ACOUSTIC MONITORING OF SINGING INSECTS

†Todor Ganchev, ‡Ilyas Potamitis, †Nikos Fakotakis

†Department of Electrical and Computer Engineering, University of Patras, Greece, tganchev@wcl.ee.upatras.gr,
‡Department of Music Technology and Acoustics, Technological Educational Institute of Crete, potamitis@stef.teicrete.gr

ABSTRACT

This work reports recent results towards the development of a pilot automatic identification system for insect species. We propose a novel sound parameterization technique that is combined with state-of-the-art classification methods that dominate speaker recognition technology. The categorization of acoustic emissions of insects takes place on the levels of specific singing insect, family, subfamily, genus and species. We evaluate the performance of our approach on a large and well-documented catalogue of recordings of crickets, katydids and cicadas. Identification accuracy that exceeds 98% on the levels of singing insect and family, and 86% on the level of specific species out of 313 species is reported.

Index Terms— Biomedical acoustics, biological control systems, animals

1. INTRODUCTION

Although a considerable number of studies have been devoted to the acoustic communication of insects (see [1] and the references therein), automated species identification has been considered just a marginal field of pattern recognition while literature on this subject is sparse. In brief, acoustic identification of insects is based on their ability to generate sound either deliberately, as a means of communication (see section 2), or as a by-product of eating, flying or locomotion. Provided that the bioacoustic signal produced by insects follows a consistent acoustical pattern that is species-specific, it can be exploited for detection and identification purposes.

The practical significance and potential of automated identification of insect species lies in the following facts:

a) Insects have great economic importance as beneficial organisms in agriculture and forestry (they play a significant role in the food chain of other species and the fertility of plants). However, a number of insect species also are a negative contribution to agricultural economy as they constitute a devastating threat to plants and crops.

b) The manual detection and identification process is in most cases a highly complex procedure, because insects are heard more often than seen or trapped (especially those that live in complex environments or demonstrate nocturnal activity).

c) The development of human expertise to capture taxonomic information is costly both in time and money and requires the construction of expensive reference collections of fragile insect specimens and comprehensive literature sources [2].

d) Non-experts have great difficulty practicing taxonomy while participating in the construction of biological inventories even for routine identifications.

e) The diversity of whole animal communities is endangered by urban expansion. The existence and density of the population of certain species is directly dependent on pollution levels, climatic change and urban design. Therefore, inventorying and monitoring of certain species is a way to identify disturbance and biodiversity unbalance in a non-intrusive way.

It is well documented that insect songs provide a reliable taxonomic clue and can be used to measure biodiversity [3]. In this article, we propose a signal parameterization method inspired by the feature extraction process of speaker/speech recognition and designed explicitly for the needs of acoustic insect recognition. Moreover, the classifiers that have been tested consist the core of automatic recognition equipment. However, the signal processing methods involved in these speech processing-related tasks are adapted here for the specific task of insect recognition. Our motivation for investigating the parallelism between speech/speaker recognition and insect recognition lies on the fact that besides the difference in the sound production mechanism, bioacoustic signal recognition, in general, shares common ground with speech recognition. Thus, utilizing the experience accumulated in contemporary automatic speech recognition technology, which has reached the maturity of launching commercial products, we aim at a boost of performance of the existing automatic bioacoustic recognition equipment.

The novel application area that is related to this work includes:

a) automatic environmental monitoring and inventorying of the biological diversity of a designated area,

b) viability analysis of endangered populations,

c) habitat health assessment and deterioration as certain species are indicators of habitat quality and conservation,

d) detection and early warning of pests that are dangerous for agriculture,

e) the recognition and taxonomy of a wide range of taxa by non-specialists.

The present work reports identification results on the louder insects (i.e., crickets, cicadas and katydids). We aim at identifying specific families and subfamilies of insects, as well as identifying the particular species. We evaluate our equipment on the singing insects of the North America collection (SINA) [4] that have been tagged by scientists of considerable experience in identifying the taxonomy of insects. Progressively, this equipment will be extended to other living organisms that are able to produce consistent acoustic patterns. The long-term goal is a network of autonomous recording stations reporting data by employing wireless streaming to a central processing point.

2. HOW AND WHY DO INSECTS COMMUNICATE

The sound production mechanism in insects can be summarized as: muscle power contraction leading to mechanical vibration of the sound-producing structure and finally to acoustic loading of this source and sound radiation [5-6]. Sound is produced by insects in five different ways [7]:

1. **Stridulation:** the friction of two body parts; usually heard as

chirping, i.e., (crickets, katydids, grasshoppers, bugs, beetles, moths, butterflies, ants, caterpillars, beetle larvae, others)

2. **Percussion:** by striking some body part, such as the feet (band-winged grasshoppers), the tip of the abdomen (cockroaches), or the head (death-watch beetle) against the substrate usually heard as tapping or drumming

3. **Vibration:** the oscillation of body parts such as wings; usually heard as humming or rumbling by vibrating some body part, eg., wings, in air (mosquitoes, flies, wasps, bees, others)

4. **Tymbal Mechanism:** the quick contraction and release of tymbal muscles (vibrating drum-like membranes); usually heard as a series of clicking sounds (cicadas, leafhoppers, treehoppers, spittlebugs)

5. **Air Expulsion:** the ejection of air or fluid through a body constriction; usually heard as a whistle or hiss (short-horned grasshoppers).

There is a specific number of behavioural modes that have been observed in connection with sound production in insects. In particular, males, females and immature insects produce acoustic emissions that can be classified in four distinct categories [7]:

a) *The congregational songs.* The congregational song is a song produced in chorus and its main purpose is to cause male and female adults to congregate (cicadas).

b) *The calling songs.* These constitute the first step in pair formation and are used to attract females at long range into close proximity (e.g. crickets and cicadas produce 'mating' songs). Some females also produce a sound that will help the male to locate her (slant-faced grasshoppers) or in response to the males (katydids)

c) *The courtship songs.* These are produced at short range by males and aim to attract a responsive female before mating. Singing males switch from calling singing to courtship singing as soon as a female approaches within one meter (cicadas).

d) *The protest squawks.* These sounds declare disturbance either because the insect is captured or disturbed in flight or the male insect wants to let other males know that they are in his territory, (generally called 'warning', 'intimidation' or 'fight' sounds). This sound response can also be used to warn other insects of danger. The squawks are more or less arythmical. They are however species-specific as the frequencies and the rate of vibration of the tymbals (in cicadas) and stridulation organs (crickets, katydids) are distinct.

2.1. Crickets

Male crickets produce sounds by stridulating (by rubbing their wings together). They produce a short repertoire of consistent acoustic patterns, which are characterized by a modulation around a dominant frequency. Their sound pattern consists of pulsations, well localized both in time and frequency. In some species these impulsive sounds form packets (phrases), which are repeated rhythmically (see Fig. 1(a)). Finally, the pulsations per unit time are dependent on the environmental settings (e.g., temperature, humidity) while the fundamental remains fairly unchanged even in different behavioural modes.

2.2. Cicadas

Cicadas possess an auditory apparatus, while their hearing is tuned to receive the frequency range they produce. The acoustic pattern of cicadas is characterized by groups of pulses with a distinguishable amplitude modulation pattern (see Fig. 1(b)). The emitted sound covers the frequency range [2 kHz, 22 kHz]. The sound is
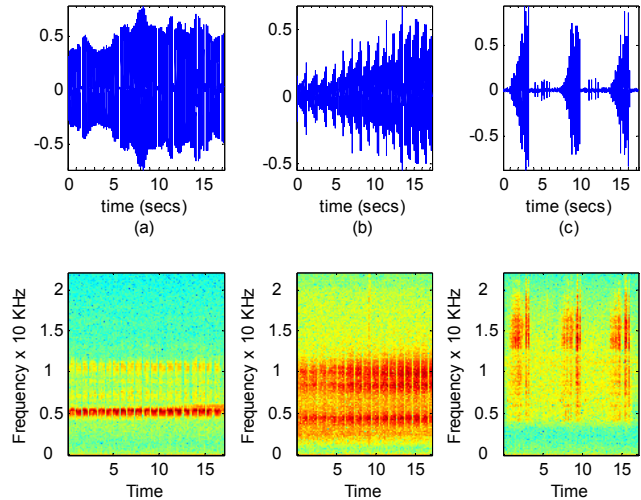

Figure 1. Typical sound patterns of insects. *Top row:* (*a*) cricket (Anaxipha n. sp. A), (*b*) cicada (Neocicada hieroglyphica), (*c*) katydid (Amblycorypha cajuni). *Bottom row:* corresponding spectrograms of the time-domain signals.

produced by males vibrating their tymbal mechanism. Females do not produce consistent acoustic emissions like males; instead, they produce the so-called 'female wing flick signals'. 'Wing flick' signals are produced through quick vibration of the wings. The vibrations produce short-time, broad-band acoustic signals in response to male calls. The timing of this signal in relation to the male call is species-specific. Males perceive both the visual and acoustic clues of the wing-flick.

2.3. Katydids

Katydids are classified in the phylum Arthropoda, class Insecta, order Orthoptera, family Tettigoniidae and are also called long-horned grasshoppers and bush crickets. Katydids are nocturnal singers. They utilize a stridulation mechanism and, in some cases, may produce tones by exciting a resonance in a tegmen (one of the front pair of wings). Each wing stroke produces a pulse of sound in a katydid's chirp. The male's calling sound is a regular repetition of multi-pulse chirps or phrases and the females chirp (in some species females can also stridulate) in response to the song of the males [8-9]. The sound pattern is comprised of a sequence of clicks with relatively short inter-click intervals where a click is a single transient-like acoustic event (see Fig. 1(c)). Some calling songs of males contain several components produced in varying temporal sequences.

Some of the most essential acoustic clues for differentiation among similar families, subfamilies, and species of insects are: a) dominant harmonic, b) rhythm and duration of pulsations, c) spread of spectral energy around the dominant harmonic, d) energy of the overtones.

3. INSECT RECOGNITION

In Fig. 2 we present a diagram illustrating the acoustic insect recognition process. This process consists of two main steps: acoustic signal parameterization and classification. While the parameterization aims at computing descriptors which account for the useful information in the signal, the classification stage compares the input feature vectors with predefined statistical models of the target classes. A final decision is made depending on the degree of proximity between the input and the models.
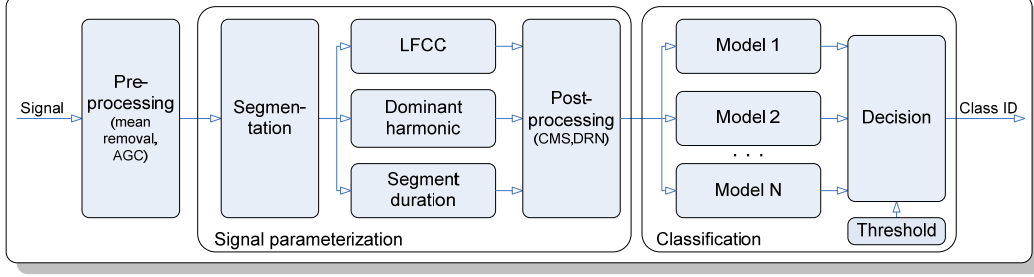
Figure 2. *Diagram of the acoustic insect recognition process for N distinct classes*

The signal parameterization approach is based on variable-size framing, which considers each *active* part of the signal (corresponding to bursts of pulsations) as an independent event. In the following, we describe the variable-size segmentation and signal parameterization steps.

Step 1: *Pre-processing of input signal:* It consists of mean value removal and amplitude normalization through automatic gain control applied to the time-domain signal.

Step 2: *Variable-length segmentation:* It is based on an energy detector, which estimates the energy E_{use} for a frame of K successive samples as:

$$E_{use}(k) = \sum_{i=1}^{K} \left(x(kL+i) \right)^2, \quad k = 0,...,M-1, \quad (1)$$

where x is the input signal, k is the group index, L is a predefined step size which defines the degree of overlapping between two successive frames, and

$$M = \lfloor (N - K + L)/L \rfloor \quad (2)$$

is the number of frames in a recording with length N samples. The operator $\lfloor \; \rfloor$ stands for rounding towards the smaller integer value. The $E_{use}(k)$ contour is further used as input for the detector of *acoustic activity.* Since the subsequent estimates of the energy are for overlapping groups, the precision of the detected borders depends on the step size L. In the present work, we consider L=5 (equivalent to time resolution ~113 *μsec* at 44100 Hz sampling frequency), which provides a good trade-off between time resolution and computational demands. For obtaining a smooth estimation of $E_{use}(k)$ we used a group size K=110 samples, which corresponds to frame size of 2.5 milliseconds.

Step 3: *Estimation of the signal descriptors:* Each active segment is subjected to short-time discrete Fourier transform (DFT). The sample size of the DFT equals the size of the segment. When the length of a segment is smaller than 2048 samples, we perform zero padding. In order to reduce the computational demands an upper bound of 1.5 seconds per segment was set. Furthermore, we apply a filter-bank consisting of B=200 equal-bandwidth and equal-height filters on the logarithmically compressed power spectrum, considering the frequency range 2÷22 kHz. We have chosen linear spacing (equal frequency resolution) because insects, in general, can produce sounds in frequencies anywhere in the acoustic spectrum (and some at ultrasound), in contrast to the human speech signal where most of the energy is concentrated in the low-frequency formant area. The lower bound of 2 kHz was imposed to eliminate the majority of interferences from the environment. The centres of the linearly spaced filters are displaced 100 Hz one from another, and serve as boundary points for the corresponding neighbouring filters. Subsequently, the log-energy filter-bank outputs X_i are subject to the discrete cosine transform (DCT):

$$LFCC_j = \sum_{i=1}^{B} X_i \cos\left(j(i-1/2)\frac{\pi}{B} \right), \quad j = 1,...,J, \quad (3)$$

where j is the index of the linear frequency cepstral coefficients (LFCC). A series of feature selection tests have indicated that the first 24 (J=23) cepstral coefficients provide a good trade-off for the recognition task. The 0-th cepstral coefficient was excluded from the feature vector as we did not want any dependence on the field recording setup.

Finally, for each segment the composite feature vector is designed by appending the: a) *dominant harmonic f_0* that is estimated via search of the maximum magnitude in the power spectrum, b) *segment duration l_{seg}* in seconds, and c) 23 LFCCs.

Step 4: *Post-processing of the features*: Cepstral mean subtraction (CMS) is applied on the LFCCs, and dynamic range normalization (DRN) is applied on the entire feature vector.

As presented in Fig. 2, the post-processed feature vectors are fed to the classification stage. For each target class an individual model was build. In the present work we consider: a) Probabilistic Neural Network (PNN)-based [10], b) Gaussian Mixture Models (GMM)-based [11], and c) Hidden Markov Model (HMM)-based [12] classifiers. GMMs and HMMs are trained based on a standard version of the expectation-maximization algorithm [11-12].

4. EXPERIMENTS AND RESULTS

In order to provide an efficient evaluation scheme of the proposed automatic recognition system, we used several corpora of insect recordings with known and reliable identification tags [4, 13].

According to our intention to evaluate two alternative schemes (straight and hierarchic) for insect recognition, we defined datasets that serve to a number of experiments. In the straight scheme, an unknown recording is compared to the model of each species. In the hierarchic scheme (refer to Fig.3) we are trying to identify the groups and subgroups to which the unknown recording belongs. The hierarchical scheme delivers some more information for the cases when the specific species cannot be identified.

In the experimentations we aimed at the identification of:
- 3 groups of singing insects {cicadas, crickets, katydids},
- 2 *populations* of crickets {North Mexico, Japan},
- 3 *populations* of cicadas {North Florida, Michigan, Thailand},
- 2 *families* crickets of North Mexico {Gryllidae, Gryllotalpidae}; 2 *families* of katydids {Tettigoniidae, Prophalangopsidae},
- 6 *subfamilies* of North Mexico crickets {Eneopterinae, Gryllinae, Mogoplistinae, Nemobiinae, Oecanthinae, and Trigonidiinae}; 5 *subfamilies* of katydids {Conocephalinae, Copiphorinae, Phaneropterinae, Pseudophyllinae, Tettigoniinae},
- 4 *genus* of crickets from subfamily Nemobiinae {Allonemobius, Eunemobius, Neonemobius, Pictonemobius}; 7 *genus* of katydids from subfamily of Phaneropterinae {Amblycorypha, Are-
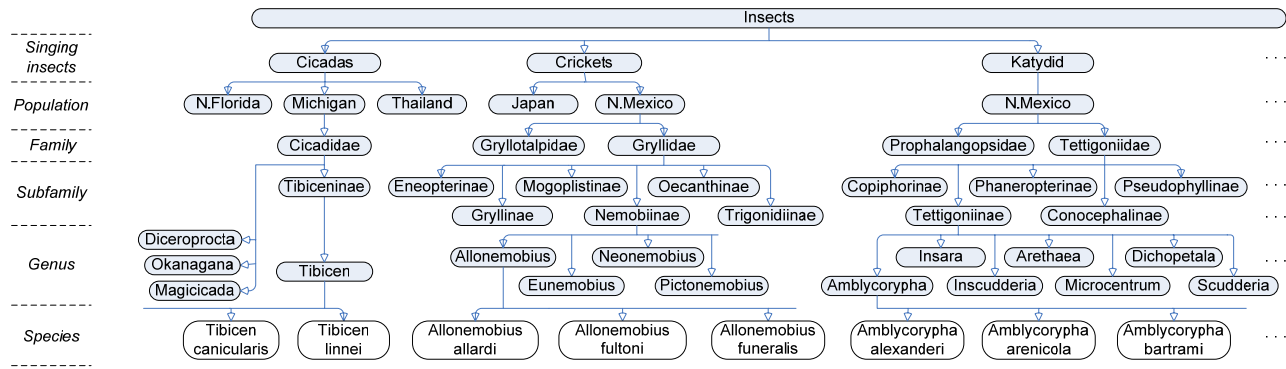
Figure 3. The hierarchical scheme for insect recognition

thaea, Dichopetala, Insara, Inscudderia, Microcentrum, Scudderia}; 4 *genus* of cicadas from family Cicadidae {Diceroprocta, Magicicada, Okanagana, Tibicen},

- 7 cricket *species* from genus Allonemobius; 14 katydid *species* from genus Amblycorypha; 7 cicada *species* from genus Tibicen,
- and a pool of 313 *species* (either cicadas, crickets, or katydids).

The train and test datasets were composed of non-overlapping parts of the aforementioned corpora [4, 13].

The experimental results are presented in Table 1. The sign "x" in some cells denotes these experiments that were not performed due to lack of data. In the straight scheme the PNN and GMM classifiers outperformed the HMM one. One reason is that when compared to PNNs and GMMs, HMMs require significantly larger amounts of training data in order to build reliable models. Although in the straight scheme the PNN and GMM classifiers made an equal number of errors they differed in the misclassified species. On the other hand, the HMMs, mainly due to their capability to model temporal sequences, outperformed the GMMs and PNNs at the most levels in the hierarchical scheme. At some levels of hierarchy (see level genus for cicadas and katydids, subfamily for crickets) the PNNs exhibited competitive performance when compared to GMMs and HMMs. However, when the performance is averaged for all experiments the GMMs surpass the PNNs.

Table 1. Experiments and results – accuracy in percentage

Straight (brute force) approach -- All species

	PNN	GMM	HMM	Comments
Species	86.3	86.3	75.2	313 species of Cicadas,Crickets,Katydids

Hierarchical approach -- 3 singing insects

	PNN	GMM	HMM	Comments
Sing.Insect	97.5	98.6	98.8	3 classes (Cicadas,Crickets,Katydids)

Hierarchical approach -- Cicadas

	PNN	GMM	HMM	Comments
Population	99.3	97.6	99.2	3 areas (N.Florida, Michigan, Thailand)
Family	x	x	x	1 family (Cicadidae)
Subfamily	x	x	x	1 subfamily (Tibiceninae)
Genus	97.9	94.4	97.9	4 genus (Magicicada,Okanagana,Tibicen,Dice
Species	85.7	100.0	83.3	7 species from genus Tibicen

Hierarchical approach -- Crickets

	PNN	GMM	HMM	Comments
Population	99.1	99.5	99.6	2 areas (Japan, N.Mexico)
Family	99.0	99.5	99.6	2 families (Gryllotalpidae, Gryllinae)
Subfamily	99.5	97.4	99.7	6 subfamilies from Family Gryllinae
Genus	90.2	93.8	93.8	4 genus from Subfamily Nemobiinae
Species	95.7	100.0	100.0	7 species from Genus Allonemobius

Hierarchical approach -- Katydids

	PNN	GMM	HMM	Comments
Population	x	x	x	1 area (N.Mexico)
Family	97.6	100.0	100.0	2 families (Tettigoniidae,Prophalangopsidae)
Subfamily	91.3	92.6	94.3	5 subfamilies from family Tettigoniidae
Genus	78.6	71.1	48.5	7 genus from subfamily Phaneropterinae
Species	71.4	85.7	50.0	14 species from genus Amblycorypha

5. CONCLUSION

In this work we address the task of the automatic acoustical identification of insects by elaborating signal parameterization methods and state-of-the-art pattern matching techniques in a manner that resembles the methodology of speaker recognition. The presented automatic identification system demonstrated to be highly accurate in recognizing the family, subfamily and specific species of insects. We postulate that this study will benefit potential non-intrusive acoustic environmental monitoring applications, as the proposed approach is directly expandable to other species.

6. REFERENCES

[1] Drosopoulos S., Claridge M., "Insect Sounds and Communication: Physiology, Behaviour, Ecology, and Evolution", Contemporary Topics in Entomology, CRC Press, 2005.

[2] Gaston K., O'Neill M.A., "Automated species identification -- why not?", Phil. Trans. Roy. Soc. B, 359(1444), pp. 655–667, 2004.

[3] Riede K., "Acoustic monitoring of Orthoptera and its potential for conservation", Journal of Insect Conserv., 2, pp. 217–223, 1998.

[4] Singing insects of North America collection (SINA) http://buzz.ifas.ufl.edu/

[5] Bennett-Clark H., "Resonators in insect sound production: How insects produce loud pure-tone songs", Journal of experimental Biology, 202, pp. 3347-3357, 1999.

[6] Bennett-Clark H., "Songs and the physics of sound production", in Cricket Behavior and Neurobiology. Ed. F. Hubrer, T.E. Moore and W. Loher, pp. 227-261, Cornell University Press, 1989.

[7] Alexander R., "Sound production and associated behavior in insects", The Ohio Journal of Science 57(2): 101, 102, pp. 101-113, 1957.

[8] Alexander R.,"Arthropods", in T.A. Sebeok, ed., Animal communication. Indiana University Press, Bloomington, pp. 167-216, 1968.

[9] Shaw K.C., "Environmentally-induced modification of the chirp length of males of the true katydid, Pterophylla camellifolia (F.) (Orthoptera: Tettigoniidae)", Annals of the Entomological Society of America, 68 (2), pp. 245-250, 1975.

[10] Specht D.F., "Probabilistic Neural Networks", Neural Networks, 3(1), pp. 109-118, 1990.

[11] Reynolds D., Rose R., "Robust text-independent speaker identification using Gaussian mixture speaker models", IEEE Trans. on Speech and Audio Proc., 3 (1), pp. 72-83, 1995.

[12] Rabiner L., "A tutorial on hidden Markov models and selected applications in speech recognition", in Readings in Speech Recognition, ed. by Waibel A., and Lee K., (Kaufmann, San Mateo), pp.267-296, 1990.

[13] Insect Sound World. Songs of crickets and katydids from Japan. http://mushinone.cool.ne.jp/English/ENGindex.htm