Automatic Acoustic Identification of Insects Inspired by the Speaker Recognition Paradigm

Ilyas Potamitis[†], *Todor Ganchev*[‡], *Nikos Fakotakis*[‡]

[†] Department of Music Technology and Acoustics,
Technological Educational Institute of Crete, Daskalaki-Perivolia, 74100, Rethymno, Crete, Greece
potamitis@stef.teicrete.gr

[‡] Department of Electrical and Computer Engineering,
University of Patras, 26500 Rion-Patras, Greece
{tganchev, fakotaki}@wcl.ee.upatras.gr

Abstract

This work reports our research efforts towards developing efficient equipment for the automatic acoustic recognition of insects. In particular, we discuss the characteristics of the acoustic patterns of a target insect family, namely the cricket family. To address the recognition problem we apply a feature extraction methodology that has been inspired by well documented tactics of speech processing, which were adapted here to the specifics of the sound production mechanism of insects, in combination with state-of-the-art speaker identification technology. We apply this approach to a large and well documented database of families and subfamilies of cricket sounds, and we report results that exceed 99% recognition accuracy on the levels of family and subfamily, and 94% on the level of a specific insect out of 105 species. We deem this equipment will be of practical benefit for non-intrusive acoustic environmental monitoring applications as it is directly expandable to other insect species.

Index Terms: bioacoustics, insects, identification

1. Introduction

There are more than 900,000 known insect species and there may be as many as ten times that number yet to be identified, forming the largest ecological group on Earth. Beyond the scientific interest of investigating the diversity of biological organisms, insects have great economic importance as beneficial organisms in agriculture and forestry (insects play significant role in the food chain of other species and the fertility of plants). However, a number of insect species also have negative contribution to agricultural economy as they constitute a threat to plants and crops.

Insects are mainly identified by their appearance and sound production that are species-specific. The detection and species recognition of insects are usually carried out manually, using trapping and observation methods. The 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). Moreover, 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. Non-experts have great difficulty practicing taxonomy while participating in the construction of biological inventories, even for routine identifications.

Recent progress in computer technology as well as in signal processing and pattern recognition has introduced the possibility of automatically identifying species primarily on the basis of capturing subtle differences by means of image [1-3] and acoustic signal processing [4-6]. However, the application of pattern recognition and machine learning techniques to this kind of problem is still in its infancy.

In brief, acoustic identification of insects is based on their ability to generate sound either deliberately as a means of communication or as a by-product of eating, flight or locomotion. Provided that the bioacoustic signal produced by insects follows a consistent acoustical pattern that is species-specific, it can be employed for detection and identification purposes.

In the present contribution, we address a bioacoustic signal classification problem by exploiting technology that was initially developed for speech recognition, and afterwards successfully adapted for the needs of speaker recognition. In particular, by adopting the statistical learning techniques inherent to speaker identification [7], we aim at categorizing acoustic recordings to specific families and subfamilies of insects, as well as at identifying the definite species. However, the applicability of these techniques and especially the feature extraction procedure is not straightforward, as the spectral patterns of insect sounds differ to a great extent from those of speech, mainly due to the different sound production process (insects do not possess vocal tract and do not use their mouth to produce vocalizations).

The selected target groups of insects to be identified are families, sub-families and specific species of crickets mainly from the singing insects of North America collection (SINA) [8]. This group has been chosen because the SINA project presents a large and representative collection of cricket sounds of North America that are identified and tagged by scientists of considerable experience in identifying the taxonomy of insect species.

The long term objectives of this work are:

- The development of a pilot automatic detection/ identification equipment capable of detecting/identifying insect species. Progressively, this equipment can be extended to other living organisms that are able to produce consistent acoustic patterns. This will allow unmanned, non-invasive acoustic surveying and environmental monitoring, which will cost-effectively assess and categorize the biological diversity of large regions.
- The recognition of a wide range of taxa by non-specialists.
- Automatic acoustic identification of pests and *selective* activation of repelling mechanisms based on ultrasound emission, which is tuned to specific insects.

2. Sound production in insects

There is a specific number of behavioural modes that have been observed in connection with sound production in insects. The first mode includes those situations in which the insects produce sounds to attract the female mates into close proximity (e.g. crickets produce the so called 'courtship' or 'mating' songs) or to cause the female to produce a sound that will help the male to locate her (slant-faced grasshoppers) or cause congregation of large numbers of males and females (cicadas). The second general behavioural mode consists of the situations in which (following [9]): a) the sound is produced as a reaction to the presence or activities of other organisms. In particular, males, females, or immature insects produce acoustic emissions in order to declare their disturbance (when captured and held or because of the presence of another organism), or warn other insects of danger, b) a male insect may sing in order to let other males know that an area is his territory, (i.e. mark their territory generally called 'warning', 'intimidation' or 'fight' sounds), c) a female produces acoustic emissions in the presence of a male of the same species.

Besides the sound generation as a means of communication, sound can be produced non-intentionally as a result of eating, flying or locomotion. 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 [10], [11].

Sounds are produced by insects in five different ways [9]:

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, such as the 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).

In Fig. 1 we depict characteristic spectrograms of some of these sound production mechanisms.

2.1 The acoustic profile of cricket sounds

Crickets (the male ones) 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. 2). Some the most essential acoustic clues for differentiation among similar families, subfamilies, and specific species are the:

a) dominant harmonic,
b) rhythm and duration of pulsations,
c) spread of spectral energy around the dominant harmonic,
d) energy of the overtones.

The most challenging task is to classify cricket species belonging to the same family and subfamily as their tonal characteristics bear resemblance to each other. In Fig. 2 we depict spectrograms of members of the same subfamily (top right and top left figures, the subfamily Trigonidiinae) and different subfamilies (bottom left and right figures, subfamily Oecanthinae and subfamily Nemobiinae). Another difficulty is that some species admit a very sparse representation of their signal in the time-frequency domain (e.g. subfamily Eneopterinae). 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.

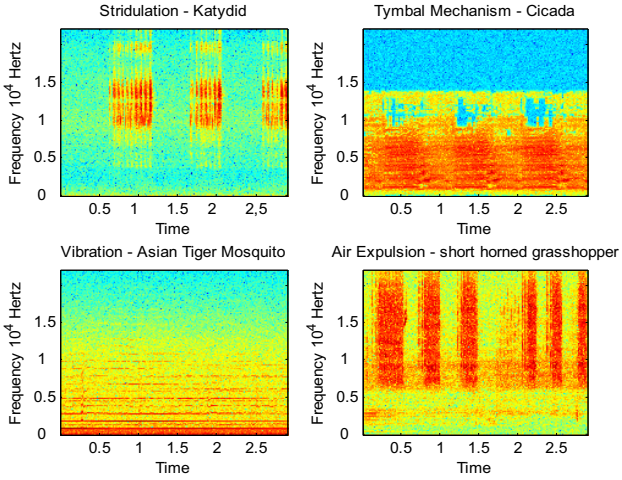

Figure 1 *Spectrograms of members of different insect families having different sound production mechanisms*
Top left: Stridulation: Katydid
Top right: Tymbal: Cicada
Bottom left: Vibration: Asian tiger mosquito
Bottom right: Air Expulsion: Short-horned grasshopper

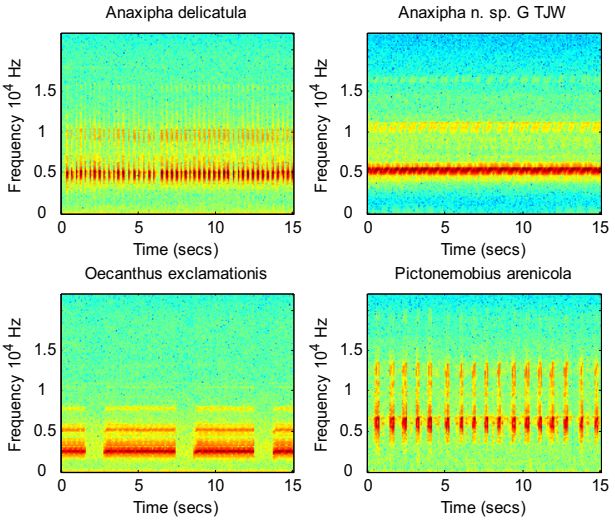

Figure 2 *Spectrograms of cricket sounds*
Top left: Subfamily Trigonidiinae, Anaxipha delicatula member
Top right: Subfamily Trigonidiinae, Anaxipha n. sp. member
Bottom left: Subfamily Oecanthinae, Oecanthus exclamationis
Bottom right: Subfamily Nemobiinae**,** Pictonemobius arenicola

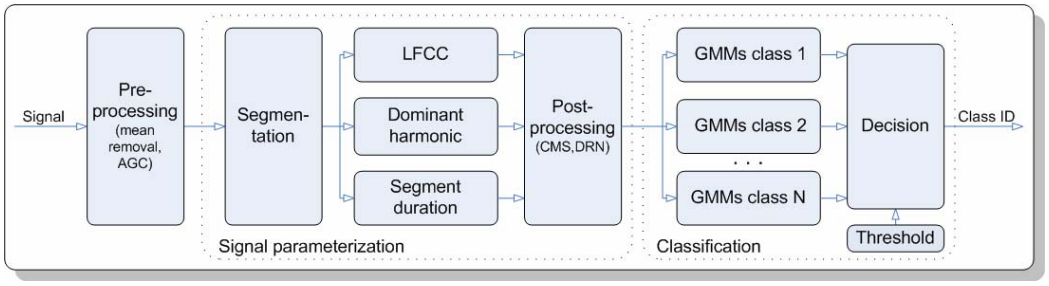

Figure 3 *Diagram of the acoustic insect recognition process for N distinct classes*

3. Insect recognition

In Fig. 3 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 models of the target classes. A final decision is made depending of the degree of proximity between the input and the models.

In this section, we discuss two signal parameterization approaches. The first one is similar to the speech parameterization concept and utilizes fixed-size frame segmentation with a fixed degree of overlap among the subsequent segments. The second one is based on variable-size framing, which considers each *active* part of the signal (corresponding to bursts of pulsations), as an independent event. Each event is treated as one integral chunk and is processed independently from all other events. In the following, we describe the steps of the variable-size frame parameterization followed by the fixed-size frame case.

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

Step 2: *Signal segmentation:* It is based on an ultra-short-time energy detector, which estimates the energy E_{use} for small non-overlapping groups of successive samples as:

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

where x is the input signal, k is the group index, L is a predefined group size, and $M=N/L$ is the number of frames in a recording with length N samples. 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 non-overlapping groups, the precision of the detected borders depends on the group size L. In the present work, we consider L=5 (equivalent to time resolution ~110 *μsec* at 44100 Hz sampling frequency), which provides a good trade-off between time resolution and computational demands. Fig. 4 provides an illustration for the case of variable-size frame segmentation.

Step 3: *Composition of the feature vector:* Each segment is subjected to short-time discrete Fourier transform (STDFT). The sample size of the DFT equals the size of the chunk. When the length of a segment is smaller than 2048 samples, we perform zero padding. This provides a frequency resolution of ~22 Hz. In order to reduce the computational demands an upper bound of 1.5 sec segments is set. The frequency range is restricted to 2÷10 kHz where most of the cricket activity takes place. Furthermore, we apply a filter-bank consisting of B=80 equal-bandwidth and equal-height filters on the logarithmically compressed power spectrum. The centres of the linearly spaced filters are displaced 100 Hz one from another, and serve as boundary points for the corresponding neighbouring filters. 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 speech signal where most of the energy is concentrated in the low-frequency formant area. The next step is the decorrelation of the log-energy filter-bank outputs X_i via the discrete cosine transform (DCT):

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

where j is the index of the linear frequency cepstral coefficients (LFCC). A series of feature selection tests have shown that the first 24 (J=23) cepstral coefficients are sufficient to carry out the recognition task and the contribution of the higher order LFCCs has a minor impact on the recognition performance. The 0-th cepstral coefficient was excluded from the feature vector as we did not want any dependence on the field recordings setup. Finally, for each segment the final feature vector is composed of a) the *dominant harmonic* f_0 that is estimated via search of the maximum magnitude in the power spectrum, b) *segment duration* l_{seg}, and c) the 23 LFCCs. The time lag between pulsations is not used as a feature mainly because crickets vary their rhythm of pulsations due to temperature variations.

Step 4: *Features post-processing*: Cepstral mean subtraction and dynamic range normalization.

In the case of fixed-size frame segmentation, we performed the previously described feature extraction steps using a frame of l_{seg}=100 msec with 90% overlapping among the successive frames. The feature extraction is applied on the active segments selected by the acoustic activity detector described in Step 2.

The normalized feature vectors obtained either for variable- or fixed-size frames are fed to the GMM-based classifier. As presented in Fig.3 for each of the target classes a GMM is considered. The class-specific diagonal covariance GMMs are trained by employing a standard version of the expectation maximization algorithm [12]. Different number of Gaussians, ranging from 8 to 256, is employed for the different classes to better fit the underlying distributions. During the classification process, we compute for each trial class-conditional probabilities from all models. The Bayesian decision rule is applied to detect the winning class [7].

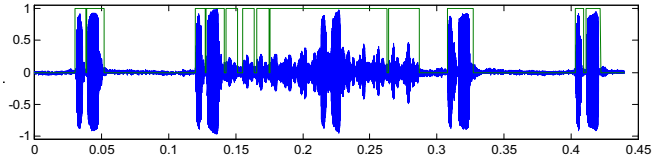

Figure 4 *An example for variable-size frame segmentation* (*the ordinate stands for magnitude and the abscissa for time in sec*)

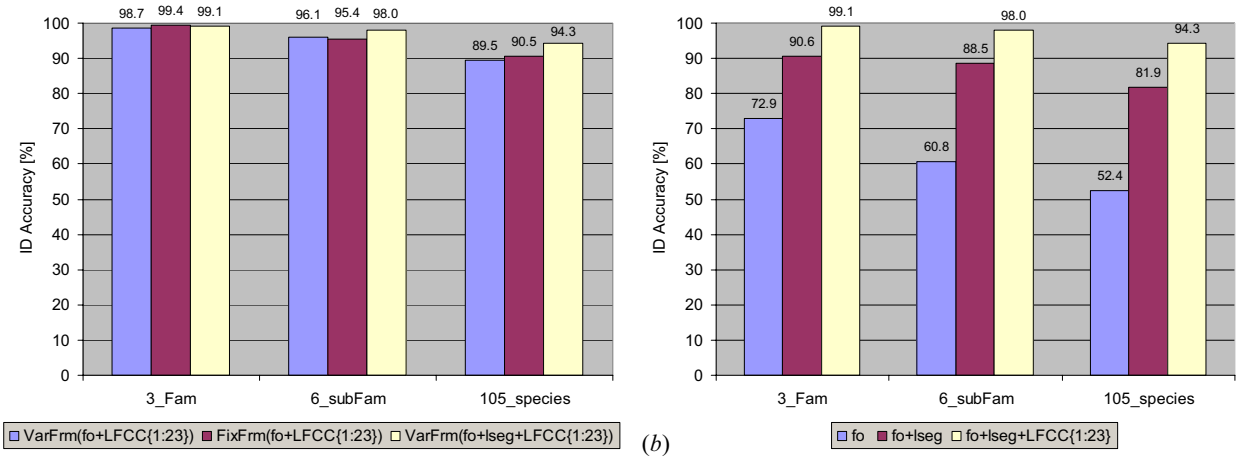

Figure 5 *Identification accuracy for the Family, Sub-Family, and Species experiments: (a) comparison between variable and fixed-size frame approaches, and (b) the contribution of the various parameters in the variable–size frame segmentation scenario*

4. Experiments and results

We were interested in identifying the entire hierarchy (*specific species*, *subfamily*, *family*, etc) to which specific recordings belong. Thus, experiments aiming at identification of:

- three different *families* {North Mexico Gryllidae (102 species), North Mexico Gryllotalpidae (4), and crickets from Japan (26)[1]},
- six *subfamilies* of North Mexico crickets {Eneopterinae (9), Gryllinae (22), Mogoplistinae (17), Nemobiinae (20), Oecanthinae (17), and Trigonidiinae (17)},
- and 105 *species* that belong to the six subfamilies of North Mexico crickets, were performed.

The experimental results are presented in Fig. 5. Specifically, Fig. 5(*a*) presents comparison between the signal parameterization approaches discussed in Section 3, and Fig. 5(*b*) presents the accuracy gain due to combining various parameters. The results depicted in Fig. 5(*a*) demonstrate the advantage of the variable-frame analysis that allows finer analysis of the spectrum of the quasi-stationary insect sounds compared to the fixed frame case. Figure 5(*b*) reports results on different features sets and how the recognition accuracy scales when we append them.

The recognition results prove that: (1) the significance of the dominant frequency as a main clue for insect recognition, and (2) segment duration and spectral energy distribution around the dominant carry useful supplementary information.

5. Conclusions

Despite two centuries of taxonomic activity the overwhelming majority of species in most ecosystems remains unidentified, while only a small number of highly trained specialists are able to make any identification [1]. In this work we address the task of the 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 is based only on the acoustic modality and demonstrated to be highly accurate in recognizing the family, subfamily and specific members of a target insect.

[1] http://mushinone.cool.ne.jp/English/ENGindex.htm

6. References

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

[2] Watson A.T, O'Neill M.A, and Kitching I.J, "A qualitative study investigating automated identification of living macrolepidoptera using the Digital Automated Identification SYstem (DAISY)", *Systematics and Biodiversity*, 1(1), 2003.

[3] Mortensen E. et al., "Pattern Recognition for Ecological Science and Environmental Monitoring: An Initial Report", N. MacLeod and M. O'Neill (Eds.) *Algorithmic Approaches to the Identification Problem: Systematics,* http://web.engr.oregonstate.edu/~tgd/ publications/index.html

[4] Chesmore E., Ohya E., "Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition", *Bulletin of Entomological Research*, 94(4), pp. 319-330, 2004.

[5] Chesmore E., "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals", *Applied Acoustics*, 62, pp. 1359-1374, 2001.

[6] Dietrich C., "Temporal Sensor fusion for the Classification of Bioacoustic Time Series", PhD thesis, University of Ulm, Department of Neural Information Processing, 2004.

[7] Reynolds D., and 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.

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

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

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

[11] 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. Ithaca, New York: Cornell University Press, 1989.

[12] McLachlan G.J., and Krishnan T., "The EM algorithm and extensions". Wiley Series in Probability and Statistics, New York: Wiley, 1997.