Bird sound recognition and classification using wavelets

Arja Selin, Jari Turunen, Juha T. Tanttu

Tampere University of Technology, Pori, Pohjoisranta 11, P.O.Box 300, FIN-28100 Pori, Finland – E-mail: arja.selin@tut.fi

Birds and especially sounds produced by birds have always fascinated people. Many biologists have found bird song to be an interesting area for example for behaviour and recognition research. Birds produce sounds generally using the sound channel. Most sounds are produced by the syrinx, which is the avian vocal organ. The syrinx is bipartite, so the bird can produce two notes simultaneously. Due to specific properties of bird sound, it is necessary to develop new signal processing techniques. Bird sounds can be tonal or inharmonic. The main idea was to study, how inharmonic and transient bird sounds can be recognized and classified efficiently. The wavelet analysis was selected due to its ability to preserve both frequency and temporal information, and to analyse signals, which contains discontinuities and sharp spikes, as inharmonic bird sounds. The essential information of the sounds is possible to present with only few features obtained from the wavelet analysis.

First the recorded soundtracks were segmented into smaller pieces, and then the sounds were preprocessed, which included for example noise reduction. The sampling rate F_s was 44.1 kHz. All sounds were decomposed into the wavelet coefficients with the wavelet packet decomposition (WPD) using the Daubechies db10 wavelet function, because it compromised the best decomposition results with selected bird sounds. The best decomposition level (N) in this case was six, and thus, the signal was split into $2^6=64$ bins. The bin 1 and the bins 33–64 proved to be irrelevant due to backround noise. Thus, the wavelet coefficients were calculated from bins 2–32, and the frequencies between $F_s/128$ to $F_s/4$ were used in the recognition. The number of the coefficients in the each bin is $L/2^N$, where L is the length of the original signal.

The four features, maximum energy, position, length, and width were calculated from these coefficients. The average energy of the wavelet coefficients of each 31 bins was calculated, and then the maximum energy value was chosen of the decomposition bin coefficient energies. The position represented the number of the bin, in which the maximum energy was located. The number of those coefficients of each bin, which exceeded the chosen threshold value, was added together and then it was averaged between the all bins that contained sound energy. This feature was called the length. The width represents the number of those bins, which sum of the energy exceeded the other chosen threshold value. Finally, all four features were normalized, in order that they were comparable with each other.

The data consisted of sounds of eight bird species. The Mallard (*Anas platyrhynchos*), the Graylag Goose (*Anser anser*), the Corncrake (*Crex crex*), the River Warbler (*Locustella fluviatilis*), the Magpie (*Pica pica*) are species with inharmonic sounds. The Pygmy Owl (*Glaucidium passerinum*), the Quail (*Coturnix coturnix*) and the Spotted Crake (*Porzana porzana*) were chosen as a reference species. All training and testing data (3132 sounds) of these eight bird species was from different tracks and that data was divided so that there were about 70% training data and 30% testing data. The two well-known neural networks, the multilayer perceptron (MLP) and the self-organizing map (SOM) were used as classifiers. The networks were first trained, when the feature vectors of four features were fed to the supervised 15-40-3 MLP and to the unsupervised 10 x 10 SOM. After that both networks were tested.

The results were quite encouraging, for the MLP classified 96% and the SOM 93,8% of the test sounds correctly. Using the MLP all test sounds of the Quail and the Spotted Crake were recognized correctly and 24 sounds of all test sounds were classified as outliers. Correspondingly, using the SOM all test sounds of the Pygmy Owl and the River Warbler were recognized correctly, and 32 sounds of all test sounds were classified as outliers. In conclusion, the wavelet transform has proved to be an efficient method taking into consideration. The tools presented in this study provide quite robust approach for recognition and classification purposes, particularly among the inharmonic and transient bird sounds.