Recognition of Bumblebee Species by their Buzzing Sound

Mukhiddin Yusupov[1], Mitja Luštrek[2], Janez Grad, Matjaz Gams[2,1]
Czech Technical University in Prague – Computer Science Department
Czech Republic

[2]Jozef Stefan Institute – Intelligent Systems Department
Slovenia
{yusupmuk@fel.cvut.cz} {mitja.lustrek@ijs.si} {matjaz.gams@ijs.si} {janez.grad@siol.com}

ABSTRACT: *The goal of our work is to help people to automatically identify the species and worker/queen type of bumblebee based on their recorded buzzing. Many recent studies of insect and bird classification based on their sound have been published, but there is no thorough study that deals with the complex nature of buzzing sound characteristic of bumblebees. In this paper, a database of recorded buzzings of eleven species were preprocessed and segmented into a series of sound samples. Then we applied J48, MLP and SVM supervised classification algorithms on some predetermined sets of feature vectors. For five species the recognition rate was above 80% and for other six species it was above 60%. At the end we consider how to further improve the results.*

Keywords: Buzzing Sound, Supervised Classification

Received: 2 July 2019, Revised 24 September 2019, Accepted 8 October 2019

DOI: 10.6025/jes/2019/9/4/131-134

1. Introduction

Bumblebees are important pollinators of many plants and their colonies are now used extensively in greenhouse pollination of crops such as tomatoes and strawberries. Some bumblebee species are declining and it is a cause for concern in Europe, North America and Asia. In Europe, around one quarter of species are threatened with extinction, according to recent studies. This is due to a number of factors, including land clearing and agricultural practices. There is a lot of research devoted to keep some bumblebee species from such decline.

Until now over 250 species are known. There are about 19 different species of bumblebee found in the UK, 68 in Europe, 124 species in China, 24 in South America and 35 in Slovenia. The colonies of bumblebees are composed of a queen and many workers. Since only experts can identify the species by looking at or listening to them and their sound, we decided to make this identification easy for all. One needs to record the buzzing and provide it to the system (program) that will process it and then tell to which species and worker/queen type this buzz corresponds to. The program is accessible from the homepage of the Department of intelligent systems at the Jozef Stefan Institute - http://dis.ijs.si/. More information can be provided from janez.grad@siol.com.

Although there are generalizations of the type of problem we are solving here, such as a system for classifying many types of insects [4], relatively little has been done previously on automatic recognition on bumblebee species with a detailed analysis of their buzzing sound. Several internet applications provide sounds and images and images of different species of birds, frogs etc., but they rely on human pattern-recognition skills to identify the species at hand and do not provide active help. Our study is related to active system help in recognizing a particular (sub)species, and in particular to other audio data classification problems like classification of general audio content [8], auditory scene recognition, music genre classification and also to the speech recognition, which have been studied relatively extensively during last few years also in our department. We here try to take advantage from these previous studies.

Some studies like [7], where they also tried to classify bee species, used different approaches. We can view these systems as solving a pattern recognition problem. In [7] the recognition of bee species is performed visually, based on its image. The task was to find relevant patterns from the image and identify similarities to specific species. However, pictures vary a lot based on different factors, and often a picture does not represent well what we see in nature. In our work the patterns are buzzing sound events produced by bumblebees. The chosen approach is recognition based on a parametric (feature) representation of the sound events. Features should be selected so that they are able to maximally distinguish sounds that are produced by different bumblebee species. Most of the recognition systems based on audio and especially human voice recognition uses Melfrequency cepstrum coefficients (MFCC) as a feature vector. There are also works where feature vectors are Linear Predictive Coding coefficients (LPC) or a set of low-level signal parameters like in [1].

This paper uses MFCC and LPC to extract the features. For the extraction of features and for other processing of audio records we used jAudio package [9]. Before feature extraction we preprocess and segment the sound recordings. We found that the segmentation is as important as the extraction of features with a strong influence on the prediction accuracy. Then we constructed the model separately with three different classification algorithms: J48, MLP and SVM. Training and evaluation of a model were performed on a stored database of fifteen species of bumblebees. The experiments were carried out using WEKA open source machine learning software. Results show that SVM has better performance than other two systems.

2. Preprocessing

Each sound record preprocessing consists of three steps: normalization, pre-emphasis and segmentation. First the normalization is applied to the record by dividing it with maximum value:

$$\tilde{a} = x(i)/max\, x(i),\ 0 \square i \square n - 1 \tag{1}$$

where $x(i)$ is the original signal, $\tilde{x}(i)$ is the normalized signal and n is the length of the signal.

After that pre-emphasis is performed in order to boost only the high-frequency components, while leaving the lowfrequency components in their original state. This approach is based on observations that sound data comes with a high frequency and low magnitude whereas the parts of the recording that we are not interested in (noise, gaps) incorporate low frequency and much higher energy. The pre-emphasis factor α is computed as

$$\alpha = e^{-2\pi Ft} \tag{2}$$

where t is the sampling period of the sound. The new sound is then computed as:

$$H(z) = 1 - \alpha z^{-1} \tag{3}$$

The last step of the pre-processing is segmentation. In this step we separate sound record into a number of samples which represent only the buzzing. Each sound record is 45 to 60 seconds long. Extracting features from the whole sound record, firstly, increases the computational complexity and, secondly, affects the accuracy of the recognition. We do not need to calculate the features for the silent, noisy and other irrelevant parts of the record.

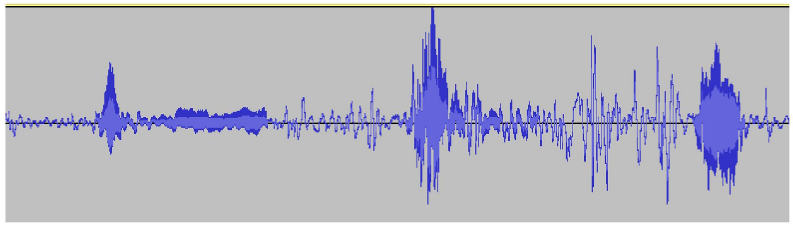

Figure 1. Representation of audio record of humilis queen species in time domain

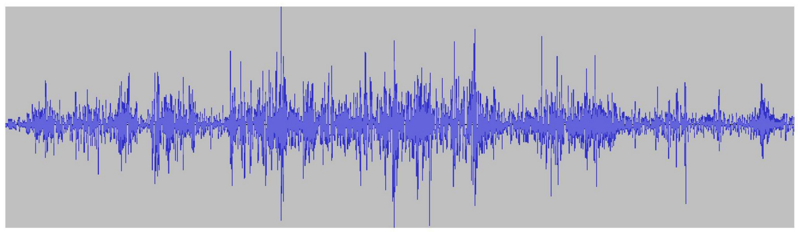

Figure 2. Representation of audio record of sylvarum worker species in time domain

However, spectral changes of signal parts are rather diverse and detection of boundaries of such samples is difficult because adjacent samples of separate buzzings can overlap in time and frequency. Moreover, due to the starting point of buzz is being slow it may occur below the background noise level. In Figure 1 we can see the representation of the sound record of humilis queen. It is difficult to recognize there are three separate relevant parts and everything in between with low frequency components as not relevant.

In Figure 2 it is even more problematic to say where exactly buzzing of the sylvarum worker starts and only in 20% of the recording there is the buzzing signal we are interested in. During the investigation of spectrum of each bumblebee species we also found out that buzzing of the same species can vary based on the state of the bumblebee or during the buzzing of one species some other ones can join and as a result we will have a combination of buzzes. Same species makes one buzz when for example it is working and has some other different buzz when it is angry.

We have to take into account various factors in devising a segmentation method, since unsuccessful separation of a record would result in unsuitable candidate samples and subsequently parametric representation would be different than for real signal data. That is why for the current version of our work we decided to segment the audio recordings manually by an audio editor program so that we could see the result of recognition based purely on real signal data. On one hand, this decision of manual separation obliges us to use in a testing phase of the model only noisiness records where most parts of the record consists of signal data. But on the other hand we analyzed how recognition accuracy changes when we change the strategy for segmentation, since by visually looking at the spectrum of record it is easier to segment it. In this current state of the work we segmented the recording manually into samples of 1-4 seconds of length and the parts which have less than 1 seconds of buzzing duration we combined with adjacent samples.

3. Feature Extraction and Model Construction

For each sample segment we calculated MFCC and LPC features. These are features that are mostly used in audio based classification tasks. Samples are processed in a window-by-window manner. The size and the overlapping factor of windows are the two key parameters in feature extraction in general in any audio/signal processing task. We found that the window size of 2014 and the overlapping factor of 30% gives us the feature vectors, which subsequently resulted in the best recognition model. For each window we have several MFCC or LPC values. It is better to represent each window with one feature value by aggregating all the values in a window, so we applied the aggregation by computing the mean value for each window.

In this work we considered three classification algorithms for building the model and these are the J48 Decision Tree, MLP Neural Network and SVM algorithms. Models were built by these algorithms to classify among the 15 different bumblebee cases, most common in Slovenia, i.e. central Europe. Classifying between a worker and a queen is not difficult and on average 90% for all species are easily identified, but we are more interested in knowing the exact type of species like hortoum, hypnorum and pratorum in addition to the status of a bumblebee in colony. So for each species we have two cases, either queen or worker, altogether 15 classes. The fact that the number of records for each class of species in our testing and training dataset are not evenly distributed caused slight inconvenience for us to build a good model. Also, this is one of the reasons why we have different rates of accuracy for all species. 5 of the bumble species we recognized with above 80% of accuracy and 2 of them had a rate of 95%. In Table 1 we provide the rates of recognition for each model built separately on MFCC and LPC feature values with the three algorithms.

	LPC	**MFCC**
J48	56%	56%
MLP	56%	60%
SVM	57%	64%

Table 1. Evaluation of the rates of recognition accuracy for each built model

In practical terms, when the system proposed three mostprobable classes, the accuracy rose to over 90% overall, enabling users to distinguish between the three proposed potential solutions visually. This is the way the system works at the moment.

4. Conclusion

In future we want to make the segmentation step to separate record of samples automatically in a system by incorporating all we learned from recordings and patterns of the 11 bumblebee species of both types. Also, we are going to build model using HMM and deep learning, because in many works related to audio classification HMM and deep learning produce best results. Then we intend to compare its result with the ones we obtained from SVM.

References

[1] Fagerlund, Seppo. (2007). Bird Species Recognition Using Support Vector Machines. *EURASIP Journal on Advances in Signal Processing*, vol 2007, Article ID 38637. doi:10.1155/2007/38637

[2] Leqing, Zhu., Zhen, Zhang. (2010). Insect sound recognition based on SBC and HMM. 2010 International Conference on *Intelligent Computation Technology and Automation*.

[3] Briggs, Forrest., Raich, Raviv., Xiaoli, Z., Fern. Audio Classification of Bird Species: a Statistical Manifold Approach.

[4] Diego, F., Silva, Vinícius, M. A., de Souza, Eamonn Keogh, Daniel, P. W., Ellis. Applying Machine Learning and Audio Analysis Techniques to Insect Recognition in Intelligent Traps.

[5] Ruben Gonzalez. Better than MFCC Audio Classification Features.

[6] Bueno, J. F., Francoy, T. M., Imperatriz-Fonseca, V. L., Saraiva, A. M. Modeling an automated system to identify and classify stingless bees using the wing morphometry – A pattern recognition approach.

[7] Fujisaki, H., Ohno, S. (1996). Analysis and Modeling of Fundemental Frequency Contour of English Utterances. Proc. EUROSPEECH'95. 2, 634-637. Philadelphia.

[8] Erling Wold, Thom Blum., Douglas Keislar, James Wheaton. Content Based Classification, Search, and Retrieval of Audio.

[9] McEnnis, Daniel., McKay, Cory., Fujinaga, Ichiro, Depalle, Philippe. jAudio: A feature extraction library.