A robust algorithm for word boundary detection in the presence of noise," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 3, pp. 406–412, 1994.
- [9] Y. Xing, X. Wu, and Z. Xu, "Multiclass least squares auto-correlation wavelet support vector machines," *ICIC Express Letters*, vol. 2, no. 4, pp. 345–350, 2008.
- [10] F. Beritelli, "Robust word boundary detection using fuzzy logic," *Electronics Letters*, vol. 36, no. 9, pp. 846–848, 2000.
- [11] L. Liu, J. He, and G. Palm, "Effects of phase on the perception of intervocalic stop consonants," *Speech Communication*, vol. 22, pp. 403–417, 1997.
- [12] N. Erdöl, C. Castelluccia, and A. Zilouchian, "Recovery of missing speech packets using the short-time energy and zero-crossing measurements," *IEEE Trans. on Speech and Audio Processing*, pp. 295–303, 1993.
- [13] D. Mitrovic and M. Zeppelzauer, "Discrimination and retrieval of animal sounds," in *Proc. the IEEE Multimedia Modelling Conference*, 2006.
- [14] G. G. and Z. Li., "Content-based classification and retrieval by support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, pp. 209–215, 2003.
- [15] R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [16] L. Donoho, "De-noise by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, 1995.
- [17] D. M. Klapachinski, C. R. E. Lima, C. A. Alves Kaestner, "Spectral noise gate technique applied to birdsong preprocessing on embedded unit," in *Proc. 2012 IEEE International Symposium on Multimedia*, 2012, pp. 24–27.
- [18] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 1830–1839, 2008.
- [19] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
- [20] C.-C. Chang and C.-J. Lin. (2001). LIBSVM: A library for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [21] M. R. Peterson, T. E. Doornik, and M. L. Raymer, "GA-facilitated KNN classifier optimization with varying similarity measures," *IEEE Congress on Evolutionary Computation*, vol. 3, pp. 2514–2521, 2005.

- [22] J. D. Kelly and L. Davis, "Hybridizing the genetic algorithm and the k nearest neighbors classification algorithm," in *Proc. 4th Int. Conference: Genet. Algorithms & Applications*, pp. 377-383, 1991.
- [23] L. Medsker and J. Liebowitz, *Design and Development of Expert Systems and Neural Networks*, New York: Macmillan, 1994.
- [24] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [25] S. Fagerlund, "Bird species recognition using support vector machines," *EURASIP Journal on Applied Signal Processing*, 2007.



You-Jia Chen is pursuing his doctoral degree at the Department of Electrical Engineering, National Dong Hwa University, Taiwan. His research interests include computer communication networks, data mining, applications of machine learning techniques and e-learning.



Heng-Ming Chen is pursuing his doctoral degree at the Department of Electrical Engineering, National Dong Hwa University, Taiwan. His research interests include computer communication networks, data mining, applications of machine learning techniques and e-learning.



Chao-Yi Chen is pursuing his master's degree at the Institute of Electronics Engineering, National Taiwan University, Taiwan. His research interests include computer communication networks, data mining and applications of machine learning techniques.



Chenn-Jung Huang received the B.S. degree in electrical engineering from National Taiwan University, Taiwan and the M.S. degree in computer science from University of Southern California, Los Angeles, in 1984 and 1987, respectively. He received the Ph.D. degree in electrical engineering from National Sun Yat-Sen University, Taiwan in 2000. He is currently a professor in the Department of Computer Science & Information Engineering, National Dong Hwa University, Taiwan. His research interests include computer communication networks, data mining, and diagnosis agent for e-learning.



Sheng-Yuan Chien is pursuing his master's degree at the Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan. His research interests include computer communication networks, data mining and applications of machine learning techniques.

Data Mining and Algorithm

