Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 51806, 9 pages
doi:10.1155/2007/51806

Research Article

Wavelets in Recognition of Bird Sounds

Arja Selin, Jari Turunen, and Juha T. Tanttu

Department of Information Technology, Tampere University of Technology, Pori, P.O. Box 300, 28101 Pori, Finland

Received 9 September 2005; Revised 30 May 2006; Accepted 22 June 2006

Recommended by Gerald Schuller

This paper presents a novel method to recognize inharmonic and transient bird sounds efficiently. The recognition algorithm consists of feature extraction using wavelet decomposition and recognition using either supervised or unsupervised classifier. The proposed method was tested on sounds of eight bird species of which five species have inharmonic sounds and three reference species have harmonic sounds. Inharmonic sounds are not well matched to the conventional spectral analysis methods, because the spectral domain does not include any visible trajectories that computer can track and identify. Thus, the wavelet analysis was selected due to its ability to preserve both frequency and temporal information, and its ability to analyze signals which contain discontinuities and sharp spikes. The shift invariant feature vectors calculated from the wavelet coefficients were used as inputs of two neural networks: the unsupervised self-organizing map (SOM) and the supervised multilayer perceptron (MLP). The results were encouraging: the SOM network recognized 78% and the MLP network 96% of the test sounds correctly.

1. INTRODUCTION

Nearly all birds make different kinds of sounds which are used in communication with other conspecifics and also between different species. Sounds are only produced when needed, and so all the sounds have some meaning [1, 2]. Most sounds are produced by the syrinx, which is the avian vocal organ [3]. In most species the syrinx is bipartite, so the bird can produce two notes simultaneously [4, 5]. Bird sounds can be tonal or inharmonic, which is one way to divide the bird species into groups. Inharmonic sounds are often transient and their frequency contents are very near each other. Bird vocalization contains both songs and calls. Calls are shorter and simpler than songs, and both sexes produce them throughout the year. It seems that most birds have from 5 to 15 distinct calls, and the functions of them can be, for example, flight, alarm, excitement, and so on. Some birds can have several different calls for the same function, whereas some birds use very similar calls in different circumstances to mean different things. In addition, in many species there is high individual and regional variability in phrases and song patterns [6–9]. Thus, two kinds of bird sound variability have to be taken into account in the classification. One is the variation of different sound types and another is the variation across geographic regions and among individuals.

Human ear and brain constitute an effective voice recognition system. For the human ear it is relatively easy to notice even subtle differences in sounds, whereas for the computer the recognition task is much more difficult. In bird sound research, the typical methods of classification have been listening and visual assessment of spectrograms. However, human decision is always subjective. So, the automatization of this classification process would be an important new tool for bioacoustic research [10]. Automatic classification offers new possibilities for the identification of vocal groups of birds, and may also give new tools for the classification of the sounds of other animals.

Classification of bird sounds has been studied a lot and its application range includes, for example, bird census and taxonomy [11–13]. Nevertheless, only a few studies exist where the identification of bird species by their sound is made automatically [14–19]. Most of these studies, for example, [14, 17], have focused on tonal and harmonic sounds, and are based on conventional spectral analysis methods. These methods are not well matched to inharmonic and transient sounds. In [19] inharmonic bird sounds have been classified using 19 low-level parameters of syllables. It seems, however, that the number of parameters is probably too high for an efficient recognition algorithm.

The aim of our study was to develop a computationally effective recognition method for inharmonic bird sounds,

and to investigate the applicability of the wavelet analysis for this task. The wavelet analysis has gained a great deal of attention in the field of digital signal processing [20]. It has many advantages, for example, its ability to find out both frequency and temporal information, and to analyze signals which contain discontinuities and sharp spikes. These properties are appropriate for inharmonic and transient bird sounds. In the wavelet packet transform the original signal is converted into wavelet coefficients. The orthogonal wavelet packets can be designed by hierarchical association of PR (perfect reconstruction) paraunitary filter banks [21]. Because the number of the coefficients is usually large after the decomposition and because using all wavelet coefficients as features will often lead to inaccurate results, the extraction of the most important features is essential. The feature extraction from wavelet coefficients has been studied, for example, in [22, 23]. In spite of the many advantages of the wavelet transform, it also has a disadvantage: it is time dependent. To avoid this problem, four shift invariant parameters were used as features in this study.

Artificial neural networks (ANNs) are being applied to pattern recognition and have successfully been used in the automated classification of acoustic signals including animal sounds [24–27]. The ANNs have also been used in the classification and recognition of bird sounds [28–30]. In this study, two commonly known neural networks, the unsupervised self-organizing map (SOM) and the supervised multilayer perceptron (MLP), were selected as the classifiers due to their ability to compensate discrepancies among the data. The distinguishability of bird species was first examined with the SOM, which is essentially a clustering algorithm, and after that the sound data was classified using the MLP.

2. METHODS

The model of the whole recognition process is presented in Figure 1. During the preprocessing the noise was reduced from the soundtracks. Then the soundtracks were segmented into smaller pieces which are called sounds in the sequel. During the postprocessing the sounds were checked manually. All the sounds were decomposed into the wavelet coefficients using the wavelet packet decomposition (WPD). The features were calculated from these wavelet coefficients and the feature vectors were composed. The feature vectors of the training data were introduced to the MLP and the SOM networks during the training phase. Finally, both networks were tested on separate testing data and the recognition results were examined. Altogether, the phases of the recognition process were automatic, except the checking of the sounds, which was made manually.

2.1. Preprocessing, segmentation, and postprocessing

During the preprocessing the zero mean data was normalized in the range $[-1, 1]$, and the low-frequency wind noise was reduced using a long moving average filter. Because the noise level varied a lot between the sound tracks, the noise threshold level was calculated adaptively from long-term

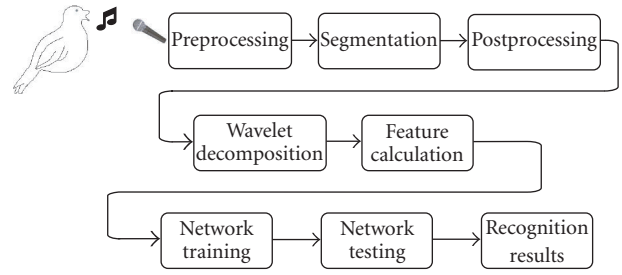

Figure 1: The recognition process.

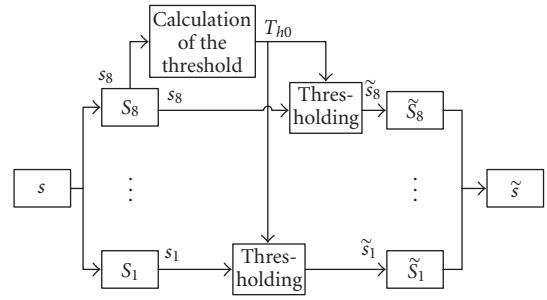

Figure 2: The noise reduction using the filter bank.

mean energy value during the segmentation. The soundtracks were extracted automatically into smaller pieces identifying the beginning and ending of each call. The soundtrack was clipped if the onset of the sound exceeded the adaptive threshold level and the end of the sound dropped under that threshold value.

During the postprocessing the interfering broadband noise was reduced from the sound signal, s, using the eight-band filter bank (cf. Figure 2).

The outputs $\tilde{s}_i(n)$ from the thresholding blocks were calculated as

$$\tilde{s}_i(n) = \begin{cases} 0 & \text{if } s_i(n) < T_{h0}, \\ \operatorname{sgn}(s_i(n))\left(|s_i(n)| - T_{h0}\right) & \text{else} \end{cases} \quad \text{for } i = 1, \ldots, 8, \tag{1}$$

where the threshold value T_{h0} was defined as 2 times the standard deviation of the output s_8 after preliminary tests. Reduction of the noise emphasized the essential information of the bird sound. At the end of the postprocessing all sounds were checked manually and verified consistently. A few sounds were recorded in a very noisy environment or they were in inseparable groups, and were therefore rejected during the manual checking.

2.2. Wavelet packet decomposition

The wavelet packet analysis was used for the signal decomposition [31, 32]. In the WPD the signal s is split into approximation (A) and detail (D) parts. Due to the downsampling, aliasing occurs in the WPD tree. This aliasing changes the

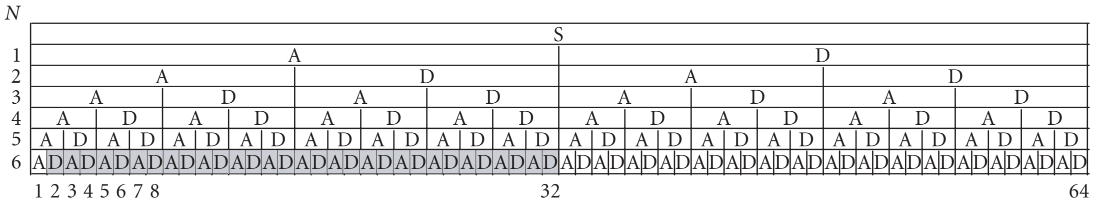

FIGURE 3: The symmetric wavelet decomposition tree. The grey bins are used in the proposed method.

frequency order of some branches of the tree [33]. The symmetric wavelet decomposition tree is illustrated in Figure 3, where the WPD tree is put in an increasing frequency order from the left to the right.

The preliminary tests showed that the best decomposition level (N) was six. Thus, the signal s was split into $2^6 = 64$ parts, which are called bins in the sequel. The bin number 1 contained so low frequencies that proved to be irrelevant for the recognition. Because the bins 33–64 also proved to be irrelevant, the wavelet coefficients were calculated from bins 2–32 marked grey in Figure 3.

There are several wavelet families that have proved to be particularly usable [34]. The Daubechies wavelet family (dbN) was selected, because in it both scaling and wavelet functions are compactly supported and they are orthogonal. The 10 dB was selected for the wavelet function, because the preliminary tests showed that it compromised the best decomposition results of the tested alternatives with the selected bird sounds.

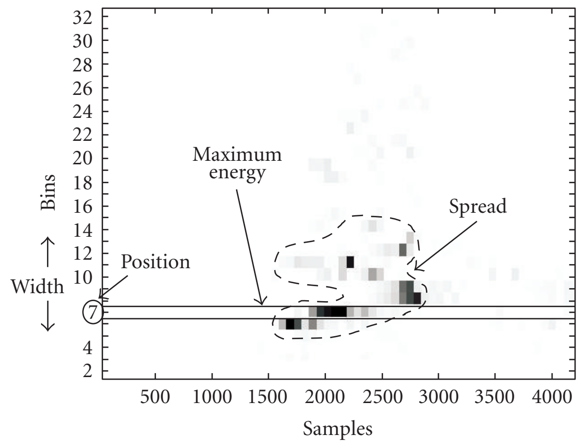

FIGURE 4: The four shift invariant features: maximum energy, position, spread, and width. The larger absolute values of the wavelet coefficients are presented with the darker color.

2.3. *Features*

As mentioned before, the main disadvantage of the wavelet transform is its time dependence. That is why the four shift invariant parameters were selected as features. These four features, *maximum energy*, *position*, *spread*, and *width* are illustrated in Figure 4.

The number of the WPD coefficients of each bin is denoted as n_c. The bin energy $E_B(r)$ of the wavelet coefficients c of bin r was defined as

$$E_B(r) = \sum_{n=1}^{n_c} c^2(n,r), \quad r = 2,3,\ldots,32, \tag{2}$$

and the average energy $\widetilde{E}_B(r)$ of each bin r was defined as

$$\widetilde{E}_B(r) = \frac{E_B(r)}{n_c}. \tag{3}$$

The largest average energy value

$$E_m = \max_r \left(\widetilde{E}_B(r)\right) \tag{4}$$

was then searched, and it is called the *maximum ene*rgy E_m of the sound. The *position* P represents the number of the bin r, in which the maximum energy was located.

The *spread* S was calculated as

$$S = \frac{1}{\#J} \sum_{(q,r)\in J} c^2(q,r), \tag{5}$$

where q is the number of the sample and r is the number of the bin. J is a set of index pairs (q,r) for which $c^2(q,r) > T_{h1}(r)$. In (5) $\#J$ is the number of elements (cardinality) of the set J. So, the spread S is a sum of the average energies of those coefficients whose energy exceeded the threshold value T_{h1}. After the preliminary test with the data the threshold value $T_{h1}(r)$ was calculated as

$$T_{h1}(r) = \frac{\widetilde{E}_B(r)}{6} \tag{6}$$

from the average energy $\widetilde{E}_B(r)$ of bin r.

The fourth feature, the *width* W represents the number of bins which satisfy the inequality

$$E_B(r) > T_{h2}, \tag{7}$$

where the threshold value T_{h2} was selected as 1.3 after preliminary tests with the data.

Finally all four features were normalized, in order to be comparable with one another. The normalization levels were defined after preliminary tests with the data. The maximum energy E_m was normalized as

$$\widetilde{E}_m = \frac{E_m}{n_B}, \tag{8}$$

Table 1: Selected set of bird sounds used in this study.

Scientific abbr.	Scientific name	English name	Sound type	MLP training	SOM training	Testing
ANAPLA	*Anas platyrhynchos*	Mallard	Inharmonic	138	113	60
ANSANS	*Anser anser*	Greylag goose	Inharmonic	135	113	59
COTCOT	*Coturnix coturnix*	Quail	Tonal	190	113	83
CRECRE	*Crex crex*	Corncrake	Inharmonic	443	113	110
GLAPAS	*Glaucidium passerinum*	Pygmy owl	Pure harmonic	113	113	48
LOCFLU	*Locustella fluviatilis*	River warbler	Inharmonic	890	113	328
PICPIC	*Pica pica*	Magpie	Inharmonic	203	113	97
PORPOR	*Porzana porzana*	Spotted crake	Tonal	166	113	69
—	—	—	—	2278	904	854

where n_B is the number of the coefficients of the bin which exceeded the T_{h1}. The position P was normalized as

$$\widetilde{P} = \frac{P}{2^N/4} = \frac{P}{16}. \tag{9}$$

The spread S was normalized as

$$\widetilde{S} = \frac{S}{100} \tag{10}$$

and the width W as

$$\widetilde{W} = \frac{W}{20}. \tag{11}$$

Thus, $31 \times n_c$ WPD coefficients were reduced to four normalized features: maximum energy $\widetilde{E}_m$, position $\widetilde{P}$, spread $\widetilde{S}$, and width $\widetilde{W}$. These four features formed the final feature vector for recognition. The main reason for the normalization was the SOM, which yields better recognition results if the inputs are in the same scale. In addition, the training time of the SOM network is shorter with normalized inputs.

2.4. Classifiers

Two commonly known neural networks, unsupervised self-organizing map (SOM) [35] and supervised multilayer perceptron (MLP) [36], were used as classifiers. The neural networks were selected due to their ability to compensate discrepancies in the data. This is one way to deal with the individual and regional variability of bird vocalizations. The motivation for using unsupervised and supervised networks was to verify the predefined decisions of the supervised MLP against the unsupervised SOM, and to compare their relative performance. In the SOM the four-dimensional data was mapped into two-dimensional space. The SOM clusters the data so that neighbouring clusters are quite similar, while more distant clusters become increasingly diverse [35]. The low and high variability between the sounds of the species can be seen from the compactness of the clusters. Thus, in this study the distinguishability of the species was first examined with the SOM, and after that the classification was made with the MLP.

In the SOM training the calculated feature vectors were introduced to a 10×10-size SOM network. The other sizes, for example, 6×6, 8×8, and 12×12, of the network were also tested. However, the chosen size yielded best recognition results. The SOM network was trained for up to 3000 epochs using the training data (cf. Table 1). The results did not improve although the number of the epochs was changed.

After preliminary tests, the selected MLP architecture was 4-15-40-3. Each output was finally rounded to 0 or 1, and then three output bits of each sound were converted into numbers 1–8, which was enough for classes of eight bird sounds. The MLP network was trained for up to 65 epochs and the mean square error goal was 0.0001. After the training, it became obvious that all the nodes, and the weighting and bias parameters of the MLP network were needed, which means that none of the outputs of the nodes was too close to zero. Both networks were tested on separate testing data after the training.

3. THE BIRD SOUND DATA

Our main purpose was to study the efficient recognition of inharmonic or transient bird sounds. The sampling rate of the sound data, F_s, was 44.1 kHz and 16-bit accuracy was used. The data was analyzed in the Matlab environment [37], and the Wavelet Toolbox [34] was utilized. The idea was to choose such bird species whose sounds are inharmonic and sounds which resemble one another. This is the reason why the inharmonic sounds of the mallard, the greylag goose, the corncrake, the river warbler and the magpie were selected. The sounds of the quail and the spotted crake are tonal, but contain some transient features, for example, irregular pitch period. The pure tonal territorial song of the male pygmy owl was chosen as a reference sound.

In the classification, the variation of different sound types in every species has to be taken into account by examining each sound type separately. That is why only one type of call of each species was used in this study. However, several types of calls of the greylag goose were included, because these calls are very similar to one another. Hence, it was

tested how the greylag goose can be recognized using many types of calls. In addition, a sufficient number of recordings of those eight species was available quite easily and the quality of the recordings was sufficient. The data of the selected eight species is summarized in Table 1. The table contains scientific abbreviations and names, English names, and sound types. Also the number of sounds in the training and testing is indicated.

The sounds were recorded in Finland by Pertti Kalinainen, Ilkka Heiskanen, and Jan-Erik Bruun. There were totally 3132 sounds which were divided into training data (2278 sounds) and testing data (854 sounds). The training and testing data were from different tracks. It turned out that if there were the same number of training data of each group, the SOM network yielded better results. Thus, in the case of the SOM network the training data was reduced to 113 samples per species.

The typical spectrograms and corresponding wavelet coefficient figures of eight species that were used in this study are presented in Figure 5. As can be seen, the wavelet transform compresses the energy of the coefficients more than traditional Fourier transform in spectrograms. Only the very essential information is preserved after the WPD.

4. RESULTS

4.1. Results using the SOM

The clustering result of the SOM network after training is illustrated in Figure 6.

The areas marked with letters present how sounds of each bird species were situated in the 10×10 SOM network (cf. Section 2.4) after the overlapping nodes had been analyzed. The SOM network was examined node by node and the outliers were labelled. The species which had most sounds in a particular node won and the possible other sounds were classified as outliers. If two or more different species had the same number of sounds in a particular node, all were classified as outliers. If no species won, the node was classified as unspecified. If no sound is situated in the node, it was classified as empty node. Unspecified nodes are marked with black color and empty nodes with grey color in Figure 6. In the SOM, compact clusters represent the species with little variation between sounds, and, respectively, the scattered clusters represent the species with large variation. As it can be seen, for example, the test sounds of the river warbler (R) form a compact and uniform area, whereas the sounds of the greylag goose (G) spread out in a broad area. The SOM clustered 87% of training sounds correctly.

The confusion matrix of Table 2 illustrates the recognition result of the SOM network after the trained network had been tested on the test sounds. The rows of the confusion matrix show how each species is recognized. All the test sounds of the river warbler (LOCFLU) were recognized correctly, as can be seen from the diagonal of the matrix. Altogether, 7% of the test sounds were unspecified and 15% were recognized wrongly. It should be noticed that only 51% of the sounds of the greylag goose were recognized correctly, and 23% of the sounds were recognized unspecified. That might result from the fact that several types of calls of the greylag goose were included in the study. Altogether, 92 sounds of all 854 test sounds were recognized wrongly. A total of 78% of the test sounds were recognized correctly with the SOM network.

4.2. Results using the MLP

Table 3 contains the recognition result of the MLP network. All the test sounds of the quail (COTCOT) and the spotted crake (PORPOR) were recognized correctly. Again, the recognition result of the sounds of the greylag goose was poor, and the reason might be the same as with the SOM network. Twenty-four sounds of all the test sounds were recognized wrongly. Altogether, 96% of the test sounds of the eight bird species were recognized correctly with the MLP network.

5. DISCUSSION AND CONCLUSIONS

Our purpose was to study how inharmonic and transient bird sounds can be recognized efficiently. The results of this study are very encouraging. The results indicate that it is possible to recognize bird sounds of the test species using neural networks with only four features calculated from the wavelet packet decomposition coefficients.

Segmentation plays an important role in sound recognition, because incorrectly segmented sounds will probably be classified wrongly. In most cases, segmentation is the most complicated and challenging part of the whole recognition process. However, it is quite difficult to make it totally automatic. Noise reduction goes hand in hand with successful segmentation. The segmentation is even more difficult if the sound tracks are very noisy. In this study the segmentation and noise reduction were implemented so that the original sound information of the target species remained as intact as possible. After the automatic segmentation, all the sounds were checked manually. The noise reduction was done using an eight-band filter bank, which reduced the irrelevant noise information and emphasized the essential information of the bird sound. The main purpose of the preprocessing was to control the signal quality so that all sounds were comparable with each other.

The selection of the wavelet function and the decomposition level are the most important phases of the WPD. In this study the 10 dB was selected for the wavelet function and the level of the decomposition was selected to be six after preliminary testing. The preliminary tests were used because the authors do not know any reliable algorithm for selecting the wavelet function and the decomposition level properly. The preliminary tests indicated that the 10 dB wavelet function and the 6th decomposition level compromised the best decomposition results with selected bird sounds.

The four features were calculated from the wavelet packet decomposition coefficients. Many kinds of other features were calculated from the coefficients and they were also tested. However, the chosen four features: maximum energy,

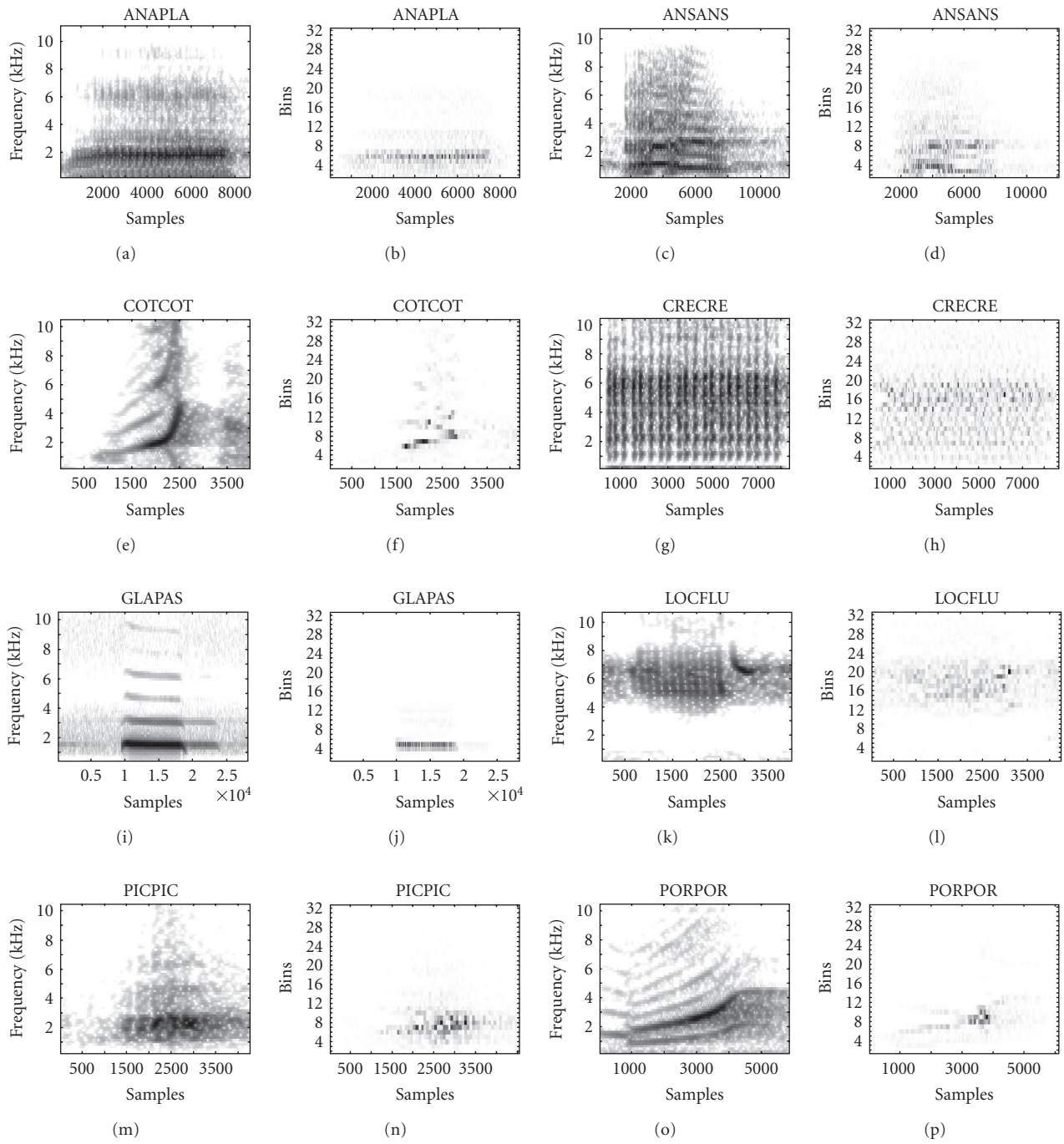

Figure 5: (a), (c), (e), (g), (i), (k), (m), and (o) typical spectrograms and (b), (d), (f), (h), (j), (l), (n), and (p) corresponding wavelet coefficients of the eight species used in this study are presented. The frequency and bins are bounded to 11.025 kHz (Fs/4), because at the higher frequencies there was no essential information. In the spectrograms the darker colors represent the higher energies of the sound. Correspondingly, the larger absolute values of the coefficient are presented with the darker color in the adjacent wavelet coefficient figures. The range of the coefficients is $[-5, 5]$.

position, spread, and width, described and separated the sounds of the eight bird species best.

The data of the eight bird species that was used in this study was divided so that there were about 70% training data and 30% testing data. Both networks, the SOM and the MLP, were first trained and then tested on separate data. The training data contained very probably sounds of seven mallard, nine graylag goose, three quail, eight corncrake, five pygmy owl, two river warbler, six magpie, and three spotted crake individuals. The testing data was selected from tracks different from the training data and it was also very probably from different individuals. So, the testing data consisted of

Table 2: The confusion matrix in percentage terms when using the SOM network.

%	ANAPLA	ANSANS	COTCOT	CRECRE	GLAPAS	LOCFLU	PICPIC	PORPOR	Unspecified
ANAPLA	**78**	20	0	0	0	0	0	0	2
ANSANS	24	**51**	0	0	0	0	0	2	23
COTCOT	0	0	**87**	0	0	0	8	4	1
CRECRE	0	0	0	**83**	0	0	1	0	16
GLAPAS	0	15	0	0	**75**	0	0	0	10
LOCFLU	0	0	0	0	0	**100**	0	0	0
PICPIC	1	0	2	1	0	0	**58**	38	0
PORPOR	0	0	0	0	0	0	9	**91**	0

Table 3: The confusion matrix in percentage terms when using the MLP network.

%	ANAPLA	ANSANS	COTCOT	CRECRE	GLAPAS	LOCFLU	PICPIC	PORPOR
ANAPLA	**98**	2	0	0	0	0	0	0
ANSANS	2	**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**

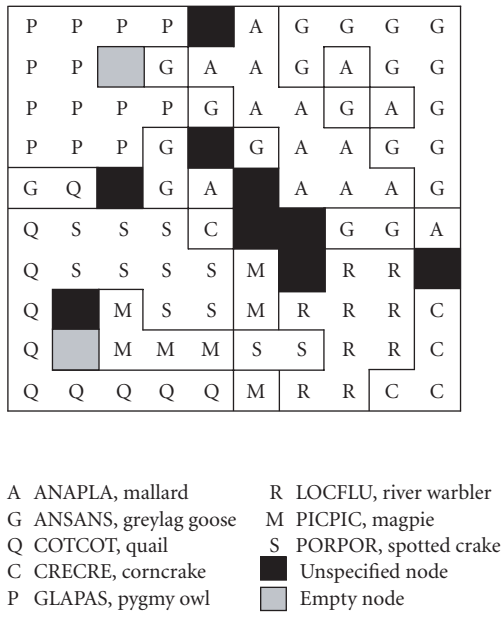

Figure 6: The clustering result of the 10×10 SOM network after training.

sounds of two mallard individuals, four graylag goose, two quail, two corncrake, and two pygmy owl individuals, and one river warbler, one magpie, and one spotted crake individuals.

In conclusion, the SOM classified 78% and the MLP 96% of the test sounds correctly. After the testing of both networks, all wrongly recognized sounds were manually examined and labelled. The test result showed that 24 sounds were recognized wrongly using the MLP network. In the SOM network 39 of test sounds were unspecified and 92 sounds were recognized wrongly. After plotting and examining all the wavelet packet coefficient figures of the misrecognitions, the reason for the most wrong recognitions became obvious. Firstly, the coefficient pattern of the misrecognitions was shifted so that two features, the position and the width, were strayed. Secondly, the wrong recognition resulted presumably from false segmentation or low signal-to-noise ratio.

The proposed method provides quite a robust approach to sound recognition, particularly to the inharmonic and transient bird sounds. The variability among the bird sounds within and between the species was taken into account using neural networks in the classification. The sounds of the selected eight species vary only slightly. Also, the variation across geographic regions was insignificant, because all the sounds were recorded in Finland.

In conclusion, the results presented in this paper are very encouraging. They indicated that it is possible to recognize bird sounds using neural networks with only four features calculated from the wavelet packet coefficients. Although the neural networks have many benefits, such as their ability to learn and therefore generalize the variability of the data, there is a long way to go before the recognition system beats the human ear. When using neural networks in the pattern

classification, there has to be a fixed number of classes into which activations are classified. Hence, the disadvantage of the neural networks is the fixed number of output classes, that is, closed set of species. When more species need to be classified, the network has to be retrained all over again before it can be tested on a new set of birds.

Although the tested algorithms proved to be quite robust recognition methods for a limited set of birds, the proposed method cannot beat a human expert listener. A human expert listener can identify birds with almost 100% accuracy by using a priori knowledge and environmental or other context-dependent information for classification, whereas our proposed method uses only a short recording without any other information. In [19] the inharmonic bird sounds were recognized with nearest neighbor classifier using Mahalanobis distance measure with 74% accuracy, whereas in this study the SOM classified 78% and the MLP 96% of the inharmonic bird sounds correctly. On the other hand, the results are quite incomparable to other methods, because the test set of birds was limited and the features were calculated differently.

The method tested in this study is intended for automatic monitoring of birds that are living in a predefined area or night time active birds or migratory birds whose probability of existence is known beforehand. The continuous monitoring of the same birds is costly and time-consuming. Thus, the aid of automatic recognition in field work might be desirable. The algorithm must be fine-tuned in a way that it recognizes the predefined and limited set of birds correctly either leaving out or storing the uncertain or unknown sounds for manual checking.

Automatic recognition presents a new method for identifying and differentiating bird species by their sounds, and may offer new tools also for bird researchers. However, the automatic recognition of bird species is by no means an easy task. The fact that sounds and calls vary among species and the same species might have many call types make automatic recognition even more difficult. In this demanding task the wavelet transform has proven to be an efficient method to be taken into consideration.

6. ACKNOWLEDGMENTS

The authors would like to thank Pertti Kalinainen, Ilkka Heiskanen, and Jan-Erik Bruun for their recordings and Docent Mikko Ojanen for his helpful comments on biological issues. The authors also wish to thank the reviewers for their encouraging comments and suggestions. This Research was funded by the Academy of Finland under research Grant 206652 and by the Ulla Tuominen's Foundation.

REFERENCES

[1] C. K. Catchpole and P. J. B. Slater, *Bird Song: Biological Themes and Variations*, Cambridge University Press, Cambridge, UK, 1995.

[2] D. E. Kroodsma, *The Singing Life of Birds: The Art and Science of Listening Birdsong*, Houghton Miflin, Boston, Mass, USA, 2005.

[3] C. H. Greenewalt, *Bird Song: Acoustics and Physiology*, Smithsonian Institution Press, Washington, DC, USA, 1968.

[4] S. A. Zollinger, T. Riede, and R. A. Suthers, "Production of nonlinear phenomena in the Northern Mockingbirds (*Minus polyglottos*)," in *Proceedings of the 1st International Conference on Acoustic Communication by Animals*, pp. 283–284, College Park, Md, USA, July 2003.

[5] R. A. Suthers, G. Beckers, S. A. Zollinger, E. Vallet, and M. Kreuzer, "Mechanisms of vocal complexity in birds," in *Proceedings of the 1st International Conference on Acoustic Communication by Animals*, pp. 237–238, College Park, Md, USA, July 2003.

[6] J. W. Bradbury, "Parrots and technology," in *Proceedings of the 1st International Conference on Acoustic Communication by Animals*, pp. 29–30, College Park, Md, USA, July 2003.

[7] M. C. Baker and D. M. Logue, "Population differentiation in a complex bird sound: a comparison of three bioacoustical analysis procedures," *Ethology*, vol. 109, no. 3, pp. 223–242, 2003.

[8] J. G. Groth, "Call matching and positive assortative mating in red crossbills," *The Auk*, vol. 110, no. 2, pp. 398–401, 1993.

[9] M. S. Robb, "Introduction to vocalizations of crossbills in Northwestern Europe," *Dutch Birding*, vol. 22, no. 2, pp. 61–107, 2000.

[10] V. B. Deecke and V. M. Janik, "Automated categorization of bioacoustic signals: avoiding perceptual pitfalls," *Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 645–653, 2006.

[11] A. M. Elowson and J. P. Hailman, "Analysis of complex variation: dichotomous sorting of predator-elicited calls of the Florida scrub jay," *Bioacoustics*, vol. 3, no. 4, pp. 295–320, 1991.

[12] J. G. Groth, "Resolution of cryptic species in appalachian red crossbills," *The Condor*, vol. 90, no. 4, pp. 745–760, 1988.

[13] S. F. Lovell and M. R. Lein, "Song variation in a population of Alder Flycatchers," *Journal of Field Ornithology*, vol. 75, no. 2, pp. 146–151, 2004.

[14] A. Härmä, "Automatic identification of bird species based on sinusoidal modelling of syllables," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 5, pp. 545–548, Hong Kong, April 2003.

[15] A. Härmä and P. Somervuo, "Classification of the harmonic structure in bird vocalization," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 5, pp. 701–704, Montreal, Quebec, Canada, May 2004.

[16] N. Mesgarani and S. Shamma, "Bird call classification using multiresolution spectrotemporal auditory model," in *Proceedings of the 1st International Conference on Acoustic Communication by Animals*, pp. 155–156, College Park, Md, USA, July 2003.

[17] J. T. Tanttu, J. Turunen, A. Selin, and M. Ojanen, "Automatic feature extraction and classification of crossbill (*Loxia spp.*) flight calls," *Bioacoustics*, vol. 15, no. 3, pp. 251–269, 2006.

[18] P. Somervuo and A. Härmä, "Bird song recognition based on syllable pair histograms," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 5, pp. 825–828, Montreal, Quebec, Canada, May 2004.

[19] S. Fagerlund and A. Härmä, "Parametrization of inharmonic bird sounds for automatic recognition," in *proceedings of the 13th European Signal Processing Conference (EUSIPCO '05)*, Antalya, Turkey, September 2005, Proceedings on CD-ROM.

[20] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, 1991.

[21] A. K. Soman and P. P. Vaidyanathan, "Paraunitary filter banks and wavelet packets," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, pp. 397–400, San Francisco, Calif, USA, March 1992.

[22] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 83–88, 1999.

[23] R. Learned, "Wavelet packet based transient signal classification," M.S. thesis, Massachusetts Institute of Technology, Cambridge, Mass, USA, 1992.

[24] S. M. Phelps and M. J. Ryan, "Neural networks predict response biases of female tungara frogs," *Proceedings of the Royal Society—Biological Sciences (Series B)*, vol. 265, no. 1393, pp. 279–285, 1998.

[25] V. B. Deecke, J. K. B. Ford, and P. Spong, "Quantifying complex patterns of bioacoustic variation: use of a neural network to compare killer whale (Orcinus orca) dialects," *The Journal of the Acoustical Society of America*, vol. 105, no. 4, pp. 2499–2507, 1999.

[26] J. Placer and C. N. Slobodchikoff, "A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs," *Behavioural Processes*, vol. 52, no. 1, pp. 1–9, 2000.

[27] A. Thorn, "Artificial neural networks for vocal repertoire analysis," in *Proceedings of the 1st International Conference on Acoustic Communication by Animals*, pp. 245–246, College Park, Md, USA, July 2003.

[28] A. L. McIlraith and H. C. Card, "Birdsong recognition using backpropagation and multivariate statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2740–2748, 1997.

[29] A. M. R. Terry and P. K. McGregor, "Census and monitoring based on individually identifiable vocalizations: the role of neural networks," *Animal Conservation*, vol. 5, no. 2, pp. 103–111, 2002.

[30] P. Somervuo and A. Härmä, "Analyzing bird song syllables on the self-organizing map," in *Proceedings of the Workshop on Self-Organizing Maps (WSOM '03)*, Hibikino, Japan, September 2003, Proceedings on CD-ROM.

[31] A. Boggess and F. J. Narcowich, *A First Course in Wavelets with Fourier Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 2001.

[32] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa, USA, 1992.

[33] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*, Academic Press, Boston, Mass, USA, 1992.

[34] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelet Toolbox for Use with Matlab*, MathWorks, Natick, Mass, USA, 2000.

[35] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, Germany, 2001.

[36] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College, New York, NY, USA, 1994.

[37] MathWorks, "Matlab Software Homepage," June 2005, http://www.mathworks.com.

Arja Selin was born in Janakkala, Finland, on May 2, 1970. She received her M.S. degree in 2005. Currently she is preparing her doctoral thesis in signal processing and pattern recognition.

Jari Turunen received his M.S. and Ph.D. degrees in 1998 and 2003, respectively, from Tampere University of Technology. He currently works as a Senior Researcher at Tampere University of Technology, Pori. His current research interests cover topics such as speech and signal processing.

Juha T. Tanttu was born in Tampere, Finland, on November 25, 1957. He received his M.S. and Ph.D. degrees in electrical engineering from Tampere University of Technology in 1980 and 1987, respectively. From 1984 to 1992, he held various teaching and research positions at the Control Engineering Laboratory of Tampere University of Technology. He currently holds Professorship of Information Technology at Tampere University of Technology, Pori.