SIGNAL PROCESSING FOR VOCAL RECOGNITION OF STURNUS VULGARIS

Jana Lázničková
Bachelor Degree Programme (1), FEKT VUT
E-mail: xlazni09@stud.feec.vutbr.cz

Supervised by: Petr Marcoň
E-mail: marcon@feec.vutbr.cz

Abstract: This paper describes the issue of sturnus vulgaris detection in the vineyards in order to scare these animals more effectively. The analysis and classification of bird singing is difficult because many problems can appear. One of the problems is background noise e.g. sounds of cars, trees, and also the singing of various bird species at once. Another problem is different types of bird songs. For example, an alarm melody, search for food, and also communication between the birds during a flight. This article presents a solution to one of these problems in case when only audio recordings are available.

Keywords: sturnus vulgaris, raspberry, MFCC, python, spectrogram

1 INTRODUCTION

Starlings in the vineyards cause significant damage every year. Winegrowers use several scaring devices to reduce the invasions of starlings. However, devices on the market often disturb surroundings with excessive noise. There is an idea to scare starlings before they land in the vineyards, therefore, their presence would have to be detected by vision (camera) or noise (microphone).

This paper is focused on the creation of a Mel-spectrogram, which facilitates the subsequent use of CNN (convolutional neural network) to detect the singing of starlings. Many papers using robust methods of deep learning to identify the sounds of birds have been published. To increase the sensitivity of the detection, the recorded audio must be pre-processed.

Humans perceive only a narrow range of frequencies and have a poor perception of the linear scale. Moreover, people perceive differences at lower frequencies more effectively than at higher frequencies. To better visualize differences at higher frequencies, the conversion into the Mel-scale is made according to the literature [1].

\[ f_{mel} = Mel(f) = 2595 \log_{10}(1 + \frac{f}{700}) \]  

2 ALGORITHM FOR DATA PREPARATION FOR CNN

The algorithm using freely available libraries of the python programming language is used for sound detection. The algorithm has been modified in order to filter out the interfering components to facilitate the detection of the starling sound. This algorithm is implemented in the Raspberry Pi4 (RPi4) microcomputer. This microcomputer was chosen because of the required computing power and also because it is easy to mount, to transfer, and it supports real-time data transfer well.

The recorded birdsong is sampled and then transferred from the time domain to the frequency domain using the Fourier transform (FT). The result of the Fourier transform is the spectrum of the signal (spectrum = expression of amplitudes and phases of individual harmonic components in the frequency domain). A Short-time Fourier transform (STFT) is applied to this spectrum and the result
of this transformation is a spectrogram. The Matplotlib library is used to plot the spectrogram of the sampled signal. STFT is provided by Librosa library.

The spectrogram is depicted for 10 seconds, so as much useful information as possible from the sound recordings is retained. After that, a Mel-spectrogram is created from this spectrogram. The Mel-spectrogram is a spectrogram where the y-frequency axis is converted to a Mel-scale. A high-pass filter was used to remove low-frequency noise [3] because birds sing at high frequencies. The required frequencies from 1400 Hz is transmitted.

Mel-Frequency Cepstral Coefficients (MFCC) [1] are coefficients that make up the Mel-Frequency Cepstrum MFC. These coefficients are acquired by applying discrete cosine transform (DCT) to the Mel-spectrogram. The plotting of the Mel-spectrogram and the calculation of MFCC coefficients allow creating the CNN more effectively. The transfer of the data from a Mel-spectrogram to CNN requires all of them to be at the same size. The input signal can be received by RPi4 in real-time, and also with a directional microphone. Similar solutions are known, e.g. for the detection of other species of birds [4] or bats [5].

The entire process of audio recording processing can be described by one block diagram:

![Block diagram of audio recording](image)

**Figure 1:** A block diagram of audio recording

### 3 ASSESSMENT OF MEASURED SIGNALS

Figures 2 to 4 show the results of the analysis of the starling song. Figure 2 shows only the monophony of a starling. The linear frequency spectrogram of its sound is visualized in Figure 3. Figure 4 shows the Mel spectrogram, which is based on a linear frequency spectrogram. This spectrogram is converted to a Mel-scale, and then a high-pass filter is applied to it. Figure 5 shows the spectrogram of a golden eagle for comparison with a starling. The difference between spectrograms 4 and 5 is obvious.
Figure 2:  The audio recording of sturnus vulgaris.

Figure 3:  The spectrogram of sturnus vulgaris (linear scale).

Figure 4:  The Mel-spektrogram of sturnus vulgaris.

Figure 5:  The Mel-spektrogram of golden eagle.
4 CONCLUSION

The paper describes processing of birds sound recordings to increase the sensitivity of detection algorithms. A filter filtering out the low-frequency noise was introduced in the process. The conversion of a linear spectrogram to a Mel spectrogram and the subsequent calculation of MFCC coefficients is necessary to smoothly follow the CNN [6] to detect the sound of starlings in the vineyards. More work on the project will follow. In terms of software, CNN will be used and in terms of hardware, RPi4 will be attached to the rotating mechanism, which will be controlled by a microcomputer and according to the directional microphones will identify not only the sound but also the position of the source of the sound.

ACKNOWLEDGEMENT

This paper was created with the support of the Zéta 4 TAČR program, specifically the project TJ04000441: Starling scare system based on a passive optical locator.

REFERENCES


