Classification of Bird Species by Using Key Song Searching: A Comparative Study *

Andreas Franzen
School of Electrical Engineering
Chalmers University of Technology, Sweden
e8frasse@etek.chalmers.se

Irene Y.H. Gu
Dept. of Signals and Systems
Chalmers University of Technology, Sweden
i.gu@ieee.org

Abstract — *To better understand the environment we are living, especially animals and birds around us, we need to study their behavior and the way of their communication. This paper addresses the problem of classifying bird species of interests using the digital signals of recorded bird songs. First, through comparisons of speech and bird song signals and the experiments, we propose a simple model (similar to that of the speech) for generating synthetic bird songs. We then propose a key-bird-song searching method for the recognition of bird species of interest. This is possible since bird songs appear to be simpler as compared to human speech. A hierarchical classification method is then suggested. In the coarse level of the classification, only candidate songs from those birds whose time-dependent coupled sound patterns are 'close' to that of the species of interest are chosen as the candidates. In the fine level, time-frequency 'formant' trajectory-related features from the candidate songs are used for the classification. A case study is conducted for the recognition of a selected bird species, the Great Tit. Preliminary experimental results from using five different bird species and 87 songs have shown promising results in recognizing the selected bird species of interest, with less than 3% of classification errors.*

Keywords: bird species classification, bird song synthesis, key song search, bird song recognition.

1 Introduction

To better understand the environment we are living, especially animals and birds around us, we need to study their behavior and the way of their communication. This paper addresses the problem of classifying bird species of interest through analyzing the digital signals of recorded bird songs.

Several previous work have been reported on bird song analysis and identification. These methods include: using Wigner-Ville distribution for spectral analyzing of bird songs[7]; using artificial neural networks, statistical analysis and Bayesian classification methods [4, 5, 6, 8] for identification of bird species; and using dynamic time warping and hidden Markov model [1, 3] for bird song recognition.

For classification of bird species, recognizing different types of bird songs plays an important role. To understand which features best represent the bird song signals, the paper will first make a comparative study between speech and bird songs. Based on that, a simple bird song production model will be proposed. From our experiments, we have discovered that the same (or, a slightly simplified) speech model is also valid for modelling the bird songs. This discovery has made it possible to generate synthetic bird songs from a mathematical model.

Next, we propose a novel method for classifying bird species by searching the key songs that characterize the bird species of interest. This idea is very similar to those used in speech recognition, however, applying to bird songs. The method is new (from the best of our knowledge) as far as bird species recognition is concerned.

Since bird song is simpler than human speech, less features will be required as compared to speech recognition. The features selected for bird song classification include the time-dependent bird song patterns (i.e., periodically coupled sound bursts), and the time-trajectories of pseudo 'formants' of the bird song. A hierarchical classification method is then proposed. In the first (coarse) level of the classification system, candidate songs are selected from those birds whose time-dependent coupled sound features are 'close' to that of the species of interest. In the 2nd (fine) level, a detailed feature matching is performed by using a set of time-varying pseudo formant (i.e. time-varying frequency peaks) trajectories and the associated profiles.

The remaining of the paper is organized as follows. In Section 2, a brief description of the bird species in Sweden is given. In Section 3, a bird song production model is proposed through comparative studies and experiments. In Section 4, we first examine the waveforms and spectrograms of several bird species. Based on the observations, we then propose a key-bird-song recognition method. In Section 5, a case study is presented with

*0-7803-7952-7/03/$17.00

the great tit being the bird species of interest. Seven features are identified for the songs of great tits. In Section 6, experiments are described and some preliminary results are included. Finally, conclusions are given in Section 7.

2 Bird Types and Songs

There are many different species of birds living around us. As shown in the pictures of Fig.1, typical species of birds in Sweden include [2]:

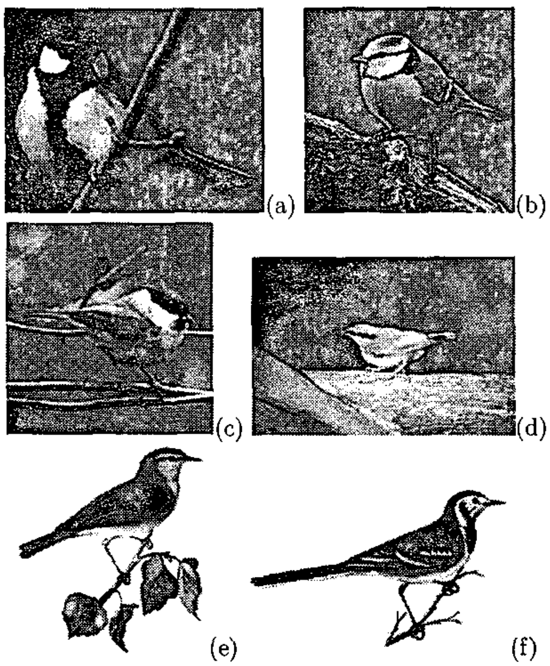

Figure 1: Picture of typical birds in Sweden. (a) Great tit; (b) Blue tit; (c) Marsh tit; (d) Nuthatch; (e) Willow warbler; (f) White wagtail.

- *Great tit* (Swedish name: Talgoxe): Great tits are the largest tit in Sweden and is a very common type of birds. Usually their songs are loud and clear, and are often heard during the spring time.
- *Blue tit* (Swedish name: Blåmes): Blue tits are easily distinguished by their blue crown. A blue tit is also much smaller than the great tit.
- *Marsh tit* (Swedish name: Entita): Marsh tits are fairly common birds, however can easily be mixed up with willow warblers.
- *Nuthatch* (Swedish name: Nötvecka): Nuthatches are easily recognizable by their way of climbing tree trunks. A nuthatch has the ability of climbing down a tree with the head first, which other Swedish birds do not have.
- *Willow Warbler* (Swedish name: lövsångare): Willow Warbler is known as the most common bird in Sweden. Their songs, softer and more fluent, are quite different from the other birds listed in this paper.
- *White wagtail* (Swedish name: Sädesärla): White wagtails stay in Sweden during the spring and summer time, and migrate to Germany and Northern Africa regions before the winter. A white wagtail bird often chases insects on the roads and has the characteristics of frequently wagging its tail.

3 Modelling of Bird Song signals: A Comparative Study

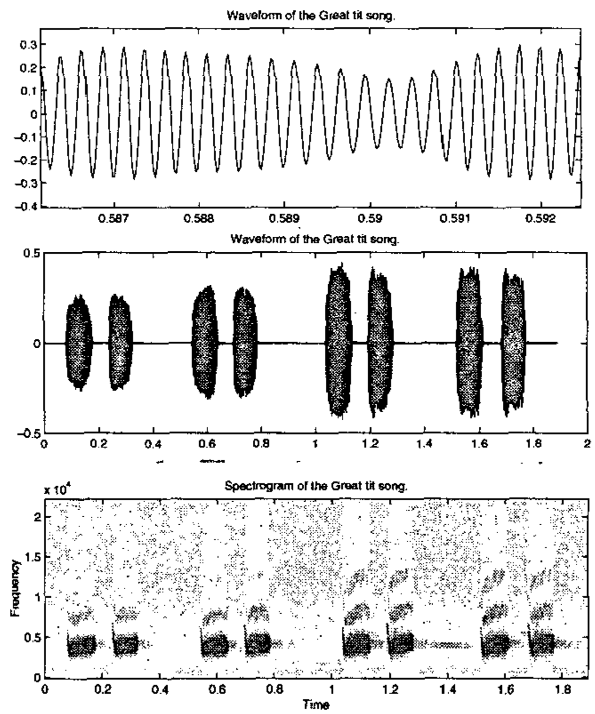

Figure 2: Signals from a recoded song of a great tit. From top to bottom: (a) Zoomed waveform from the bird song in (b); (b) digital waveform of a bird song; (c) corresponding spectrogram.

In this section, an attempt is made to find a proper signal processing model for the digital signals of bird songs. Before modelling the bird songs, let us first examine the waveforms and spectrograms of a bird song. Fig.2 shows the waveform and spectrogram of a typical song from a great tit. As can be observed from Fig.2, the waveforms of bird songs bear some similarities to that of human speech, especially they mostly contain quasi-periodic signals (similar to the voiced speech signals corresponding to vowels in speech). However, there are some differences. For example, the waveforms of bird songs appear to have amplitude modulations (which are not very obvious in the speech waveforms). Further, the bird song signal has a wider frequency range and a much higher fundamental ('pitch') frequency than those in the human speech. As can be seen from Fig.4(c) the 'pitch' (or fundamental) frequency of bird is about 4000Hz, much higher than that of the human speech (usually in the range of several hundred Hz).

The basic idea here is to use a simplified version of speech model as the model for bird song production. The simplification is made at the model input where only the voiced excitation is used (for most of bird songs). Fig.3 shows the proposed bird song production model, where a simple human speech production model is also included for comparison.

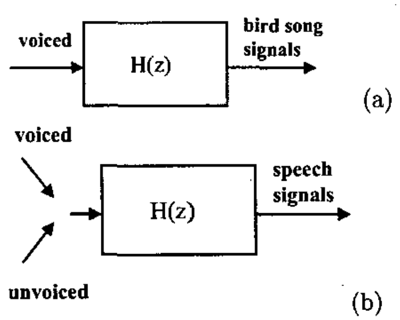

Figure 3: (a) The proposed bird song production model. (b) a simple speech production model (for comparison). The filter transfer function in the above models are defined as $H(z) = b_0 / \left(1 + \sum_{i=1}^{N} a_i z^{-i}\right)$.

In order to test whether this model is valid, the inverse models are firstly used to estimate the LPC parameters and the residuals of time-varying models with a natural bird song as the input. Repetitive impulses (with the estimated fundamental frequencies of the bird song) are then input to the estimated models to generate a synthetic bird song. Our test results show that the model indeed requires much higher 'pitch (or, fundamental)' frequencies for bird songs as comparing to that for the human speech. Further, our tests show that the resulting synthetic bird songs have a rather good quality, and sound very similar to the natural songs. As an example, Fig.4 shows the waveform, the spectrogram and the pitch contour of the synthetic bird song from a great tit. By comparing the original and the synthetic bird songs, we can see that the proposed bird song production model is valid.

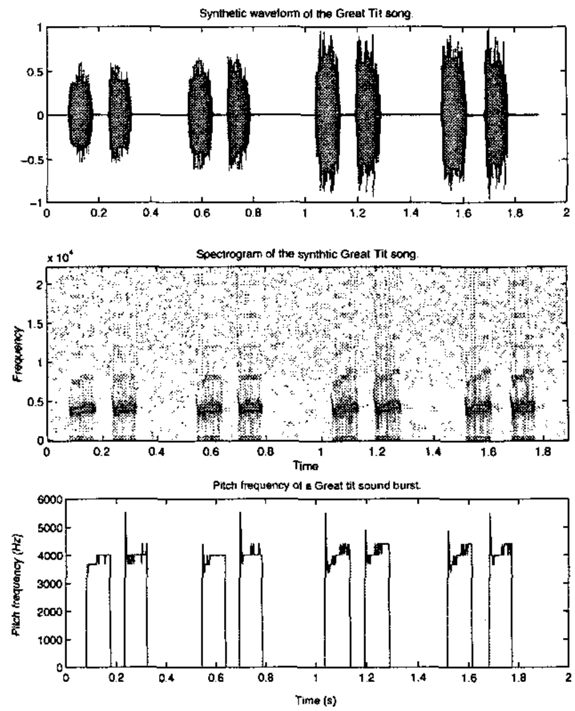

Figure 4: The synthetic signal of the bird song corresponding to the original one in Fig.2(b). From top to bottom: (a) synthetic waveform; (b) corresponding spectrogram; (c) estimated pitch frequency contour used as the model input.

4 Recognizing Bird Species of Interest via Key-Song Searching

4.1 Observations and the basic ideas

In order to recognize a certain type of birds (or a bird species of interest), we introduce a novel method for classifying bird species by searching key songs that characterize the bird species of interest. The basic idea here again bears some similarities to speech recognition where key speech sometimes is required to be picked up in order to trace a particular piece of information of interest. The idea is novel as far as the bird song recognition is concerned. By taking into account that bird songs are simpler than the human speech, such a key-song-searching approach may lead to a simple way of bird species recognition. The advantage of this method is that, instead of requiring the extraction of features for all bird species and the successful classification of all birds, the problem is now reduced to a much smaller sub-problem as the searching of key features associated with one species and subsequently recognizing it.

Before applying any strategies for key song searching, let us first examine the different features revealed in the waveforms and spectrograms of some bird songs. Fig.5

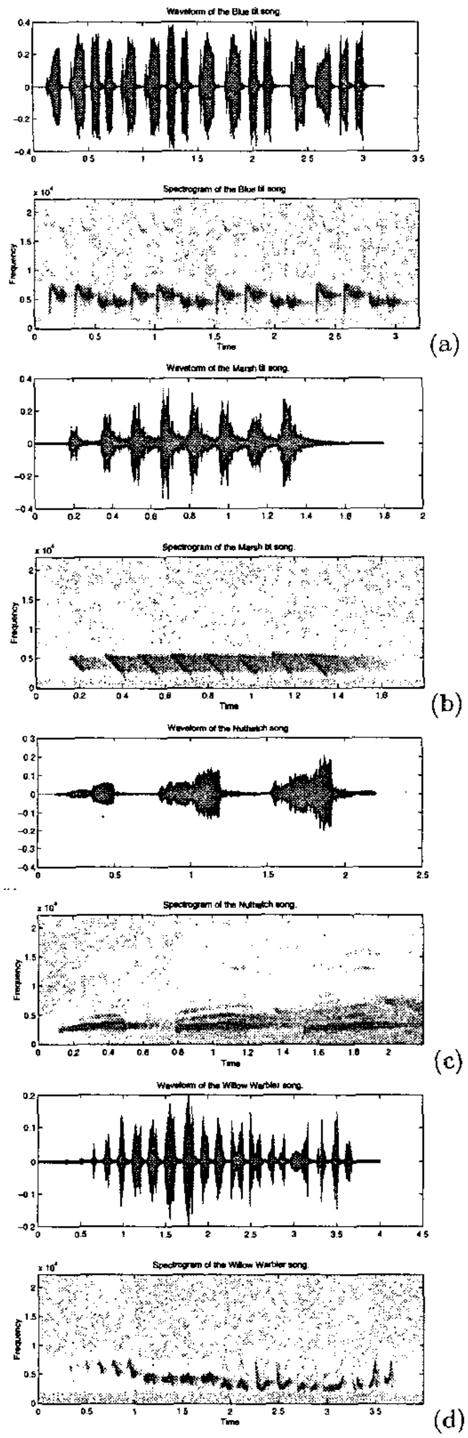

Figure 5: Waveforms (odd rows) and Spectrograms (even rows) of bird songs from four different bird species. (a) Blue Tit; (b) Marsh Tit; (c) Nuthatch; (d) Willow Warbler.

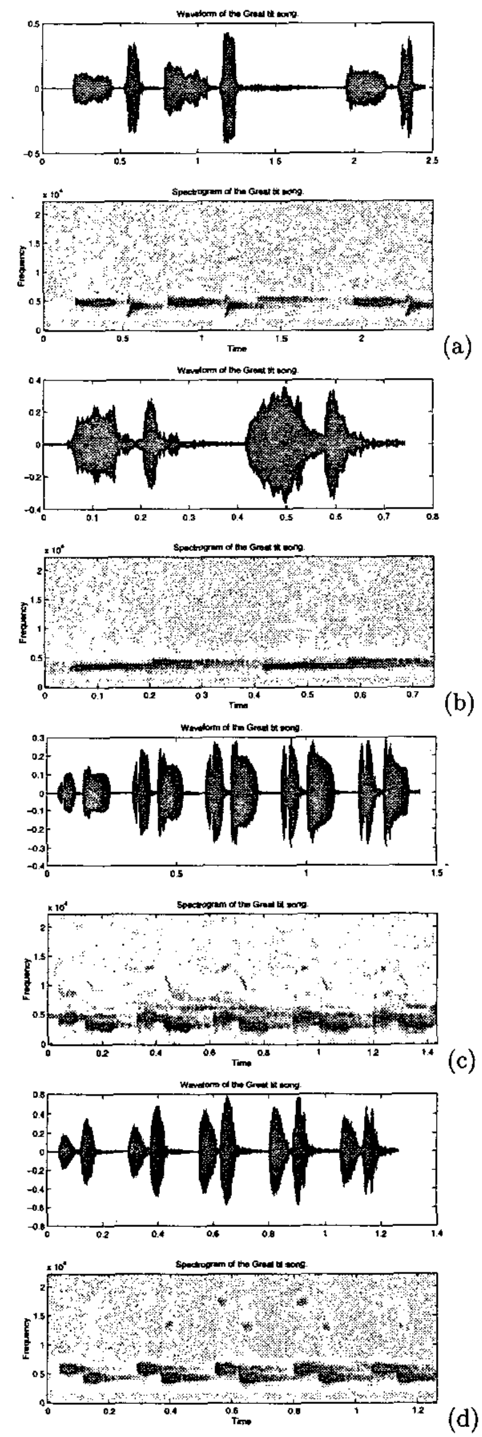

Figure 6: Waveforms (odd rows) and spectrograms (even rows) of different bird songs from the Great Tits.

shows the waveforms and spectrograms of bird songs selected from four different bird species, while Fig.6 shows the waveforms and spectrograms of different songs from one selected bird species, the great tit.

From Fig.5 one can see that the songs from different bird species have many differences. From the waveform shapes of different bird species, one can observe that each species tends to sing songs with groups of coupled bursts (e.g. in pairs, trios or quartets). Further, a song usually contains a periodic repetition of these (groups of) bursts. This *time-domain feature* may play an important role for humans to recognize the bird species. Next to this, there are many different *spectral features.* For example, the first 'formant' (or, the resonant frequency of bird's vocal tract) in the lowest frequency side is often much stronger than the remaining of the formants at the higher frequency side. Further, different bird species tends to have very different coupled formant frequencies as well as different changing trends in formant trajectories.

From Fig.6 one can see that different songs from great tits vary greatly, however, some common features exist. For example, the songs from the great tits are most likely to appear as periodic repetitions of paired bursts. The frequency of the 1st formant in one of the bursts is often (although not always) higher than the other one in the pair.

4.2 The proposed method

Based on the above observations, we propose a key bird song searching method for recognition of one bird species of interest.

A hierarchical classification method is proposed as follows. In the first level (coarse classification), candidate songs are extracted from those birds whose time-dependent coupled sound features are 'close' to that of the species of interest. In the 2nd level (fine classification), features related to the time-varying 'formant' (or spectral peaks) trajectories and the associated profiles are used. The selected features in 2 levels include:

- level-1 (for coarse classification): using time-dependent group sound patterns, i.e., periodically-coupled sound bursts in bird songs;
- level-2 (for fine classification): using time-varying trajectories of coupled 'formants' (spectral peaks) in bird songs.

The feature used in level-1 is a significant feature since periodically-coupled sound bursts plays an important role for humans to recognize the bird species of interest. Most songs from other bird species are expected to be rejected in the coarse level as not being the selected bird species (about 77% in our case study, see Section 6), leaving only a small number of candidate songs for a further classification in level-2.

5 Case Study: Feature Extraction for the Recognition of Great Tits

In the following, a case study is conducted for bird species recognition. *In our study we choose the great tit as the bird species of interest.* Emphasis is put on the most common male great tit songs consisting of pairs of bursts.

5.1 Selection of features

The selected features should mainly characterize the songs from the species of interest, i.e., the great tit in this case. Since most bird songs from various species appear to contain periodic repetition of groups of bursts, it is important to first segment these bursts in the time-domain before extracting the features.

Segmentation

Let us consider bird songs from great tits: a song from a great tit contains repetitive pairs of bursts, each pair can be described by four segments, leading to 4 time durations: the time duration (TD) of first burst ($b1$), of first silence ($s1$) , of 2nd burst ($b2$), and of 2nd silence ($s2$) (denoted respectively as TD_{b1}, TD_{s1}, TD_{b2}, TD_{s2}).

Features

For recognizing the great tit species, seven different features are selected for each pair of bursts from a given song as follows:

Feature-1 (x_0) is associated with the *minimum and maximum ranges of time-durations* obtained from the statistics of supervised training from the songs of a chosen species. For example, for pairs of bursts associated with great tits, the ranges of time duration consist of $[TD^{b1}_{min}, TD^{b1}_{max}]$, $[TD^{s1}_{min}, TD^{s1}_{max}]$, $[TD^{b2}_{min}, TD^{b2}_{max}]$, and $[TD^{s1}_{min}, TD^{s1}_{max}]$, where TD_{min} and TD_{max} denote the minimum and maximum time duration from the training bird songs of each species.

Feature-2 (x_1) is associated with the spectrum smoothness around the 1st dominant formant. This is defined as the averaging number of local spectral extrema with the small window around the formant.

Feature-3 (x_3) is associated with the normalized dynamic range of the dominant formant, defined as the averaging normalized magnitude difference between the peak of the formant and its left valley.

Feature-4 (x_3) is associated with the level of domination and the sharpness of the dominant formant. This is defined as the width around the formant frequency within which the spectrum contains a pre-selected percentage of total energy. The more dominant formant is, the narrower this width is.

Feature-5 (x_4) is associated with the change (increase or decrease) of formant frequency trajectories. The standard deviations of the dominant formant frequency whose values exceed a certain threshold are averaged.

Feature-6 (x_5) is associated with the combination of time-durations in each burst group. For example, for a group equals to a pair, the feature is defined as

$$x_5 = \frac{TD_{s1}}{TD_{s2}} \, min\{\frac{TD_{b1}}{TD_{b2}}, \frac{TD_{b2}}{TD_{b1}}\} / \left(1 + c|\bar{f}_{b1} - \bar{f}_{b2}|\right)$$

where $\bar{f}_{b1}$ and $\bar{f}_{b2}$ are the averaging dominant formant within the 1st and 2nd burst of the pair, and c is a scaling factor. For the Great Tit, the 2nd silence duration is usually longer than the 1st one (i.e. $\frac{TD_{s1}}{TD_{s2}}$ should be small). Further, the two bursts usually have different time durations (i.e., $min\{\frac{TD_{b1}}{TD_{b2}}, \frac{TD_{b2}}{TD_{b1}}\}$ should be small). The formant frequencies in the two bursts are usually different (i.e., $|\bar{f}_{b1} - \bar{f}_{b2}|$ should be large). Therefore, the smaller the x_5, the more likely the song belongs to the Great Tit.

Feature-7 (x_6) is associated with difference between periodic groups within each song. For each song, there is a periodic repetition of groups of bursts. As these repetitions bear much similarity, the difference should be small over different neighboring groups. This is measured by the difference of dominant formant frequency trajectories between the two immediate neighboring groups averaged over the bursts and the groups within each song.

Feature-1 (i.e. x_0) is used in the level-1 of the system. While feature-2 to feature-7, defined as the elements in the feature vector $x = [x_1 \; x_2 \; x_3 \; x_4 \; x_5 \; x_6]^T$, are used in the level-2. It is worth to notice that even though a song from a unknown bird species may have different number of bursts in the group other than two, all song bursts are treated in pairs (as those in the great tits).

6 Experiments and Preliminary Results from the Case Study

For our case study, i.e. recognizing songs from the great tit species from songs among a total of five bird species. The sampling rate for digital bird song waveforms is f_s=44.1 KHz. After segmenting a giving bird song (that belongs to one of the five bird species) into pairs of bursts, seven features described in Section 5.1 are extracted from each pair of bursts in the song.

The supervised training

Table 1 lists the names of five bird species, the number of birds and songs, and the number of paired bursts from each bird species used for the supervised training.

Table 1: The names of bird species; the number of birds, songs and pairs of bursts in each species that were used for the supervised training.

Name of species	No. birds	No. songs	No. pairs
Great tit	8	8	29
Blue tit	12	40	158
Marsh tit	2	10	23
Nuthatch	4	16	75
Willow warbler	1	6	40

To observe whether the selected features are effective, the separability of these features for different bird species are examined. Since the dimension of feature space is six, we shall only observe feature spaces up to a dimension of three by combining several features. This will give us a rough idea on how good these features are. Fig.7 shows the histograms of six different features for five bird species. Fig.8 shows several 2D feature spaces obtained by combining 2 features as specified, and Fig.9 shows one of the 3D feature spaces by combining 3 features as specified. Further, by using feature-1, x_0, the system was able to retain all paired song bursts (29 pairs) from the great tits, and to remove about 77% (or, 228 pairs out of the total 296 pairs) pairs of bursts from the remaining four other bird species.

From Fig.7, one can see that each single feature x_i, i=1,...6, provides some separability for the great tit and the remaining four species. From Fig.8, one can see that by combining several features, the feature spaces for different bird species are more separable than that in Fig.7. From Fig.9, one can see that there is a great improvement of separability of feature spaces for great tit and the remaining 4 species (in fact, a complete separation is achieved for the given training data). Since all 6 features provide some (correlated) information on bird species, feature optimization can be applied to minimize the feature correlation (e.g., by using singular value decomposition) and to subsequently reduce the feature space dimension by only using the principal features.

The testing process

Different songs from the five species were used for the testing process as compared to those used in the training process: for the great tits, half of the songs were used for the training, and the remaining half were used for the testing; for the other four bird species, all songs were newly collected. Table 2 shows the number of birds and songs, and the number of paired bursts from each bird species used for the testing process.

The test results are shown in Table 3. As can be seen from Table 3, all the great tits were correctly recognized (i.e., 100% recognition rate) in our tests. For the songs from the remaining four bird species, the total false alarm classification rate (i.e., songs from other species being wrongly recognized as the great tit) is 2.79%. These results from the testing process using preliminary amount of bird songs have shown that the

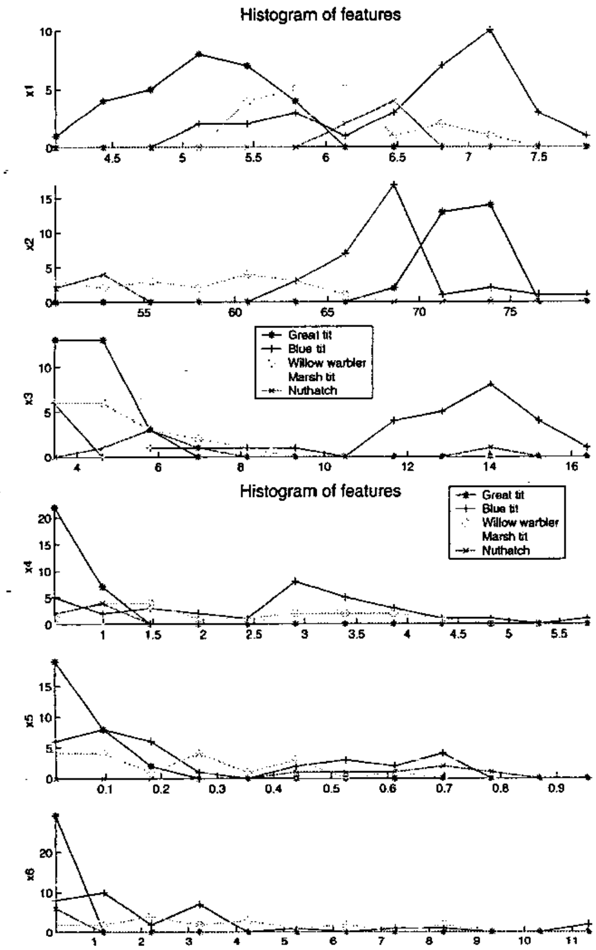

Figure 7: Histograms of song features x_1, x_2, x_3, x_4, x_5, x_6 from five different bird species. Features in the histograms are marked by: '*' in black color for great tits, '+' in blue color for the blue tits', 'o' in green color for willow warblers, '.' in yellow color for marsh tits, and 'x' in red color for nuthatches.

Table 2: The names of bird species; the number of birds, songs and pairs of bursts in each species that were used for the testing processes.

Name of species	No. birds	No. songs	No. pairs
Great tit	8	8	38
Blue tit	4	14	54
Marsh tit	7	32	119
Nuthatch	16	28	133
Willow warbler	3	5	53

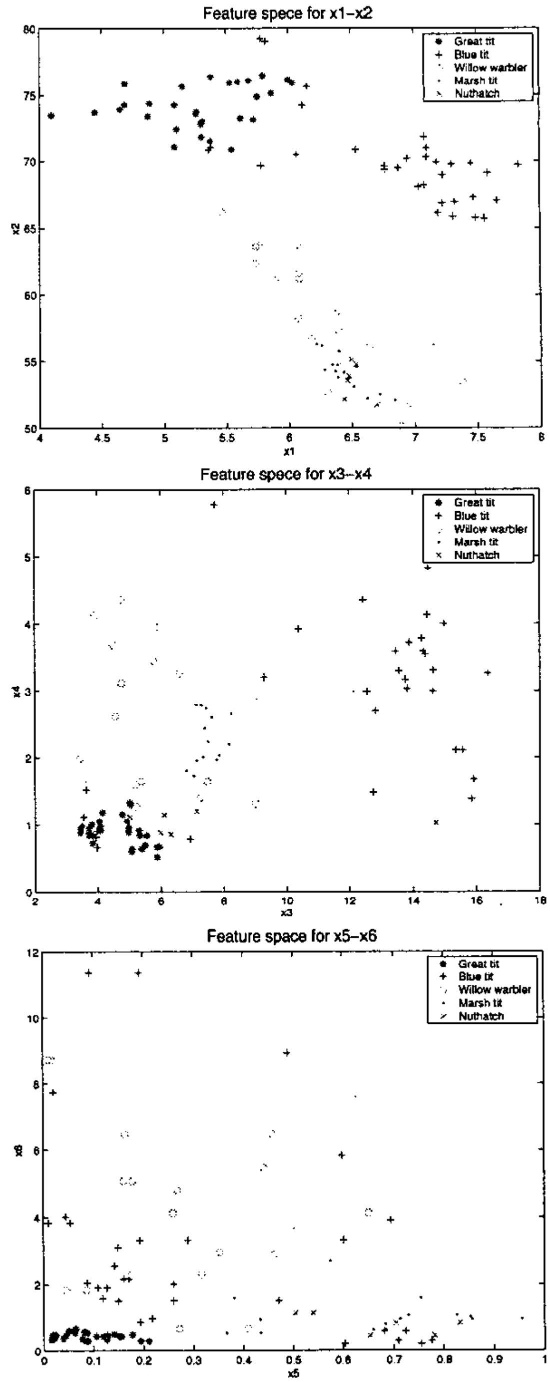

Figure 8: Examples of 2D feature spaces obtained by combining 2 specified features defined in Section 5.1. (a) feature space (x_1, x_2); (b) feature space (x_3, x_4); (c) feature space (x_5, x_6).

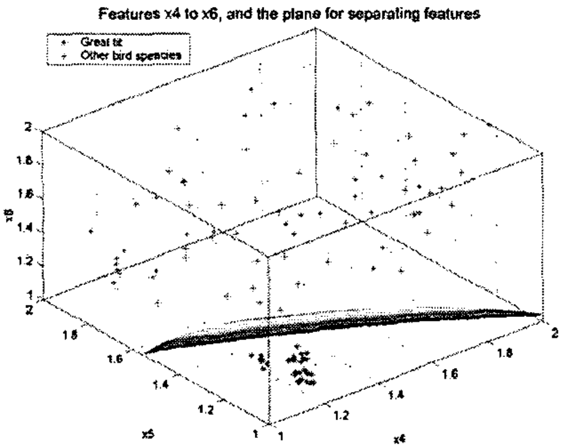

Figure 9: Examples of a 3D feature space obtained by combining 3 specified features (x_4, x_5, x_6) defined in Section 5.1. Features are marked by: '*' for great tits, '+' for the remaining four species.

Table 3: The results from the classification using the data listed in Table 2.

name of species	classified as great tits (%)	classified as non-great tits (%)
Great tit	100.0	0.0
Blue tit	0.0	100.0
Marsh tit	21.6	78.4
Nuthatch	10.0	90.0
Willow warbler	0.0	100.0

proposed method is promising for classifying a selected bird species of interest.

7 Conclusions

In this paper we have proposed a simple bird song production model and a key-song-searching method for recognition of bird species of interest. Our investigations have been conducted through comparing bird song processing and speech processing. Our experiment results have shown that the model is effective in generating synthetic bird songs resembling the original ones. A hierarchical classification method is proposed for the recognition of bird species of interest by key song searching. A case study for recognizing the great tits has been conducted, with seven features defined and extracted. Examining the feature sub-spaces using 80 bird songs from five different species for the training has shown good separability between the great tits and the remaining four other bird species. Our preliminary testing results using 87 different bird songs from five species have demonstrated that the proposed key-song searching method is promising for the successful classification of bird species of interest (with a correct classification rate of 100% and a false alarm classification rate of less than 3%).

Acknowledgement

The bird pictures in Fig.1 were collected from the following website: http://atschool.eduweb.co.uk/jblincow/ngallery.htm, http://www.talgoxen.com, and http://www.camacdonald.com/ birding/Barker. The bird songs used in our tests were obtained from the CD-roms of 'All the bird songs of Britain and Europe (4)' from Jean C. Roche., and of 'Fågelsång i Sverige' from Benny Andersson, Lars Sevensson and Dan Zetterström.

References

[1] S.E. Anderson, A.S.Dave, D. Margoliash, "Template-based automated recognition of birdsong syllables from continuous recordings", J. Acoust. Soc. Am., Vol.100(2), pp.1209-1219, 1996.

[2] Mark Beaman and Steve Madge, The Handbook of Bird Identification for Europe and the Western Palearctic, Princeton University Press, 1998.

[3] J.Kogan, D.Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study", J. Acoust. Soc. Am., Vol.103(4), pp.2185-2196, 1998.

[4] A.L. McIlraith, H.C. Card, "Bird song identification using artificial neural networks and statistical analysis", IEEE 1997 Canadian Conf. on Electrical and Computer Engineering (CCECE '97), pp.63 - 66 vol.1, 25-28, May, 1997.

[5] A.L. McIlraith, H.C. Card, "A comparison of backpropagation and statistical classifiers for bird identification", Vol.1 , pp.100-104, in Proc. International Conference on Neural Networks, June 1997.

[6] A.L. McIlraith, H.C. Card, "Birdsong recognition with DSP and neural networks", in proc. IEEE conf. Communications, Power, and Computing, (WESCANEX '95), pp.409 -414, vol.2, May 1995.

[7] C.Rogers, "High resolution analysis of bird sounds", in proc. IEEE international conf. Acoustics, Speech, and Signal Processing, (ICASSP-95). pp.3011-3014 vol.5, 1995,

[8] K. Sasaki, M.Yamazaki, "Vector compression of bird songs spectra in water sites by using the linear prediction method and its application to an automated Bayesian species classification", pp.1083 - 1088, in proc. 38th Annual Conference of the SICE, 28-30 July 1999.