De-noising aspects in the context of feature extraction in automatic bird call recognition

Jari Turunen, Arja Selin, Juha T. Tanttu, Tarmo Lipping

Tampere University of Technology, Pori, Pohjoisranta 11, P.O.Box 300, Fin-28101, Pori, Finland – E-mail: juha.tanttu@tut.fi

From technological point of view, noise exists in all sampled and stored signals. In the case of studio quality recordings, thermal noise is generated in electrical components. However, usually the noise level is too low to affect the end-use of the recordings. In the case of recordings made in nature, like bird call recordings, for example, different types of noise are present. These include wind noise (low-frequency and broadband), noise generated by trees, sounds made by other birds and animals etc.

During recording sessions, some of the noise effects can be reduced using proper screening and recording equipment. Most noise types are part of the sound environment and belong to the listening experience, however, in automatic bird call recognition noise deteriorates feature extraction. There are common methods to overcome this problem including de-noising the sound samples to reduce the effects of noise or using feature extraction methods that themselves reduce or are insensitive to the effects of noise, (wavelet decomposition or harmonic component trajectory tracking, for example). Some birds, like the Raven (*Corvus corax*) have very rich vocabulary of calls with most of the calls containing inharmonic and transient sounds. The background noise can conceal some of the low amplitude features of these calls.

In this paper we demonstrate the effects of noise reduction in automatic bird call recognition of 8 bird species (the Mallard (*Anas platyrhynchos*), the Greylag Goose (*Anser anser*), the Quail (*Coturnix coturnix*), the Corncrake (*Crex crex*), the Pygmy Owl (*Glaucidium passerinum*), the River Warbler (*Locustella fluviatilis*), the Magpie (*Pica pica*), the Spotted Crake (*Porzana porzana*)). Wavelet packet decomposition using Daubechies 10 wavelet function and 6 level decomposition was used for feature extraction. Four features including the duration and bandwidth of the high-energy area of the time-scale plane as well as the maximum energy and the scale of the maximum energy were calculated. These features were fed in to the Multilayer Perceptron (MLP) network.

Three pre-processing scenarios were compared: filter bank based mild noise reduction, Finite Impulse Response (FIR)-lowpass filter (with 10 kHz cutoff frequency) and no filtering. The training data consisted of 2278 samples and test data from 854 samples. From the test data 95.1%, 96% and 100% of the samples were recognized correctly using no filtering, the wavelet decomposition and the FIR filter, respectively. It can be seen from the results that the wavelet decomposition used for feature extraction is itself a robust decomposition method against the noise, however, the pre-reduction of noise level affects positively the recognition results. FIR filtering produced the best recognition results at the expense of totally destroyed sound environment. The filter bank method improved the recognition results at the same time preserving most of the sound environment.

Poster