RAZPRAVE IV. RAZREDA SAZU	XLVII-3	205–212	LJUBLJANA 2006

DE-NOISING ASPECTS IN THE CONTEXT OF FEATURE EXTRACTION IN AUTOMATED BIRD SOUND RECOGNITION

VPLIV ZNIŽEVANJA ŠUMA NA EKSTRAKCIJO ZNAČILNIH PARAMETROV ZVOKA PRI AVTOMATIČNEM PREPOZNAVANJU PTIČJIH GLASOV

JARI TURUNEN, ARJA SELIN, JUHA T. TANTTU & TARMO LIPPING

ABSTRACT

De-noising aspects in the context of feature extraction in automated bird sound recognition

In this paper we demonstrate the effects of noise reduction in wavelet based bird sound recognition. The non-harmonic bird sounds are difficult to separate from the background noise in order to maintain necessary features for identification. We tested three options: (i) no filtering, (ii) adaptive filter bank and (iii) lowpass Finite Impulse Response (FIR) filter. The following identification results were obtained (i) 95.1%,(ii) 96% and (iii) 100%.

Key words: Bird sound recognition, adaptive filtering, de-noising.

IZVLEČEK

Vpliv zniževanja šuma na ekstrakcijo značilnih parametrov zvoka pri avtomatičnem prepoznavanju ptičjih glasov

V članku je prikazan učinek zniževanja šuma na avtomatično prepoznavanje ptičjih glasov ob uporabi "wavelet" analize. Ptičje glasove brez harmonične strukture je težje ločiti od hrupa oz. šuma ozadja in ohraniti značilnosti zvoka za prepoznavanje. Preizkusili smo tri možnosti, brez filtriranja, z uporabo adaptivne zbirke filtrov in z nizkopasovno prepustnim FIR filtrom. Prepoznavanje napevov je bilo uspešno v naslednjih odstotkih: v prvem primeru 95,1 %, v drugem 96 %, in v tretjem 100 %.

Ključne besede: prepoznavanje ptičjih glasov, adaptivno filtriranje, zniževanje šuma.

Addresses – Naslovi

Jari Turunen
Tampere University of Technology
Pori, Pohjoisranta 11
P.O.Box 300
FIN-28101 Pori
Finland
E-mail: jari.j.turunen@tut.fi

Arja Selin
Tampere University of Technology
Pori, Pohjoisranta 11
P.O.Box 300
FIN-28101 Pori
Finland
E-mail: arja.selin@tut.fi

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

Tarmo Lipping
Tampere University of Technology
Pori, Pohjoisranta 11
P.O.Box 300
Fin-28101 Pori
Finland
E-mail: tarmo.lipping@tut.fi

INTRODUCTION

From technical 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. Usually the noise level is too low to affect the end-use of the recordings. However, in the recordings made in nature, like bird sound recordings, 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, the effects of some types of noise can be reduced using proper screening and recording equipment. Most noise types are part of the sound environment and belong to the listening experience. It is difficult to strike a good balance between noise reduction and sound environment preservation because in automatic bird sound recognition, noise complicates the feature extraction. Common methods to overcome this problem exist, 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. Example of noise insensitive feature extraction methods are wavelet decomposition or harmonic component trajectory tracking. Some birds, like the Raven (*Corvus corax*) have a very rich vocabulary of inharmonic sounds. The background noise may conceal some of the low amplitude and high frequency features of these sounds.

The classification method used in this paper is based on wavelet decomposition (SELIN 2005). Due to its fine resolution analysis in both time and frequency domains it is important to separate the sounds from background noise while preserving the fine structure of the bird sound data as well as possible.

In this paper, we tested classification of eight bird species without filtering and with two different filtering methods. The baseline classification system and the selection of bird species is explained in the paper of SELIN et al. (2006).

BIRD SPECIES

The effects of noise reduction in automated bird sound recognition were tested with 8 bird species. The selected bird species are listed in the Table 1. All sounds were recorded in Finland by Pertti Kalinainen, Ilkka Heiskanen and Jan-Erik Bruun using 44.1 kHz sampling frequency.

Sounds of the Mallard, the Greylag Goose, the Corncrake, the River Warbler and the Magpie were selected due to their inharmonic transient sounds that can be mixed in the background noise. The sounds of the Quail and the Spotted Crake are tonal, but contain some transient features. The Pygmy Owl was selected for reference purposes.

BASELINE RECOGNITION METHOD

The features from the bird sounds were calculated using wavelet packet decomposition. Daubechies 10 wavelet function and 6-level decomposition was used for feature extraction (Daubechies 1992, Selin 2005). Each bird sound was analyzed and four parameters were computed from each sound. The parameters were:

- duration of the sound (spread)
- active bandwidth of the sound (width)
- maximum average energy (energy) and
- its position (position)

Duration and active bandwidth were measured from the wavelet decomposition matrix, as seen in Fig. 1, against predefined energy threshold. The maximum energy was the highest of averaged energies from rows of the wavelet decomposition matrix. The maximum scale energy was the position (row number) of the maximum energy.

These features were the inputs of the feedforward Multilayer Perceptron (MLP) (4-15-40-3 (binary output)) network. The learning procedure was standard backpropagation, and number of training epochs was 65 with shuffled training data (Haykin 1998). The binary output produced the class number with three bits (8 classes).

FILTERING METHODS

There are several possibilities to perform noise reduction. To test the effects of noise reduction on classification results, three pre-processing scenarios were compared:

i) No filtering
ii) Adaptive filter bank based noise reduction
iii) Finite Impulse Response (FIR)-lowpass filter (with 10 kHz cutoff frequency)

"No filtering" means that the data was fed to the network without any modifications.

The filter bank noise reduction scheme divides the sound, into eight uniformly spaced frequency bands. The signal in the highest frequency band is assumed to contain mostly noise. We selected sound sample based adaptive threshold for the noise removal, which was two times the standard deviation of the signal in the highest frequency band. If the absolute value of the amplitude was bigger than the threshold level, the threshold was subtracted from the amplitude. Otherwise the amplitudes below threshold level were zeroed out. This operation was performed to all eight frequency bands using the same noise threshold level from the highest frequency band. After this operation the frequency bands were added together to form one signal. The adaptivity was gained using different threshold values for each sound frame.

The lowpass filter was a 124-tap FIR filter and its frequency response is shown in Fig. 2.

RESULTS

The training data consisted of 2278 samples and test data from 854 samples. The number of sound samples used for training and testing is presented in Table 1.

The results of the tests are summarized in Tables 2-4.

DISCUSSION

From the test data 95.1%, 96% and 100% of the samples were recognized correctly using no filtering, the adaptive filter bank and the FIR filter, respectively. Although the feature extraction based on wavelet packet decomposition is itself a robust method against the noise, as it can be seen from baseline or 'no filtering' results, it is clear that in this case both filtering scenarios enhanced the classification results.

The adaptive filter bank preserves the sound spectrum by carefully reducing the background noise. This procedure also enhances the results when compared against the baseline. FIR filtering produced the best recognition results. However, FIR filter completely removes the frequency content over 10 kHz and thus removes the unnecessary components that may affect the wavelet based feature extraction, while the adaptive filter bank will try to maintain the sound quality and environment by removing the estimated background noise from whole frequency area. Both methods are suitable for noise reduction with wavelet packet decomposition based feature extraction.

If the filtering should be done by preserving most of the sound environment, the filter bank succeeded well. If the only target is the classification, the FIR filter is easy to implement and computationally inexpensive. The recognition results using FIR filter will improve at the expense of totally destroyed sound environment. The filter bank method improved the recognition results while preserving most of the sound environment.

ACKNOWLEDGEMENTS

This work was supported by Academy of Finland grant decision 206652.

REFERENCES

Daubechies, I., 1992: Ten Lectures on Wavelets.- Philadelphia: Society for Industrial and Applied Mathematics.

Haykin, S., 1998: Neural Networks: A Comprehensive Foundation.- New Jersey: Prentice Hall

Selin, A., 2005: Bird Sound Classification Using Wavelets.- M.SC. Thesis. Tampere University of Technology.

Selin, A., Turunen, J. & Tanttu, J.T., 2006: Bird Sound Recognition with Wavelets.- *Advances of Bioacoustics 2, Dissertationes Classis IV: Historia Naturalis, Slovenian Academy of Sciences and Arts (Ljubljana)*, XLVII-3, 185-204.

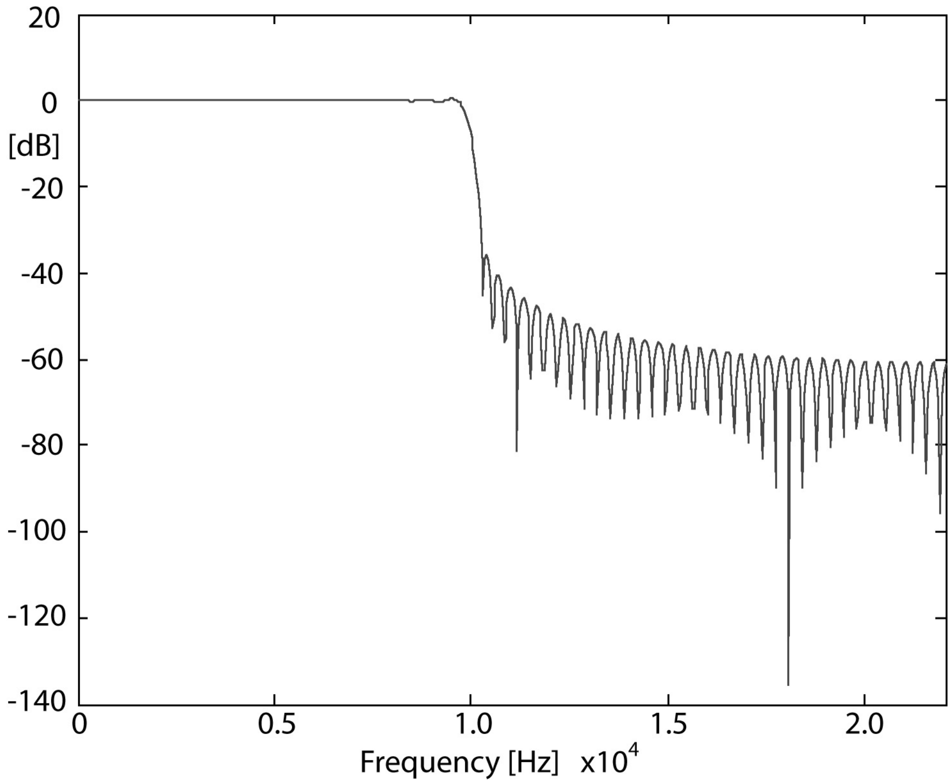

Figure 1: The frequency response of the FIR filter

Table 1: Bird species, testing and training data.

English name	Scientific name	Abbr.	Training	Testing
Mallard	*Anas platyrhynchos*	ANAPLA	138	60
Graylag Goose	*Anser anser*	ANSANS	135	59
Corncrake	*Crex crex*	CRECRE	443	110
River Warbler	*Locustella fluviatilis*	LOCFLU	890	328
Magpie	*Pica pica*	PICPIC	203	97
Quail	*Coturnix coturnix*	COTCOT	190	83
Spotted Crake	*Porzana porzana*	PORPOR	166	69
Pygmy Owl	*Glaucidium passerinum*	GLAPAS	113	48
			2278	854

Table 2: No filtering (Total 95.1%)

%	ANAPLA	ANSANS	COTCOT	CRECRE	GLAPAS	LOCFLU	PICPIC	PORPOR
ANAPLA	**100**	0	0	0	0	0	0	0
ANSANS	20	**72**	0	3	5	0	0	0
COTCOT	0	0	**97**	1	0	0	1	1
CRECRE	0	0	0	**100**	0	0	0	0
GLAPAS	0	0	0	0	**100**	0	0	0
LOCFLU	0	0	0	0	0	**100**	0	0
PICPIC	0	0	4	1	0	0	**95**	0
PORPOR	0	0	0	0	0	0	1	**99**

Table 3: Filter Bank (Total 96.0%)

%	ANAPLA	ANSANS	COTCOT	CRECRE	GLAPAS	LOCFLU	PICPIC	PORPOR
ANAPLA	**98**	2	0	0	0	0	0	0
ANSANS	1.7	**83**	1.7	5.1	1.7	5.1	1.7	0
COTCOT	0	0	**100**	0	0	0	0	0
CRECRE	1	2	0	**96**	0	0	1	0
GLAPAS	0	2	0	0	**96**	2	0	0
LOCFLU	0	0.3	0	0	0	**99.7**	0	0
PICPIC	0	0	5	1	0	0	**94**	0
PORPOR	0	0	0	0	0	0	0	**100**

Table 4: FIR-lowpass filter (Total 100%)

%	ANAPLA	ANSANS	COTCOT	CRECRE	GLAPAS	LOCFLU	PICPIC	PORPOR
ANAPLA	**100**	0	0	0	0	0	0	0
ANSANS	0	**100**	0	0	0	0	0	0
COTCOT	0	0	**100**	0	0	0	0	0
CRECRE	0	0	0	**100**	0	0	0	0
GLAPAS	0	0	0	0	**100**	0	0	0
LOCFLU	0	0	0	0	0	**100**	0	0
PICPIC	0	0	0	0	0	0	**100**	0
PORPOR	0	0	0	0	0	0	0	**100**