

**Západočeská univerzita v Plzni
Fakulta aplikovaných věd**

AUTOMATICKÁ IDENTIFIKACE A VERIFIKACE PTÁKŮ

Ing. Ladislav Ptáček

**disertační práce
k získání akademického titulu doktor
v oboru Kybernetika**

Školitel: Doc. Ing. Luděk Müller, Ph.D.

Katedra: Kybernetiky

Plzeň 2016

**University of West Bohemia
Faculty of Applied Sciences**

**AUTOMATIC BIRD
IDENTIFICATION AND VERIFICATION**

Ing. Ladislav Ptáček

**DOCTORAL THESIS
submitted in partial fulfilment of the requirements
for the degree Doctor of Philosophy
in the field of
Cybernetics**

**Advisor: Doc. Ing. Luděk Müller, Ph.D.
Department of Cybernetics**

Pilsen, 2016

Čestné prohlášení

Prohlašuji, že jsem tuto disertační práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Boršově nad Vltavou dne 30.5.2016

Poděkování

Rád bych poděkoval mému školiteli panu doc.Lud'ku Müllerovi za veškerou podporu a vstřícnost po celou dobu mého studia. Obrovský dík patří dr.Lukášovi Machlicovi, jehož pomoc byla zcela zásadní. Děkuji dr.Janu Vaňkovi, který vždy provázel své rady povzbuzujícím suchým humorem. Jsem vděčný panu prof.Josefu Psutkovi za jeho laskavou trpělivost. Díky patří dr.Pavlu Linhartovi, původci myšlenky automatického rozpoznávání ptáků.

Též bych rád poděkoval sestře Janě za pomoc se zpracováním literatury, sestře Míše za hlídání, otci Ladislavovi za povzbuzování a zejména pak mamince Janě za trvalou pomoc během celého studia. Jsem vděčný mým dítkům Kačce a Matouškovi za rozveselování během psaní. A velký dík patří mé ženě Hance za trpělivost a podporu při studiu.

Acknowledgement

I would like to express my thanks to my advisor Mr. doc. Luděk Müller for all his support and helpfulness throughout all my studies. Great thanks also belong to Dr. Lukáš Machlica, whose help was absolutely crucial. I would like to thank Dr. Jan Vaněk, who has always spiced his advise with his specific encouraging humour. I also thank Mr. prof. Josef Psutka for his kind patience. Thanks belong also to Dr. Pavel Linhart, the parent of the idea of automatic recognition.

I also want to thank my sister Jana for assistance with literature processing, my sister Míša for babysitting, my father Ladislav for support, and my mom Jana for steady help during my studies. I am grateful to my children Kačka and Matouš for cheering me up when writing. And special thanks belong to my wife Hanka for her patience and support.

Content

1	Introduction	1
1.1	Overview	1
1.2	Thesis motivation	3
2	Thesis goals	5
3	State of the Art	6
3.1	Birds	6
3.1.1	Overview	6
3.1.2	Passerine	6
3.1.3	Hearing	6
3.1.4	Vocalization	11
3.1.5	Bird song propagation	19
3.1.6	Ringling	19
3.2	Data recording	19
3.2.1	Masking	21
3.2.2	Process automation and microphone arrays	22
3.2.3	Data processing	22
3.3	Speaker recognition	24
3.3.1	Speaker identification	24
3.3.2	Speaker verification	24
3.3.3	Speaker recognition methods in ornithology	24
3.4	Recognition system overview	25
3.4.1	Parametrization	26
3.4.2	Gaussian Mixture Model (GMM)	28
3.4.3	Training	29
3.4.4	Decision	30
3.4.5	Score calibration	31
3.4.6	Evaluation	32
3.4.7	Probability model methods	33
3.4.8	GMM-UBM Speaker verification system	33
3.4.9	Expectation-maximization EM	35
3.4.10	JFA	37

3.4.11	i-Vector.....	38
3.5	Feature extraction.....	38
3.5.1	Vocal tract model, cepstral coefficients.....	39
3.5.2	Hamming window.....	41
3.5.3	Pre-emphasis.....	42
3.5.4	Mel frequency cepstral coefficients.....	42
3.5.5	Perceptual Linear Predictive analysis.....	44
3.5.6	Linear prediction cepstral coefficients.....	46
4	Development Framework.....	48
4.1	Matlab.....	48
4.1.1	Experiment manager.....	48
4.1.2	Recording classification, data set.....	50
4.1.3	Feature extraction.....	51
4.1.4	Support modules.....	52
4.1.5	Model estimation.....	52
4.1.6	Verification and identification.....	52
4.1.7	Experiment evaluation.....	52
4.2	Speaker verification tool.....	52
4.2.1	Flow diagram.....	52
4.2.2	Input and output data.....	53
4.2.3	Results.....	54
4.2.4	Using the SV tool.....	54
4.3	iVector tool.....	55
5	Bird individual identification using as-is recordings.....	57
5.1	Introduction.....	57
5.2	Bird song data.....	57
5.2.1	Chiffchaff.....	57
5.2.2	Recording.....	58
5.2.3	Recording quality.....	59
5.3	Task definition.....	61
5.4	System description.....	62
5.5	Experiment evaluation.....	64
5.5.1	Parameters.....	65
5.6	Results.....	66

5.7	Contribution	69
5.8	Summary	69
6	Identification Vectors	71
6.1	Introduction	71
6.2	Experiment evaluation.....	71
6.3	Results	72
6.4	Contribution	73
6.5	Summary	73
7	Bird Audiogram Unified Equation	74
7.1	Introduction	74
7.2	Audiogram equation definition.....	74
7.3	Result.....	79
7.4	Contribution	82
7.5	Summary	83
8	Bird Adapted Filter.....	84
8.1	Introduction	84
8.2	Parameters optimization	84
8.3	Filter distribution definition	85
8.4	Experiment evaluation.....	87
8.5	Results	89
8.6	Contribution	92
8.7	Summary	92
9	Improving automatic bird identification by data merging.....	93
9.1	Introduction	93
9.2	Method principle	93
9.3	Results	94
9.4	Contribution	97
9.5	Summary	97
10	Bird Song Database	98
10.1	Introduction	98
10.2	Requirements.....	99
10.3	Application functionality.....	100
10.3.1	Backend.....	100
10.4	Database model	100

10.5	General description.....	101
10.5.1	List of taxonomy	102
10.5.2	LOV Birds	102
10.5.3	Mass data import	103
10.5.4	LOV Individuals.....	104
10.5.5	LOV Recordings.....	104
10.6	Contribution	104
10.7	Summary	105
11	Mashona mole-rat identification.....	106
11.1	Introduction	106
11.2	Mashona mole-rat.....	106
11.3	Data	107
11.4	Vocalization.....	107
11.5	Testing procedure	108
11.6	Results	109
11.7	Contribution	109
11.8	Summary	109
12	Conclusion.....	111
12.1	Application of an automatic bird identification.....	112
12.2	Future work	113
12.3	Personal note	113
13	Appendix.....	115
	References	2

List of Figures

Figure 3.1: ISO equal-loudness curve, 40 dB.....	7
Figure 3.2 : An audiogram distortion example if one measures just a low number of frequencies. The black dotted line represents the ISO human equal-loudness curve for 40 dB. The blue solid line demonstrates how a human audiogram may look like if only six frequencies are measured: 100 Hz, 500 Hz, 1 kHz, 3 kHz, and 10 kHz.....	9
Figure 3.3: Average audibility curves human and “average bird” [CAT08].....	10
Figure 3.4: Harmonic complex discrimination [MAR04].....	10
Figure 3.5: Audibility curves (a) of song sparrows (open circles) and swamp sparrows (filled circles) compared to the power spectra (b) of their songs [CAT08].....	11
Figure 3.6: Animal body mass and frequencies of vocalization [CAT08].....	12
Figure 3.7: Cross-section of the syrinx of a brown thrasher. (T) thermistors, (MTM) medial tympani form membranes.....	13
Figure 3.8: One song of <i>Chiffchaff</i>	14
Figure 3.9: Two syllables divided into elements.....	14
Figure 3.10: Score of a Lazuli Bunting song. Created by composer Olivier Messiaen (Fr., 1908-1992).	14
Figure 3.11: Bengali Finch song structure [MAR04].....	15
Figure 3.12: Human voice. Studio record, women, Czech language, business news, impassive. Frequency range 0 – 4 kHz.....	16
Figure 3.13 <i>Chiffchaff</i> song, recorded in wood, morning. Band width 0 – 11,025 kHz.....	16
Figure 3.14 <i>Chiffchaff</i> song, recorded in suburban areas, morning. Bandwidth 0 – 22,050 kHz.....	17
Figure 3.15: <i>Chiffchaff</i> song detail, one syllable. Recorded in the woods, sample frequency 44.1 kHz.	17
Figure 3.16: Tree pipit song detail, one syllable. Recorded in the woods, sample frequency 22.05 kHz.	18
Figure 3.17: One syllable spectrogram: <i>Chiffchaff</i> (left), and tree pipit (right). The vertical bar graphs illustrate a band energy.....	18
Figure 3.18: A voice of a Forest Owlet (<i>Athene Blewitti</i>). A wide spectrum noise from 500 Hz is visible, caused by a background noise and a recording machine.....	21
Figure 3.19: Continuous <i>Chiffchaff</i> record (raw record), length 45 s.....	23
Figure 3.20: Single <i>Chiffchaff</i> song, cut off from the raw recording, length 5.5 s.....	23
Figure 3.21: General outline of the GMM-UBM recognition system.....	26

Figure 3.22: Samples in the rectangular window are weighted by the Hamming window, FFT is performed, filtration utilizing triangular filters is carried out, and a cepstral feature vector is extracted. Subsequently, the window is shifted to its new location and the extraction process is repeated.	27
Figure 3.23: Triangular filter banks spread linearly in Hz scale.	28
Figure 3.24: Given a set of one dimensional feature vectors (x-axis), the Gaussian Mixture Model with three mixture components which best describes the data set (in the sense of maximal likelihood (4)) is given by the solid line. Note that the GMM is formed from 3 normal distributions each weighted by the relative number of vectors it encloses.	29
Figure 3.25: Verification: False and correct decision.	32
Figure 3.26: Probability ratio-based speaker detection system [REY00].	34
Figure 3.27: Parameter extraction and feature vectors origination.	39
Figure 3.28: Human vocal tract.	39
Figure 3.29: Vocal tract, equivalent circuit diagram.	40
Figure 3.30: Vocal tract, simplification.	40
Figure 3.31: Deconvolution.	41
Figure 3.32: Hamming window a) Time domain, b) Frequency.	42
Figure 3.33: Mel-frequency cepstral coefficients computing, data diagram.	43
Figure 3.34: Characteristic Mel-frequency [mel] and frequency [f] domains.	43
Figure 3.35: Mel filter bank.	44
Figure 3.36.: Linear filter bank.	44
Figure 3.37: Block diagram of PLP speech analysis.	45
Figure 3.38: LPC coefficients calculation stages [BIM04].	46
Figure 3.39: LPC, cepstral coefficients.	47
Figure 4.1: Experiment manager, block diagram.	49
Figure 4.2: Experiment entities relationship.	49
Figure 4.3: Experiment parameters definition.	50
Figure 4.4: File lists definition parameters.	51
Figure 4.5: Feature extraction, block diagram.	51
Figure 4.6: An Excel file result.	52
Figure 4.7: Function of SV tool, flow diagram.	53
Figure 4.8: A block diagram of the Identity Vectors process.	56
Figure 5.1.: Chiffchaff (<i>Phylloscopus collybita</i>). © Kristyna Felendova.	58
Figure 5.2: Single song of a chiffchaff male with the “chiff” and “chaff” syllable type examples highlighted.	58

Figure 5.3: Spectrogram of real recording used for the experiments without any pre-processing (cut off songs, de-noising, etc.). The Chiffchaff song is masked by another male, different species, wind blowing noise, continuous traffic noise, etc.	61
Figure 5.4: Identification and verification.....	62
Figure 5.5: Outline of the VAD detector.....	63
Figure 5.6: Spectrogram of a recording and a result of VAD. See false songs detections at 0:14.7, 0:15.05 and a segment containing an ornithologist speech (from 0:17.2 to 0:18.0).....	63
Figure 5.7: Parametrization of the recordings. The output parameters are formed into feature vectors.	64
Figure 5.8: Two types of errors: False Acceptance and False Rejection.	65
Figure 5.9: Detail result. Accuracy for each particular bird, see Table 5.....	67
Figure 6.1: An example of a song extracted from the raw recordings. Low overlap level, standard noise, duration 4.8 sec.	71
Figure 6.2: An example of a song extracted from the raw recordings. High overlap level, high noise, duration 3.7 sec.	72
Figure 6.3: An example of a song extracted from the raw recordings. Low overlap level, low noise, duration 2.5 sec.	72
Figure 7.1: Original data example: Audiogram, B-04, Emu (<i>Dromaius novaehollandiae</i>) [DOO02b].	75
Figure 7.2: Table 16 graphical legend.....	77
Figure 7.3: An example of the fitting (Mallard Duck).	78
Figure 7.4: Final sum of squares of residuals for all five functions. Legend: f_1 Purple, f_2 Green, f_3 Blue, f_4 Orange, f_5 Yellow. Notice the yellow is even lower than purple for just five points.....	79
Figure 7.5: Final bird audiograms. The graphs display audiograms aggregated by order. Non Passeriformes (up left), Passeriformes (up right), Strigiformes (down left), and all birds (down right). All species audiograms see in Attachment.....	82
Figure 8.1: Example of cumulative sum z function.....	85
Figure 8.2.: Linear filter distribution with different overlap. A) Overlap is just a half of triangle length a . The overlap ratio $b=0$. B) Overlap is higher than a half. The overlap ratio $b<0$. C) Overlap is lower than a half. The overlap ratio $b>0$. Legend: L_x ... Triangle x left point, R_x ... Triangle x right point.	86
Figure 8.3.: BAF distribution for Passerine. Standard overlap $b = 0$. Number of filters $N=20$	88
Figure 8.4: BAF 1/3 filter distribution for Passerine. Overlap $b = -13$. Number of filters $N=20$	88
Figure 8.5: EER of different bank filter distribution: Linear, Mel, BAF ($b = 0$), and BAF 1/3 ($b = -13$).	90
Figure 8.6: Different bank filter distribution accuracy improvement. For source data, see Table 23...	92
Figure 9.1: The principal idea of the data merging method. A defined number of recordings composes the train data. The figure shows an example for merging level = 3.	94

Figure 9.2: Accuracy improvement: Data sets 1, 2, and 3.....	95
Figure 9.3: Accuracy improvement: Data sets 4, 5, and 6.....	96
Figure 9.4: Accuracy improvement: Data sets 7, 8, and 9.....	96
Figure 9.5: Accuracy improvement: Data sets 10, 11, and 12.....	96
Figure 10.1: <i>BSC infrastructure</i>	99
Figure 10.2: <i>Relational model</i>	101
Figure 10.3: <i>LOV Administration</i>	102
Figure 10.4: <i>LOV Birds</i>	103
Figure 10.5: Mass import, step 3. Green colour signs the correctly imported records. Red colour signs an import error occurs.	104
Figure 10.6: Navigate the map.	104
Figure 11.1: The Mashona mole-rat individual.....	106
Figure 11.2: The Mashona mole-rat colony in the University of South Bohemia, Faculty of Science.	107
Figure 11.3: Spectrograms of the mating calls: cluck (left), and shriek (right).....	108

List of tables

Table 1: Comparison of a human voice and chiffchaff song characteristics.....	16
Table 2: Level of threshold θ value and error rates.....	33
Table 3: Two models of speakers.....	34
Table 4: Speaker verification tool, inputs and outputs.....	54
Table 5: Output file <i>results.txt</i> example.....	54
Table 6: SV tool, configuration of the bird verification.....	55
Table 7: The Chiffchaff recording. The rating of quality is an aggregate value based on the subjective opinion of the operators based on coefficients: noise, masking by other birds, distance, and song clearness.....	59
Table 8: Parameter iteration example.....	66
Table 9: Parametrization set up values.....	66
Table 10: General experiment result for all three experiments.....	67
Table 11: Detailed result. Accuracy for particular bird. The lowest value is 60.3% (round 2, bird D) the highest 95.7% (round 1, bird G).....	67
Table 12: Results summary. Includes a data from all experiments, and reveal the FA and FR errors in detail.....	68
Table 13: Distribution of the experiment results. The results were first rounded and then assigned to a particular level.....	68
Table 14: Number of songs in dependence on the accuracy and the recording's quality.....	69
Table 15: iVectors confusion matrix.....	73
Table 16: Main parameters of 42 audiograms. Legend: BF (Best frequency) is the frequency with the best sensitivity BI (Best Intensity). LF (Low frequency) and HF (High frequency) define the bandwidth of an audiogram. CF (Center frequency) is the frequency in the middle of an audiogram. 30 dB defines the frequency an audiogram reaches 30 dB SPL sensitivity. For graphical legend, see Figure 7.2.....	76
Table 17: Function f_i coefficients for 47 species.....	81
Table 18: Species group aggregate. Final f_i coefficients for four group, based on the order.....	82
Table 19: List of optimized parameters.....	84
Table 20: EER for different bank filter distribution: Linear, Mel, BAF ($b = 0$), and BAF 1/3 ($b = -13$).	89
Table 21: Accuracy comparing. Positive value represents improvement and negative, worsening. The table gives EER differences among particular filter distributions. For source data, see Table 20.....	91
Table 22: Data merging: Experiment results, total EER. First line gives the standard EER without any data merging, labelled as EER_0 . See the EER suffix for particular merging level. For instance, the row	

EER_5 gives the EER for merging level 5. Notice the experiments were not be performed for some merging levels because of insufficient data amount (<i>IDA</i>).	95
Table 23: Data Merging Identification improvement. The table is based on the EER results from previous table. Each line contains differences between particular merging level EER and EER with no merging. For example a third line (<i>Merging level 4</i>) contains differences between EER_4 and EER_0	95
Table 24: Data merging identification improvement, summary.....	97
Table 25: <i>User access levels</i>	99
Table 26: <i>List of application tables</i>	101
Table 27: Parametrization set up values.....	109
Table 28: Mole-rat identification results.	109

Nomenclature

ANN	Artificial Neural Network
ARSBI	Automatic Recognition System of Bird Individual
ARSBS	Automatic Recognition System of Bird Species
AUE	Audiogram Unified Equation
BAF	Bird Adapted Filters
BAF 1/3	Bird Adapted Filters with special overlap
BI	Bird identification
BIV	Bird identification and verification
BV	Bird verification
CI	Call independent
CMD	MS Window Command Line
DFT	Discrete Fourier Transformation
EM	Expectation Maximization
FE	Feature Extraction
FI	Fisher Information
FA	False accept
FR	False reject
GMM	Gaussian Mixture Model
GWF	Greenwood warping function
HMM	Hidden Markov Model
IDFT	Inverse Discrete Fourier Transformation
LLR	Log-Probability Ratio
LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstral Coefficients
MAP	Maximum A-posteriory Probability
MFCC	Mel Frequency Cepstral Coefficients
ML	Maximum Probability
MLA	Marquardt-Levenberg algorithm
MLLR	Maximum Probability Linear Regression
MLP	Multilayer Perceptron
NN	Neural Networks

PLP	Perceptual Linear Prediction
PLPCC	Perceptual Linear Prediction Cepstral Coefficients
PNN	Probabilistic Neural Networks
SV tool	Tool for automatic Speaker Verification, implemented in C++ by colleagues from Faculty of applied science.
SD	Speaker Dependent
SI	Speaker Independent
SOM	Self organizing Map
SR	Speaker Recognition
SVM	Support Vector Machine
SV	Speaker verification
TISR	Text Independent Speaker Recognition
UBM	Universal Background Model
VT	Vocal Tract

1 Introduction

1.1 Overview

An ancient theme of many novels, fantasy as well as children books is to understand animals. Automatic recognition of animal sounds represents a very interesting area with a great potential. There is a great number of species, for which vocalization plays an important role. In addition, their vocal tract anatomy is similar to the human vocal tract. Animals make sounds for many purposes: defending territory, courtship, danger warning, communication, expressing emotions, etc. People do not understand animals' language. Moreover, it is a logical assumption that we will never be able to accurately interpret the meanings of animal sounds. Creating "interpretative" dictionary appears to be an unrealistic task. However, it is possible to retrieve two crucial pieces of information from their vocalization: *Identification of an individual* and *Species recognition*. Although we mainly focus on the individual identification in our thesis, many parts of our research could be also used for species recognition.

Vocalizations are often the most noticeable manifestations of avian species. For many species living in secrecy, or in structurally complex habitats (e. g. forests, bushes, reeds), listening to bird vocalizations is often the easiest, fastest, and cheapest way to detect the presence of a local species; and for this reason, it is a widely used method of species detection in bird censuses and monitoring surveys [BIB00]. The use of acoustic monitoring methods remains limited when information about individual birds (e. g. in studies of survival, site fidelity, ethological studies) is needed and the capture-mark-recapture methods remain the only way of retrieving reliable information about individual birds for ornithologists as well as behaviour and conservation biologists.

However, capture-mark-recapture (ringing, wing tags, collars, colour marks) techniques have also some disadvantages. Some species can be difficult to capture [MAC74], marked birds may avoid recapture [LIN12a], or they may avoid the site where they were captured [LAI07]. Capturing, handling, and marking likely causes stress in animals [WIN82] and may even lead to injuries, which can be a serious issue, especially for endangered species [ARM99]. Therefore, there is a strong need for non-invasive method that would allow recognition of individual birds.

Recognition of individuals within passerine species, which possess complex songs, is a challenging task as they can adjust their repertoire content over time. They may acquire new songs or syllables during their lifetime [NOT86]. They can vary the song content in respect to the audience, whether the receiver is a male or a female [BYE96], or adjust their repertoire to match that of their neighbours each year [PAY96]. Thus, song content can vary within the day, season, or from season to season in passerines [KRO04], [CAT08]. This hinders, or even prevents, the use of call-dependent individual recognition techniques. On the other hand, call-independent methods could be efficient for individual recognition of songbirds as described in this thesis.

Content-independent methods (song- or call-independent) do not compare specific vocalization structures. Instead, they extract parameters related to sound source and filter (vocal chords / syrinx and vocal tract respectively) characteristics common to all vocalizations given by a particular individual. Recently, there were some promising attempts to apply call-independent methods for individual recognition of songbirds [FOX08], [CHE10], [GRA10].

Several animal recognition systems have been proposed. Speaker identification on the closed set of African elephants was introduced by [CLE05] where the Hidden Markov Models (HMM) with Mel-frequency Cepstral Coefficients (MFCC) were used, and the animal-speaker identification reached 82.5%. [TRA05] provided song-type classification and speaker identification of Norwegian Ortolan Bunting. They used HMM with MFCC, delta, and delta-delta parameters. The achieved accuracy varies between 63.6%-92.4% for five song types, and the song-type dependent measurements reached a higher score. The songs and syllables were extracted from the records, and the task was performed on the closed set. [TRI08] used HMM for species recognition, tested on five species of antbirds. Zsebok et. al [ZSE15] deals with a species recognition. First, the recordings are manually sorted (good or poor quality). Good recordings are involved in the experiment only. They extracted songs and calculate signal spectral parameters (time and frequency characteristics). A statistic models are conducted in Matlab Statistics Toolbox. Furthermore, [FOX08] provided call independent classification on the closed set with accuracy 54.3%-75.7% and for call dependent with 69.3-97.1% accuracy, both using Artificial Neural Network (ANN) and MFCC classification. The sounds were cut off and pre-processed. [CHE10] Cheng et al. (2010), introduces individual identification based on Gaussian Mixture Model (GMM) and MFCC across 4 passerine species with an accuracy of 89.1%-92.5%. In their study, the syllables were selected from the records and then sorted; the system used the close set. Bird species classification using Gaussian Mixture Model and a Universal Background Model (GMM-UBM) on the close set was introduced by [GRA11]. The songs were extracted from recordings, and the shortest songs were discarded. Their achieved accuracy varied between 80.8% and 99.8%.

[BUD14] studied an identification of Corncrake (*Crex crex*) individuals based on the pulse-to-pulse duration (PPD) on the close set. Each syllable was measured separately, and each pulse distribution measurement was visually checked. Atypical syllables were removed by authors.

In some studies were also used the continuous as-is recordings (so-called raw, long real-field). For example, [KOG98] experimented with recognition of song elements of birds from continuous recordings. The songs were recorded under the laboratory conditions and visually checked. [POT14] introduces a system using as-is recordings for automatic species recognition based on the HMM with MFCC parametrization. The Hilbert follower was used as a Voice Activity Detector (VAD). Matching recordings were directed to the human observer and final classification. The results varied from 71.2% to 93.3%. [VEN15] proposed a robust frame selection for bird species recognition. Only best frames that represent the dominant sounds are selected and parametrized by MFCCs. These frames were selected applying morphological operators on the raw spectrogram. The results demonstrated an accuracy of 71.5 %.

[BRI12] deals with classification of multiple simultaneous bird species. One of the essential problem of as-is recordings is the audio signal contains bird sounds that overlap in time. Challenging problem is how to separate the singers. Described automation detection of bird species occurrence is based on using a tailored framework, so-called multi-instance multi-label, MIML. The experiments contain 13 species collected with unattended omnidirectional microphones. The aim of [JAN11] was to investigate automatic detection and recognition of bird sounds in noisy environment. The detection was performed by a spectral shape method to identify sinusoidal components.

The primary goal of our thesis is the design, implementation, and evaluation of new methods and algorithms for automatic recognition of birds using live recordings without the necessity of their pre-processing. In the thesis, such automated systems using the suggested methods and algorithms, are going to be called the Automatic Recognition System of Bird Individual (ARSBI), and in some sections, we also deal with Automatic Recognition System of Bird Species (ARSBS). Ornithologists of University of South Bohemia (UoSB), Faculty of Science (FoS) required¹ the ARSBI system accuracy to be at least $\eta \geq 70\%$. System accuracy η is defined as simple ratio of correctly identified birds to total amount of individuals. This value was given based on the discussion of ornithologists on the practical use of ARSBI. The value has no exact base, but it evolves from the practical experience of ornithologists and from the comparison of classical ringing vs. non-contact identification with the help of the automatic system. It cannot be excluded that the required value η will be changed according to the use in particular cases. It can be concluded that the worse conditions (bad climate, high overlapping ratio, etc.) require fewer accuracy requirements, and vice versa.

System ARSBI enables bird identification without the necessity of catching them for ringing or DNA check. The author does not set an unrealistic goal of creating a complete universal tool for identification of individuals of all species. The completion of such a system is, under the current given recognition, unrealistic, mainly because of the recording quality issues (see sections 3.2 and 5.2.3) and non-existing sufficient recording database for training (see chapter 10).

That is why one particular species was chosen in the first part of our thesis: the chiffchaff. For our purposes, the recordings we used were made by our colleagues from University of South Bohemia, Faculty of Science, see section 5.2.2. In the course of our thesis, it was also necessary to handle some minor issues. The goals of our thesis are summarized in the following chapter 2.

1.2 Thesis motivation

At the very beginning of our research, we started to cooperate with ornithologist from Faculty of science in Ceske Budejovice. The main idea originated with ornithologist Pavel Linhart, Ph.D.: to create a tool, which enables a non-contact identification of an individual –ARSBI. Together with the observed colour band combination, the ARSBI can greatly increase the probability of a bird individual identification without the necessity of its capture.

The advantages of the method are:

- Contactless identification, which has significant advantage compared to ringing.
- Increasing the exploitability of bird song recordings. Usually an ornithologist does a spectrogram visual control (examination), or he/she uses a specialized software for basic spectrum analysis as basic parameters computation (bandwidth, energy, start points, end points, bending, etc.). Both processes are automated and computing with advanced parameters should considerably increase information derived from the vocalization.

All tasks described in this thesis follow from the original ARSBI idea and related problems. Although we see that for creating a functional ARSBI, much work still has to be done which we believe is

¹ Just for the record we can bring out that the required accuracy for mole-rats identification was $\eta \geq 65\%$ $\eta \geq 65\%$ see section 11.1. The definition of required accuracies (for chiffchaf and for mole-rat) was set independently.

feasible, with some limitation, of course. In our opinion, a crucial problem is the recording quality, especially overlapping. Altogether, we are still optimistic that an ARSBI will exist in future.

2 Thesis goals

The main goals of our thesis are:

1. Creating ARSBI for chiffchaff individual identification using as-is recordings i.e. live recordings made by ornithologist in nature without any pre-processing. The purpose of this goal is to help ornithologists from University of South Bohemia, Faculty of Science with chiffchaff identification.

ARSBI solution is described in chapter 5.

2. Propose a new feature extraction optimized for a bird song. Because ARSBI is based on the techniques used for human speech recognition, its optimization for bird songs is desirable. Goal solution is described in chapter 8.

At the same time, it is necessary to solve the below tasks that are closely connected to the main goals of the thesis:

3. Utilize bird audiograms in ARSBI. Optimizing feature extraction (goal 2) is connected to the need of working with birds audiograms in programme environment (Matlab, etc.). Audiograms are available only for certain species, and only in graphic version. It was necessary to find mathematical expression of these audiograms.

Solution of the goal is described in chapter 7.

4. To build up a bird song database for scientific bird song data sharing. When working on ARSBI it was desirable to test the system in the best possible way also on other bird species other than chiffchaff. Currently there is not available a universal bird song database that would contain annotated recordings in the sufficient amount and offer a wide range of bird species representatives. That is why a decision was made to create such database.

The solution is described in chapter 4.

5. To test a State of the Art technique for bird individual identification to prove its functionality. We choose an iVector trained by speech. We were also thinking about some techniques to improve an identification accuracy.

Solutions of these goals are described in chapters 6 and 9.

6. When creating ARSBI it was taken into consideration that it can be used for other species not just birds, but also other animals. It was decided to verify the functionality of ARSBI for mole-rats. For their way of life under the ground, vocalization is extremely important for identification at this particular species.

Using ARSBI for identification of mole-rat is described in chapter 11.

3 State of the Art

3.1 Birds

3.1.1 Overview

Ornithology is a field with a long tradition. First known scientific record of ornithology stems from 1773. Until recently, the only true possibility to identify a bird was ringing. First ringing was made by Mr. H. C. Mortensen in Denmark in 1899, in the Czech Republic in 1914 (the Austro-Hungarian empire).

Some birds (lark, nightingale, robin) are able to sing two-part. Some species create the sound in a totally different way, without using the vocal tract. For example wing vibrations (mosquito), using a special membrane (cicada), friction of wings (cricket), rodents hit their head onto burrow wall (Tachyoryctes), etc.

It is impossible or at least very difficult to create something like a lexicon, speech corpus for animals. Humans do not “understand” animals. It is possible at some species, although with difficulties and with obvious objections to the imperfection of such interpretations: incompleteness, ambiguity, when one sound has more meanings, etc. [MOL08] recognizes five different barks according to its “meaning” (joy, warning, sadness...).

Human speech is unique because of the amount of information it carries. Animals with a vocal tract similar to human, we can theoretically assume that with a better equipped brain those species would produce sounds more similar to humans. Opposed to this theory is the fact, that passerine which can imitate human speech (budgerigar, cockatiel, starling, gracula) obviously do not have a more efficient brain than other species. They produce the sounds thanks to a developed musical memory (imprint of a human word). At the same time they do not realize the meaning of produced words.

The observations have shown that the bird songs and its voice change over the time (months, years). Also there are influences of the environment. So far there has not been any research if those changes influence the model of its vocal tract.

3.1.2 Passerine

For the purpose of bird recognition the important order is Passerine, Passeriformes (in Czech language „pěvci“). *Passerines* are divided into two suborders which are dependent on syrinx anatomy, song learning ability, and some others criteria:

- *Suboscines, Tyranni* (lat.), *Křiklaví* (Czech, however this name is not often used).
- *Oscines also called songbirds, Passeri* (lat.), *Zpěvní* (Czech).

Unfortunately, division of passerines into categories and subcategories is still not unified in the Czech language due to the fact that subcategory names are still changing, duplicities exist, etc.

3.1.3 Hearing

From the anatomical point of view, the vocal tract of a passerine is similar to humans. The fundamental difference is that birds have a syrinx, which is equivalent to the human voice box or

larynx. Like the larynx, the syrinx contains special membranes, which vibrate and generate sound waves when air from the lungs is forced through them [CAT08]. It allows a bird to generate two independent audio signals simultaneously. In practice, however, there are only a few „two-tone singers“.

A significant feature of a birdsong is its duration. It is common to hear a bird singing continuously, tens of seconds without interruption. It is considered that this is achieved due to the anatomy of the bird vocal tract mentioned above, where one of the tubes drives the singing while the second performs micro-breathing.

3.1.3.1 Human audiogram

The perception of sound is limited by frequency and intensity. The human frequency range depends on the physical state of hearing of the particular person and his age. Similar dependence at animals was studied for example at [KON70]. In the work of Dooling [DOO02a], it was discovered that the hearing of a bird matures within 2 to 3 weeks. After this time, the hearing properties are identical with adult individuals.

With the constant intensity but changing frequency, the sound is not perceived the same [GRE98]. So, an audiogram over a range of frequencies, perceived with the same intensity, is used for capturing this dependency. Additionally, it was discovered that the resulting curves are different for sounds with a different intensity. For simple tones (sound containing one frequency) the curves were first measured in 1933 by Fletcher and Munson; whereas, nowadays audiograms are defined by the ISO standards over a broad range of frequencies. See Figure 3.1 for an equal-loudness curve based on the ISO.

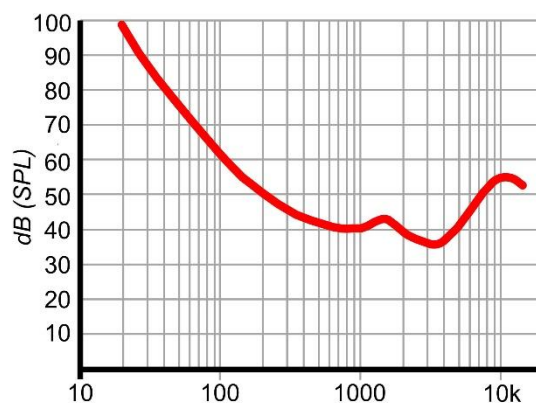


Figure 3.1: ISO equal-loudness curve, 40 dB.

Gaining the human audiogram is a routine procedure that is possible to carry out at any workplace with standard equipment and through simple communication. If the examined subject is reliable (e.g. good quality hearing is the condition for the subject's occupation: pilot, musician), it is possible to obtain substantial data with high resolution within a short time. In case a more accurate audiogram is desired, more correspondents can be tested. Thus, the reliability of gained data increases and the measurement uncertainty decreases.

The standard audiometric procedure requires the respondent to be placed in a muted room. He/she wears headphones into which sounds are randomly played with a constant frequency. The intensity

is gradually increased, and as soon as the respondent hears the sound, he/she pushes a button. The doctor changes the frequency of the testing signal and the measurement is repeated.

For measurements that are more objective or with individuals suffering a hearing disorder, excitation of the cranial bone is used. In this case, an oscillator is placed onto the cranial bone. The vibrations are thus transmitted directly into the middle ear and the eardrum and oscillation are bypassed.

Another method is reading by an electroencephalogram (EEG). This procedure is used for testing hearing disorders or at individuals that are not able to follow the measuring procedure (infants, mentally handicapped).

3.1.3.2 Bird audiogram

Based on the available research (e.g. [HEF98, [MAR04], [DOO02a], [DOO02b], [LAU07], [CAT08]), birds do not hear above 12 kHz. The basic principles of hearing are similar in birds and mammals [MAR04]. Sound causes the oscillation of air molecules, which is then transmitted to the inner ear where the hair cells invoke neuronal discharges. Sound processing in the bird brain is described in [CAT08].

An outdoor acoustic communication depends on many climatic factors. Rain, temperature differences, direction and strength of the wind affect sound speed, propagation, and attenuation. Another effect emerges in air masses with different temperatures. The masses cause sound reflections [MAR04]. The same effect occurs in water utilized by crews in U-Boats to hide the submarine from the destroyers.

Additionally, it is possible to carry on measurement with humans theoretically at an infinite number of locations. With animals, the number of measured frequencies is significantly lower. Birds typically have audiograms with four to eight measured points. Thus, retrieval of audiograms of birds has met several obstacles.

In case of birds, three main approaches exist [GRE98]. The first is a behavioural approach with the precondition that hearing is a behavioural response to sound; thus, researchers use behavioural techniques to training birds to peck the target when they hear the sound [DOO02a]. For instance, Okanoya and Dooling [OKA85] compared hearing abilities of two colonies of Canary Serinus canaries in a Belgium. They trained Canaries to peck one key when they do not hear a tone and second key when they heard it. The same operant technique was used to test high frequency hearing loss primarily above 2000 Hz of domestic Belgian Waterslager canaries (BWC) in comparison with normal hearing non-BWC [LAU07]. On one hand BWC had excellent frequency discrimination ability around 1000 Hz however their frequency discrimination in frequencies about hearing loss region was poor.

Second, the neurophysiological approach is based on measures the neuron electrical impulses in response to sound [KON70]. Konishi [KON70] determined hearing thresholds by playing sounds to anesthetized birds and then recording directly from auditory neurons in the cochlear nuclei. According to his results based on experiments with six songbirds, vocal frequencies seldom falls down under 1 kHz but all species hearing sensitivity was well below 1 kHz. As well as the differences between the lowest and highest thresholds tend to be similar among different species (i. e. about 40-50 DB).

Third, a recent method is the auditory brainstem response (ABR) which is recorded by using subdermal needle electrodes. For example, adult Budgerigars (*Melopsittacus undulatus*) were

sedated with an intramuscular injection of ketamine and diazepam prior to electrode placement [6]. Birds stayed motionless for up to 75 minutes and electrodes recorded reactions to sound stimuli from a speaker – clicks and tones. After the experiment, the birds were placed in to therapy unit where they recovered from sedation.

An audiologist typically performs test administration and interpretation. Although the ABR provides information regarding auditory function and hearing sensitivity, it is not a substitute for a formal hearing evaluation, and results should be used in conjunction with behavioural audiometry whenever possible.

3.1.3.3 Bird vs human audiogram

There are two basic differences between bird and human audiograms:

1. Small number of measured frequencies
2. Small number of measured individuals

The fact that an audiogram is made up of a limited number of points leads to inaccuracies of the collected data. A line for better evaluations of the collected data usually connects the data points. Yet, in reality, the course between such points can be different from a smooth one. See the possible indication of such distortion in Figure 3.2.

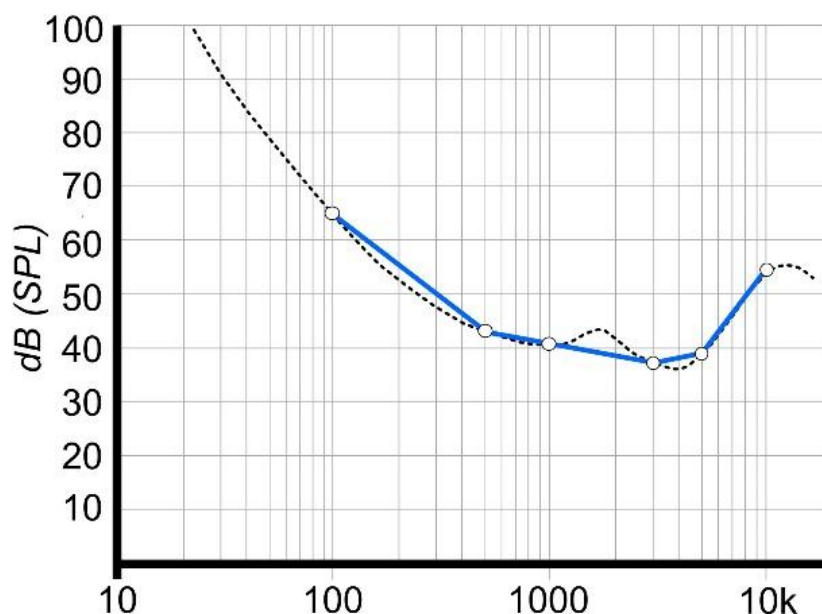


Figure 3.2 : An audiogram distortion example if one measures just a low number of frequencies. The black dotted line represents the ISO human equal-loudness curve for 40 dB. The blue solid line demonstrates how a human audiogram may look like if only six frequencies are measured:

100 Hz, 500 Hz, 1 kHz, 3 kHz, and 10 kHz.

Bird audiograms are based on a small number of *respondents* (i.e. tested birds) in contrary to human audiograms based on tens or even hundreds of interviewees. Evidently, the statistical error is high for small number of birds.

One has to consider if there is also possibility that the examined bird has a hearing different from an average individual due to mechanical damage, insufficiently developed hearing, or malformation.

Also, the error rate setting is difficult because there are no available studies dealing with hearing disorders of birds or differences of hearing sensitivity.

3.1.3.4 Bird and human hearing

Human and bird ears are variously sensitive to different frequencies. Figure 3.3 shows the dependence of both human and bird hearing on frequency.

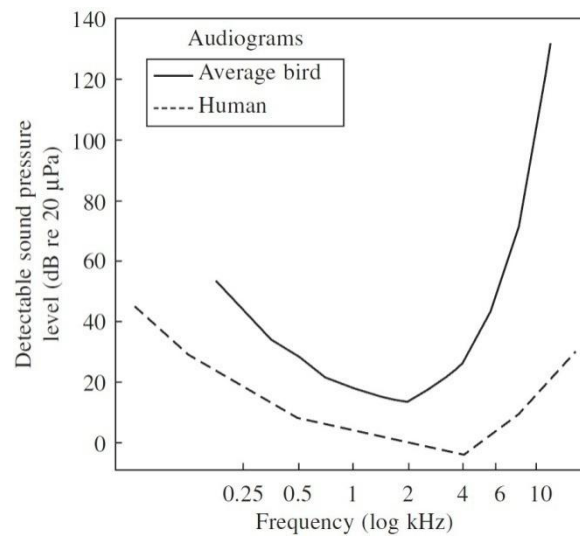


Figure 3.3: Average audibility curves human and “average bird” [CAT08].

For humans, this dependence is described by the Fletcher-Munson curves. The curve relates to the frequency band of birdsong, their communication running between 0.5 kHz and 6 kHz in average.

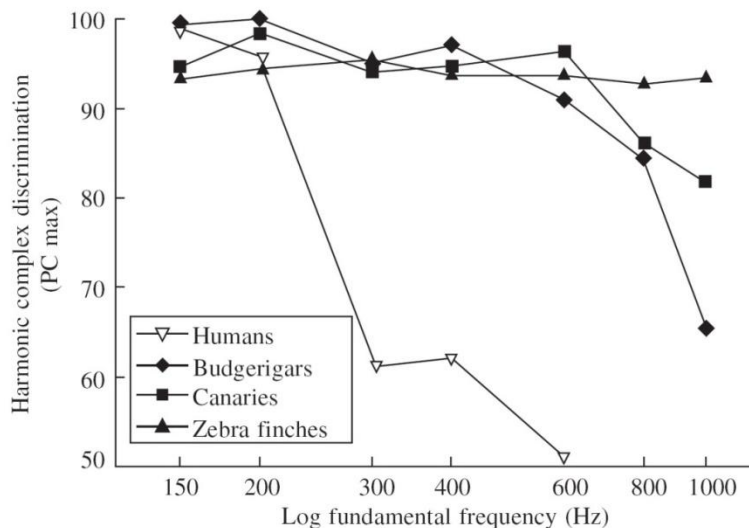


Figure 3.4: Harmonic complex discrimination [MAR04].

All birds are able to discriminate between harmonic complexes with much higher fundamental frequencies (800 to 1000 Hz) than humans; this requires temporal analysis over fundamental periods as short as 1 ms in duration. Humans are unable to discriminate between complexes with fundamental frequencies higher than about 250 Hz, i.e. about 4 ms in duration. Enhanced time processing of

complex sounds by birds, relative to humans, may be a general characteristic of the avian auditory system [MAR04].

Studies of bird hearing reveal that they do not hear well over the range of frequencies that embraces most of those used in their songs [DOO02a]. Within a narrower range of frequencies, where they hear best, the ability to discriminate between two sounds approaches the level of acuity often reported for humans. However, there is also a major difference. Birds excel in discriminating between two complex sounds, which differ only in the temporal fine structure [MAR04].

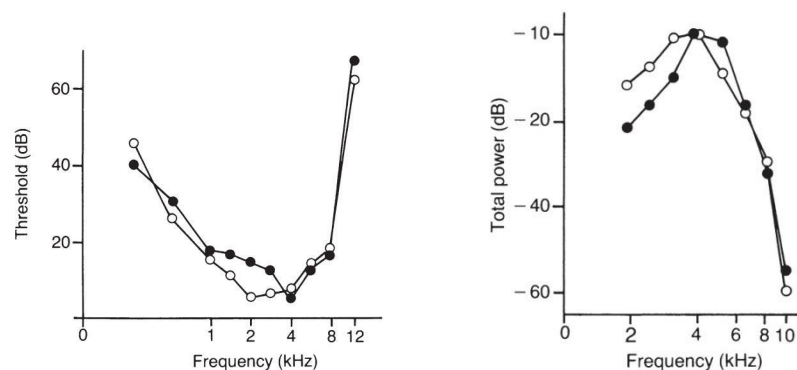


Figure 3.5: Audibility curves (a) of song sparrows (open circles) and swamp sparrows (filled circles) compared to the power spectra (b) of their songs [CAT08].

3.1.4 Vocalization

Throughout the animal kingdom, vocalization is used for various purposes. For example, it can be used to inform about sex, condition and age of the signaller [REN04], [BOU13]. It could also be useful to recognize neighbours, kin or even a particular individual [BEE85], [BAL90], [REN96] or to distinguish reproductive or dominance status [YOS09], [HOE10]. Furthermore, vocalization is useful for synchronizing members of a group [BOI95] or for warning others against danger [SIL94], [MAN02].

3.1.4.1 Bird song

Song is the natural vocalization of the passerines. Basically we recognize two main purposes of bird song: to allure female and to mark (border, defend) territorially defined districts. Male birds are usually singing because their vocalization is augmented by the male sex hormone testosterone. There are more than 500 bird species in Europe, about 400 in the Czech Republic. Ornithologists have found singing females in about 100 species. Females normally have low levels of circulating testosterone, but if these are increased then females will also often produce song [CAT08].

Vocals produced by a bird are generally divided into two categories:

- Call short signals, mostly meaning warning.
- Song songs composed of syllables, mostly territory and allure meaning.

Birds learn to sing when they are born, and as they grow, singing is greatly influenced by neighbouring bird singers. Since the 19th century, “contest canaries” have been trained by engaging a young canary near a so called “precentor”, a senior canary with a high quality song ability

(complicated songs, using many syllables, having a broad tonal range). The singing of the “learning” young canary rapidly improves thanks to the quality of the preceptor.

A bird’s repertoire highly depends on various influences such as bird mood, environmental conditions (normal, stress, and danger), day time, season, life phase (breeding season, building a nest, young bird care), temperature, and weather conditions. Moreover, a bird’s song differs from place to place because birds learn their whole life. Birds modify their song if they hear new syllables, new collocation, or a new song variation. In summary, every bird sings differently.

Chiffchaff sing during spring, when courtship dances begin. These birds sing mostly at dawn. The explanation is that a female is still slumberous and the male can get closer easily and sound propagation is easiest thanks to favourable climate and low turbulences.

3.1.4.2 Vocal tract

For humans, the frequency associated with vocal tract dimensions is fundamental. Thus, the fundamental frequency it is lower for adults and highest for children. In animals, much greater variability of the vocal box can be found. Figure 3.6 shows the dependence on animal body mass and emphasized frequencies of vocalization. With a suitably chosen scale: a line with a slope of -1 added to the graph, describing this dependency. Small animals use high frequencies while larger animals lower frequency.

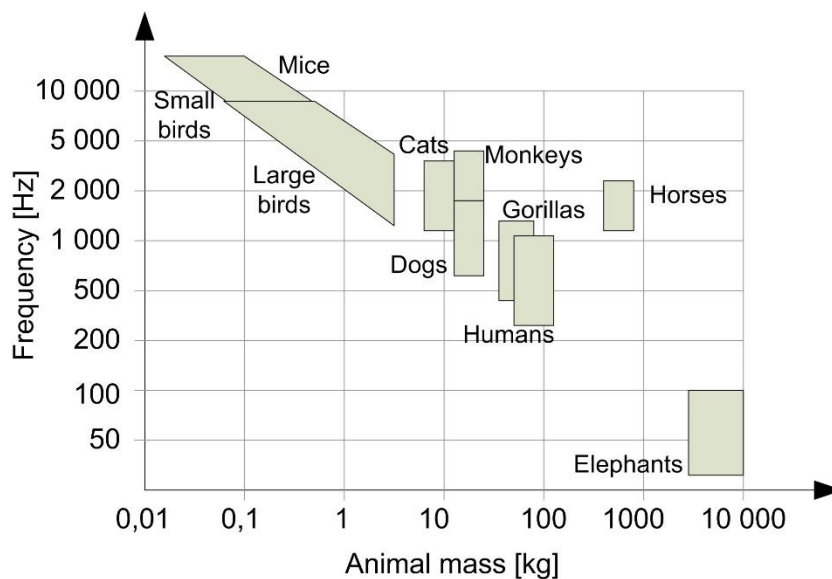


Figure 3.6: Animal body mass and frequencies of vocalization [CAT08].

The dependence for both humans and animals is related to the basic relationship between wavelength and frequency

$$\lambda = \frac{c}{f} \quad (1)$$

where c is the speed of sound. The relation approximates dry air

$$c = (331.57 + 0.607t) [\text{ms}^{-1}]. \quad (2)$$

The syrinx is the principal organ of birdsong creation. Figure 3.7 shows divided structure with two sound generators. A syrinx of double-voiced singers is described in [KRA09].

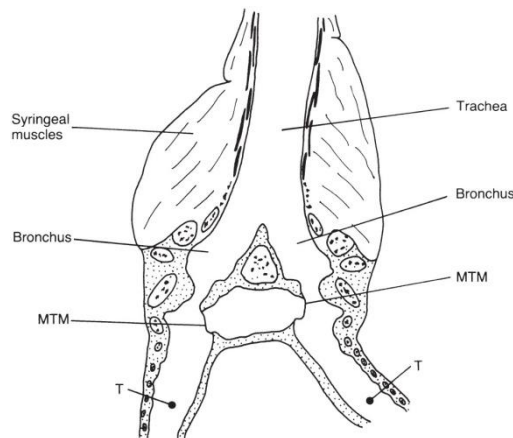


Figure 3.7: Cross-section of the syrinx of a brown thrasher.
(T) thermistors, (MTM) medial tympaniform membranes.

Unlike humans, animals are usually equipped with less noise harmonics. Some animals may produce purely sinusoidal (singers) or pure noise character (small rodents). Some birds produce the two-voiced sound, for instance [KRA09] deals with double whistle of *Centrocercus urophasianus*. Just as the human vocal tract, the principal of sound generating in birds will be approximated by convolution both generating signal $x(n)$ and impulse response $h(n)$ of the vocal tract:

$$s(n) = x(n) * h(n) \quad (3)$$

where $s(n)$ is song (speech signal), $x(n)$ is an excitation (signal source) and $h(n)$ is impulse response of vocal tract (vocal tract filter), see section 3.5.1.

In general, the vocal tract of many species of animals is similar to humans: monkeys, some singers, cetaceans. Some birds may also sing in two-tone (lark, nightingale, thrush). Some types of sound production are completely different and operate without the use of vocal tract. For example, the oscillation of the wings (mosquito), using a special membrane (cicada), rubbing the wings together (cricket), in rodents banging his head against the wall hole (Lesser Bamboo Rat).

3.1.4.3 Song hierarchy

A spectrogram of chiffchaff song is shown in Figure 3.8. The bird song is divided into four levels: *Song*, *Phrase*, *Syllable*, and *Element*. The basic bird song stands between calls and songs. The calls are short squawks emitted by birds as an emergency or warning sound. The song consists of Phrases and Syllables. The syllable is then divided into the so-called Elements; see Figure 3.8 and Figure 3.9

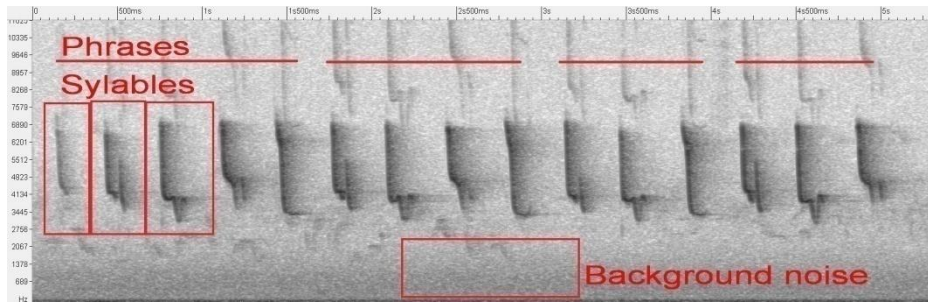


Figure 3.8: One song of *Chiffchaff*.

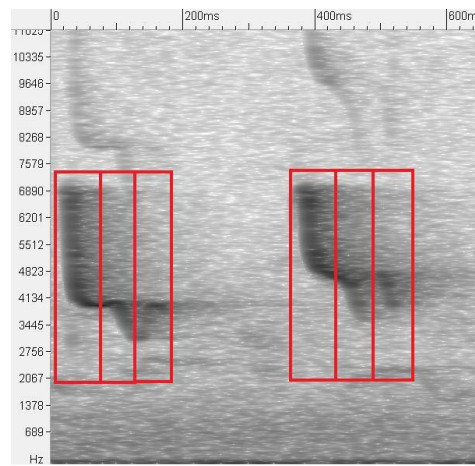


Figure 3.9: Two syllables divided into elements.

3.1.4.4 Song analysis

Ornithologists have been trying to analyse bird songs since the nineteenth century. Some tried to create a transcript of a song together with musicians, for an example see Figure 3.10. If necessary, special marks/notations can be used to describe bird song characteristics in more detail.

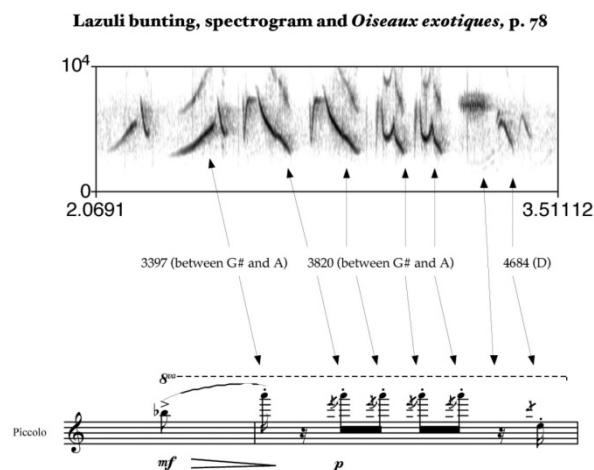


Figure 3.10: Score of a Lazuli Bunting song. Created by composer Olivier Messiaen (Fr., 1908-1992).

Since the last decade, ornithologists have used spectrogram as well as specialized software to analyse singing e.g. Avisoft-SASLab Pro (*Bioacoustics*, Germany) or Raven Pro (*Cornell Lab of Ornithology*, USA). So far, the achievement is that we are able to build up a song structure, similar

to a human word structure. An example of a song structure of *Bengali Finch* is reproduced in Figure 3.11.

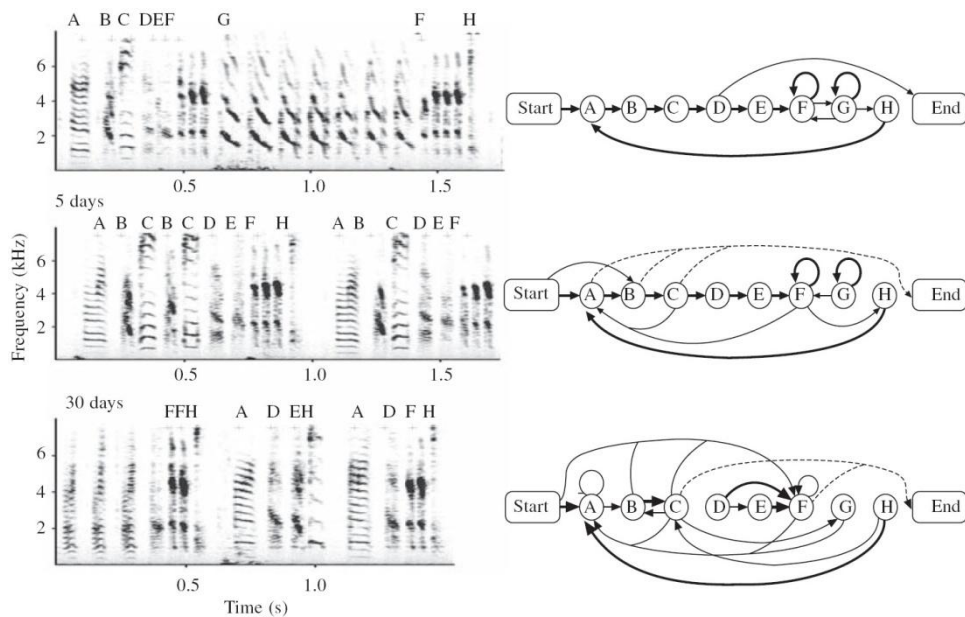


Figure 3.11: Bengali Finch song structure [MAR04].

3.1.4.5 Differences

Human voice recognition systems have been a concentration of study for many years. It is important to discover the differences between human and bird voices. List of differences taken into account are listed below:

- Limited repertoire of sound

The number of bird sounds is calculated from a few to tens of songs.

- Spectral and time-base characteristics

Approximate frequency range of human voice is 300 Hz – 3.5 kHz. The level of acoustics pressure is about 50 dB. Voice dynamic of the Czech language is about 30 dB for untrained voice. Animals produce simple sounds (squawk or croaks) or continuous sounds with some structure (bird song, whale song). The structure of a song is affected by many influences. Frequency bandwidth varies from tens of Hz (elephants) to tens of kHz (passerines). Chiffchaff produces songs between 3 and 7.5 kHz with dynamics about 20 dB. Just as a matter of interest the acoustic pressure of whale's song is reported at about 160 dB.

- Harmonics structure

Human voice contains many harmonics, lying in close frequency band.

Table 1 summarizes discovered differences between speech and bird song. Observed birds produce sound just with a few harmonics: two or three. Some animals even make pure sinusoidal or pure noise sounds (passerine, rodent). See following figures for the chiffchaff (*Phylloscopus collybita*) and tree pipit (*Anthus trivialis*) song. Notice the chiffchaff recordings were made by team of

Dr.Linhart (University of South Bohemia) and the tree pipit recordings by team of Dr.Tereza Petruskova (Charles University).

	Human Voice	Chiffchaff	Tree pipit
Frequency range	100 Hz - 3 kHz	2 kHz - 7kHz	2 kHz - 8kHz
Bandwidth	3 kHz	5 kHz	6 kHz
Average duration of one word/syllable	1 s	100 ms to 200 ms	40 ms to 200 ms
Number of Harmonics	2 - 20	1 - 3	1 - 3
Duration the signal can be considered as stationary	10 - 30 ms	5 - 20 ms	5 - 20 ms

Table 1: Comparison of a human voice and chiffchaff song characteristics.

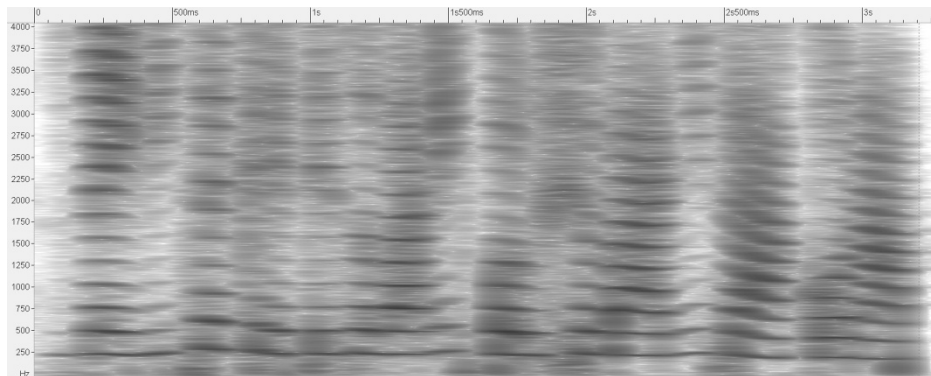


Figure 3.12: Human voice. Studio record, women, Czech language, business news, impassive. Frequency range 0 – 4 kHz.

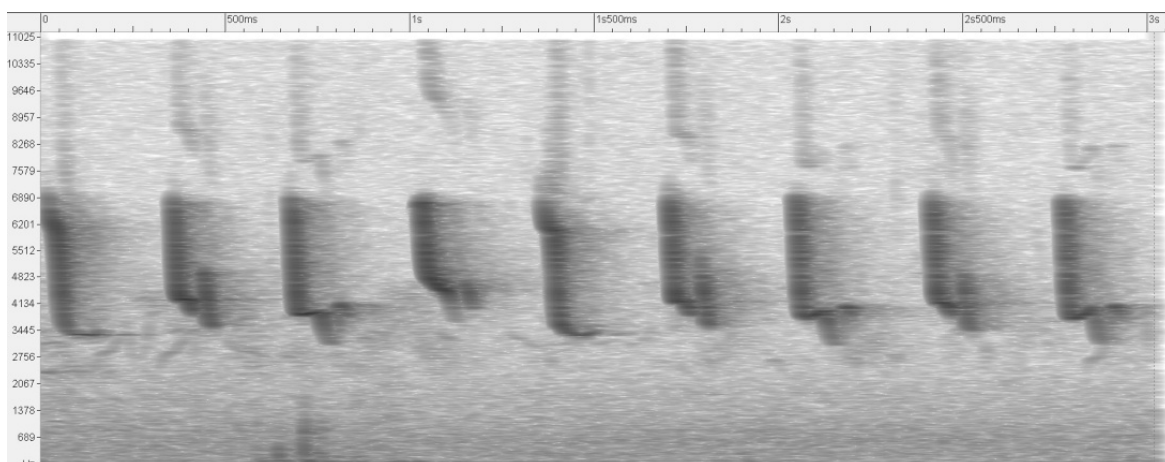


Figure 3.13 Chiffchaff song, recorded in wood, morning. Band width 0 – 11,025 kHz.

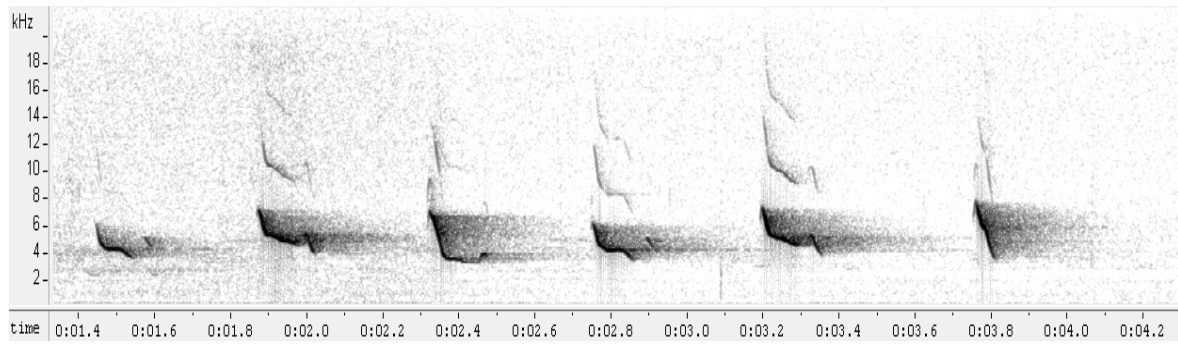


Figure 3.14 Chiffchaff song, recorded in suburban areas, morning. Bandwidth 0 – 22,050 kHz

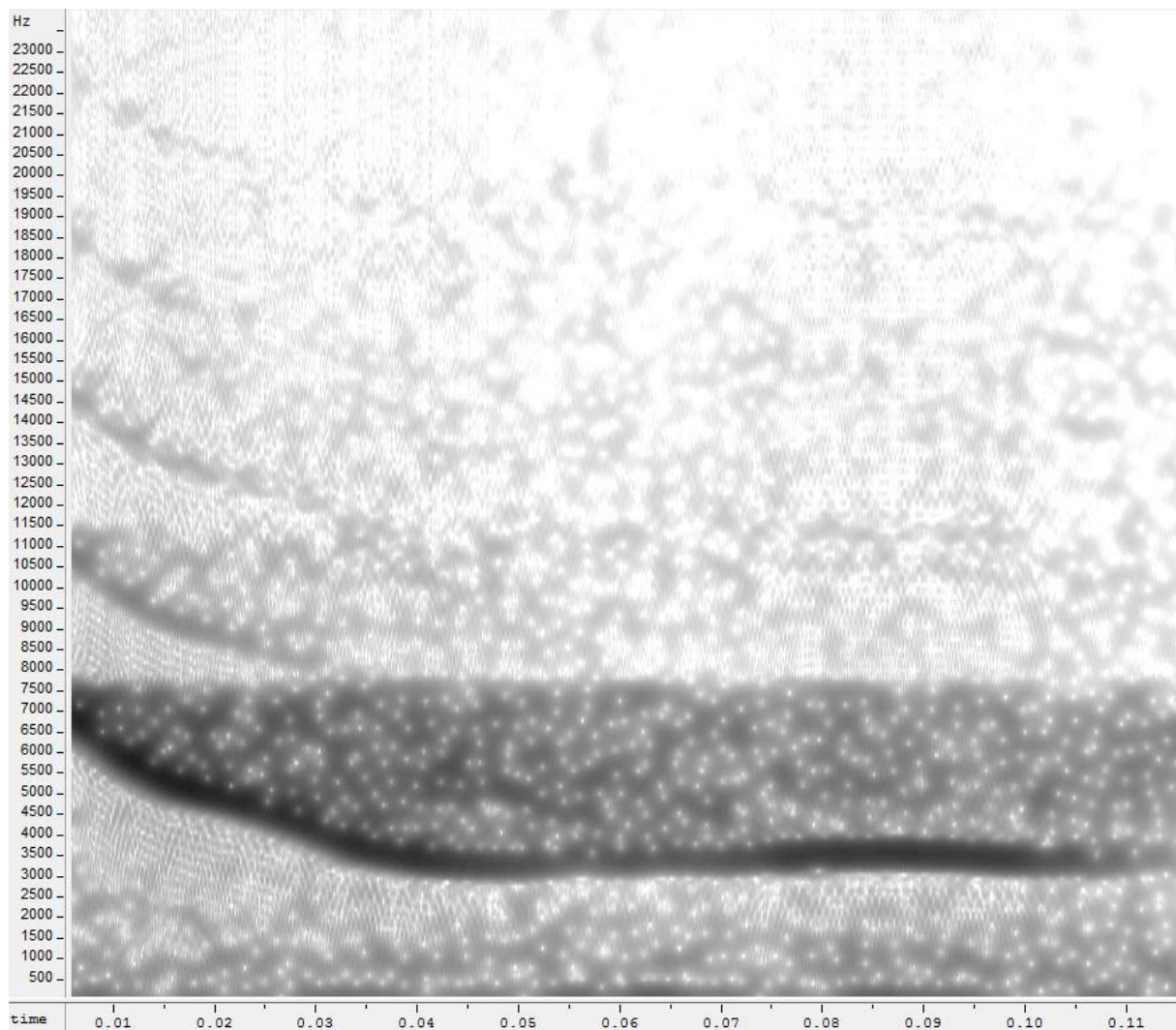


Figure 3.15: Chiffchaff song detail, one syllable. Recorded in the woods, sample frequency 44.1 kHz.

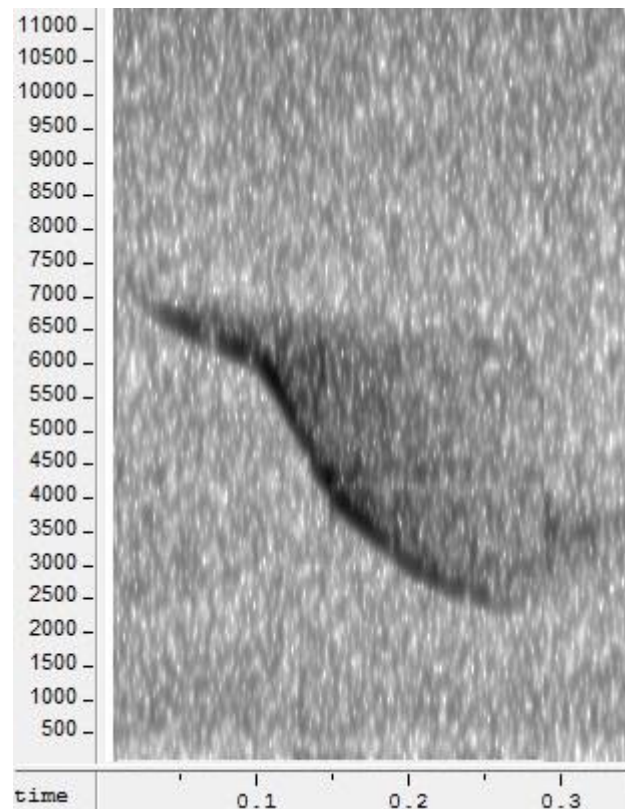


Figure 3.16: Tree pipit song detail, one syllable. Recorded in the woods, sample frequency 22.05 kHz.

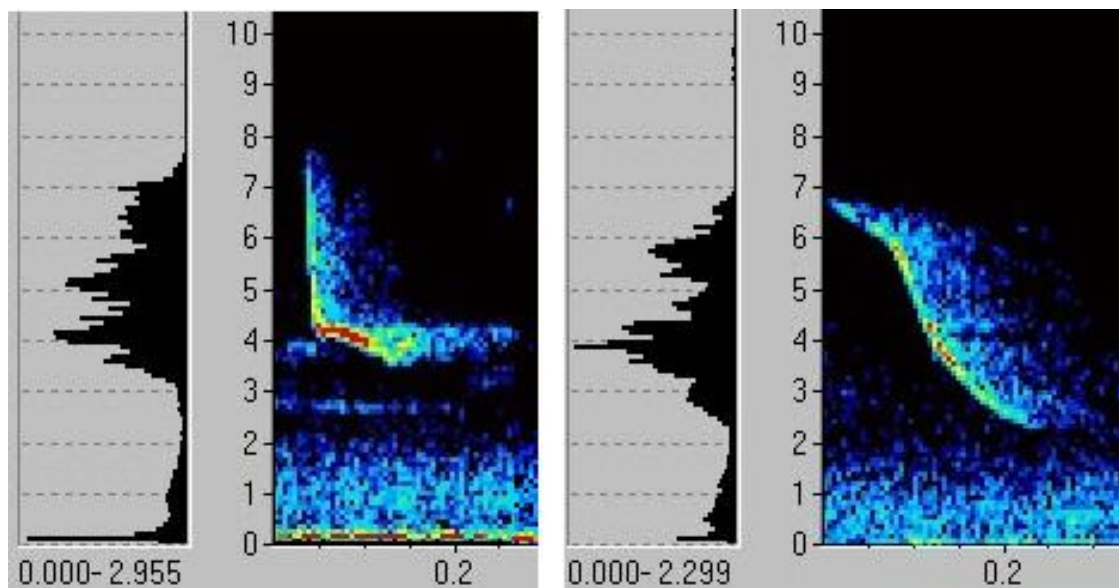


Figure 3.17: One syllable spectrogram: Chiffchaff (left), and tree pipit (right).
The vertical bar graphs illustrate a band energy.

3.1.5 Bird song propagation

Unlike humans, birds communicate over long distances and signal transmission takes a longer time. Effects like phase shift, reflections, interferences and so on, play important role in comparison to human speech. Moreover, the direct sound wave is attenuated by its propagation through the air and by hindrances (trees, walls, houses, rocks, hills, leaves, etc.). The reflected sound waves may play important role.

In summary an origin sound arrives to the bird-listener attenuated, distorted, delayed (echoes and reverberations), and phase shifted.

3.1.6 Ringing

One of the major challenges in ornithology is the task to differentiate bird individuals from each other. One of the currently used methods is bird ringing (banding). This procedure has some negative aspects:

- It is necessary to capture the bird.
- The bird is ringed for life.

Firstly, the capture is a very stressful event. What is more, if the ornithologist does not wear gloves, the bird is exposed to human contact. The bird can be kept in a net for several hours, until the zoologist arrives. It happens especially when night birds are caught. Secondly, a bird receives a ring on the body, which changes its appearance, increases its weight and sometimes hinders its movement. Furthermore, the ring may not only bother the bird itself, but there is a question whether its colour and appearance does not distract the partners or other individuals from its community.

The author of this work cooperates with ornithologists of University of South Bohemia, Faculty of Science, who have observed that Chiffchaffs which were caught, do not return to the same place so often as the other ones. The estimation of the return of the ringed birds is about 15%. Moreover, it is practically impossible to repeatedly catch a bird caught once before. The ornithologists conduct banding of warblers with one or up to three bands of different colours. They are placed on both legs to use as many combinations possible. The bands carry a unique code (1 letter and 5 digits) that enable a unique identification. In this case chipping is not possible because of the warbler's petite body structure.

It is obvious from the above mentioned that when monitoring the bird it is possible to identify individuals thanks to the bands colour combination, but only by using binoculars. This identification is not faultless, because the colour can be identified incorrectly from the distance or there might be birds from another habitat with the same colour combination from another ornithologist, etc. Such accurate determination can thus be carried out only after the bird is caught in a net and the band code is read.

3.2 Data recording

Based on the localization, bird song recordings can be crudely divided into indoor and outdoor ones. Indoor recordings take place in a laboratory or in special environment where a bird lives (zoo, aviary, botanical gardens). Outdoor recordings differ in recording time which depends on the day cycle of

the target bird. The most suitable season for recording is spring, most birds are singing at dawn, contrary to owls, for instance, whose activity is strongest at or after sunset or even at night.

When recording in the exterior it is not possible to get a recording that contains only the required songs. Birds are most active at dawn. That causes a worse quality of recordings, because they contain other songs as well, from the same, and also from different species. Mutual masking of songs is significant mainly at recordings made in the forest. In case the recording takes place near a conurbation, there are also noises of the city. The city “wakes up”, cars, trains, public transport are very noisy. Since the recording distance is up to several tens of meters, the surround noise is unavoidable.

The fundamental disadvantage of morning recordings is the clarity of the record. A significant advantage is the light and thus an easier localization of the individual. Birds that sing at day are less timid than night birds. An ornithologist can try and approach to the distance of several tens or even just to several meters. Alternatively, it can be possible to catch the bird for a accurate identification.

Birds singing during the day are not as shy as night singing ones so that ornithologist can get much closer (tens of meters). The main advantage of morning recording is sunlight and the possibility to visually identify the recorded individual(s). The main disadvantage of morning recording is:

- In the forest, many birds are singing at the same time.
- Near the town, there are urban noises.

Day recordings are similar to morning ones. They only differ by the level of noise (town traffic, animals). In addition, song activity is lower for most species during the day as compared to the morning.

One of the most important problems of outside recording is the record quality. The bird is usually far from the microphone, many birds are singing at the same time, the level of surrounding noise is usually high and unwanted sounds propagate into the records. Furthermore, if the bird moves the song frequency changes by the Doppler Effect. This implies that the quality of outside records is worse than the laboratory ones because of worse

- clearness of the singing,
- signal to noise ratio,
- ability to distinguish among the song of the target and other birds,
- unwanted sounds.

Despite these disadvantages, we prefer outside recording because the bird is recorded in a natural environment, which is crucial.

At present, many articles focus on off-line recognition system (records are stored in PC and processed). However, any on-line system would be very useful for ornithologist when performing fieldwork. We believe that an on-line system would have good applicability, but so far, no research using such system has been to our knowledge published.

While working at night, darkness is an obvious disadvantage. It disables the identification of an individual as well as its accurate localization for microphone setting. That is why stationary recorders are generally used. Another disadvantage is a smaller cadence of night birds songs and longer pauses between them. Significant parameters are also shyness and vast territories where they occur. On the other hand, the clarity of the recording tends to be better than of the morning data collection. Most

inhabitants are asleep so the level of surrounding sound is significantly lower. The recording contains only sporadic sounds of nocturnal animals and permanent sounds of the background such as rustle of the trees, insects, etc. The recordings also contain a noticeable sound of the recording device. When there are worse climatic conditions, the quality of the recording decreases rapidly.

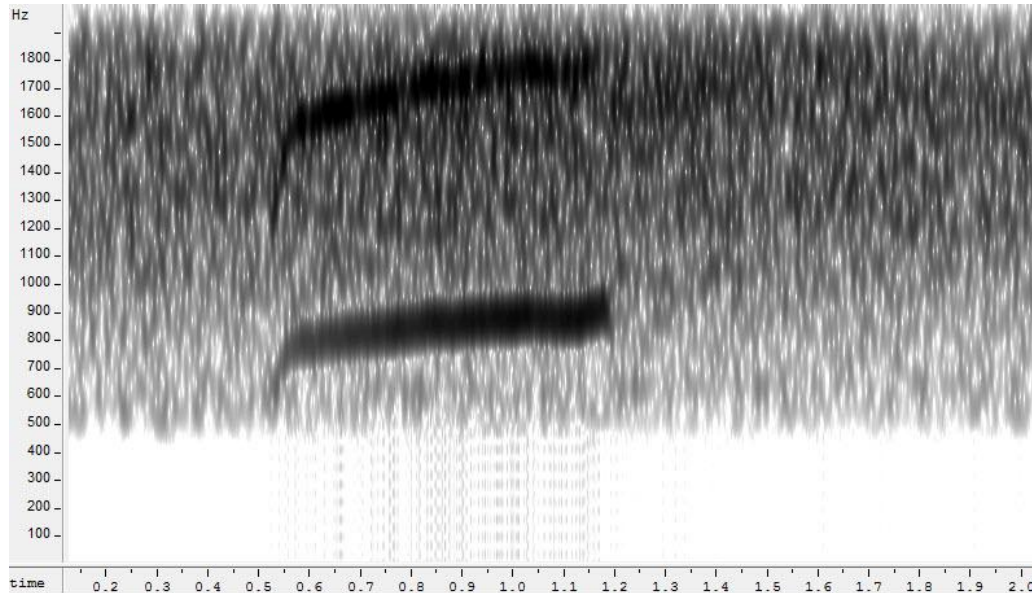


Figure 3.18: A voice of a Forest Owlet (*Athene Blewitti*). A wide spectrum noise from 500 Hz is visible, caused by a background noise and a recording machine.

Records acquired in a laboratory are exceptionally clear and clean. These can sometimes even be obtained in an anechoic room. However, just some species can be recorded by this method, usually domestic passerines (canary, budgerigar, zebra finch, starling, and parrot).

3.2.1 Masking

Consistent experiments were performed to discover masking on birds. The experiments investigated how noise interferes with song hearing. This phenomenon is difficult to study in the field [DEN98], but laboratory studies, where both signals and noise can be controlled, can provide guidelines for the effect of noise on hearing in the real world. Laboratory studies with pure tones and white noise show how intense a pure tone must be relative to the background noise in order to be heard [DOO95]

It was discovered that in the frequency region of best hearing for most birds, around 3 kHz, tone levels must be on average about 25 dB above the spectrum level of noise to be detected. Some principles regarding the masking of signals by noise were established. First, energy in the frequency region of the signal is the most effective in masking. Noise at other frequencies has much less effect. Second, the signal-to-noise ratio needed for detection stays relatively constant over a wide range of noise levels. Third, if the signal and the noise come from different directions, much less masking occurs. Finally, it is one thing to detect a signal such as a vocalization and quite another to discriminate between one vocalization and another, or to recognize a particular vocalization [MAR04].

Unfortunately, no rigorous measurement of bird masking is available as there is in human voice measurement. Only the general principles are known, which describe how birds communicate in noisy environments in nature.

3.2.2 Process automation and microphone arrays

Automatic sound recordings in the nature represent new and very interesting area for bird song research. [POT14] deals with these recordings for species recognition. Perceptual Linear Predictive cepstral coefficients (PLP CC) are extracted, as well as MFCCs. The aim of [JAN11] was to investigate automatic detection and recognition of bird sounds in noisy environment. The detection was performed by a spectral shape method to identify sinusoidal components. Ehnes and Foote [EHN15] used automated recorders. Then they visually checked spectrograms to sort the recordings into categories high/low quality. The decision was based on the spectrogram darkness (contrast between song and background). The spectrogram cross-correlation was used to calculate similarity of songs. They used software Raven Pro (*Cornell Lab of Ornithology, USA*).

Microphone arrays represent relative new approach for bird species and individual recognition. These tools enable precious source localization, which is impossible to achieve with one or two microphones. An essential overview of acoustic monitoring using microphone arrays is given in [BLU11]. Unfortunately, just a few researcher crews could use these tools because of its astronomical prices. Remote recording allows continuous area monitoring but the data evaluation is very difficult. [ULL16] deals with such type of recording using microphone array of 24 microphones. A detection system is based on spectrogram cross-correlation. Kwan et. al. [KWA06] deals with microphone array real-time monitoring system. The system automatically recognizes a large number of bird species but its accuracy highly depends on distance between bird and microphone.

3.2.3 Data processing

To work with a bird song means to work with a long recording. An ornithologist walks at the suitable position where can record for as long time as possible. The recording takes up to tens of minutes even if a stationary recorder or a tripod is utilized the duration can take hours. Theoretically, three methods are possible based on which type of recording is processed:

1. **Continuous record.** Whole recording is used with no cuts, noise cancellation etc., so-call raw recording, see Figure 3.19. An automatic VAD is required to distinguish between song and non-song segments [PTA15a].
2. **Single songs.** Single songs are cut out from the recordings, see Figure 3.20. One uses an automatic software to cut out the songs from the raw recording, e.g. Avisoft or Raven. Although accurate software setup, some mistakes occur during the cut-off process: lost songs, non-song parts classified as song, etc.
3. **Combination of both.** Single songs and continuous records are used together. Two different experiments run independently, merging the results.

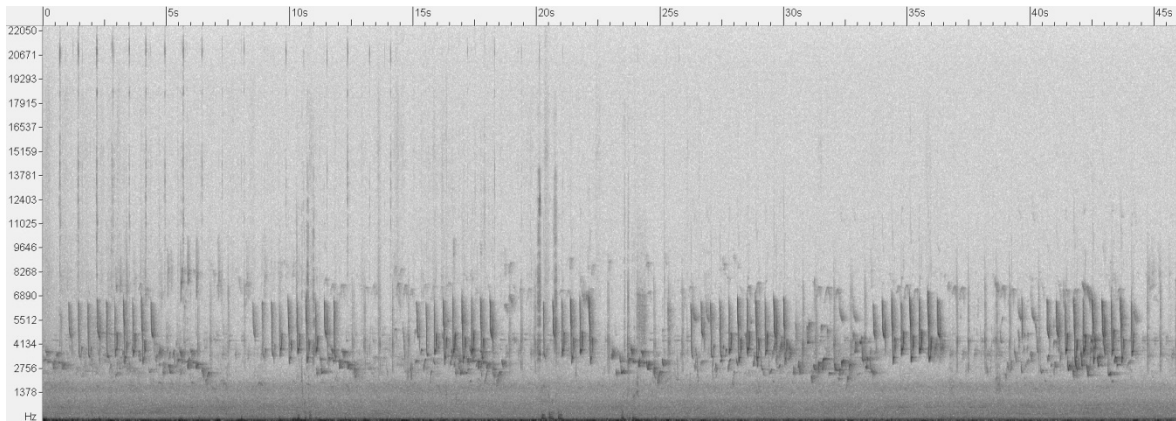


Figure 3.19: Continuous *Chiffchaff* record (raw record), length 45 s.

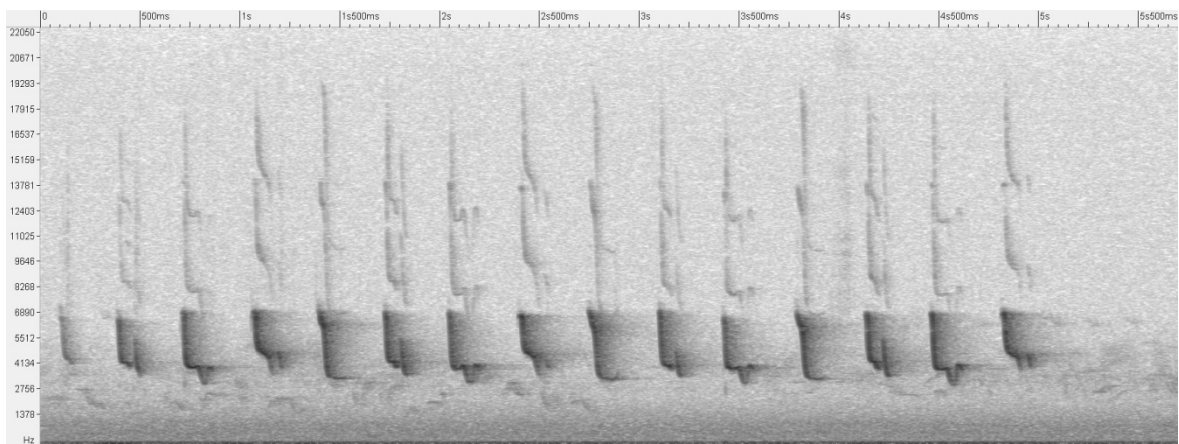


Figure 3.20: Single *Chiffchaff* song, cut off from the raw recording, length 5.5 s.

3.3 Speaker recognition

Speaker Recognition (SR) can generally be divided into two tasks: Speaker Identification (SI) and Speaker Verification (SV).

3.3.1 Speaker identification

The task for SI is to assign a given speech record to a specific speaker from a database of speakers. The aim is to identify the speaker. A typical example of using SI is authentication of a person entering a building, or on the phone (e.g. mobile banking, ticket reservation). It is important to define whether the set of speakers to be identified is closed or open. For an open-set a new speaker can appear at any time. If closed-set is considered the speaker set is finite.

First a *speaker model* of n -th speaker is defined as λ_n , and the set of all speaker models as Λ (see section 3.4.7, where we discuss the models in detail). Then for the closed-set case, models are created for all persons involved. The goal is then to identify a person by using speaker models selected from the finite set

$$\Lambda = \{\lambda_1, \dots, \lambda_L\} \quad (4)$$

In the open-case set an unknown person may appear in addition to the known ones. Then the set of models is extended by the model-set of unknown persons

$$\Lambda = \{\lambda_1, \dots, \lambda_L\} \cup \Lambda_{\text{UNKNOWN}} \quad (5)$$

However, the model of an unknown person cannot be defined as there is no data available. Instead, a speaker with the highest score is chosen. If some predefined threshold is not reached, it is concluded that no person of the given set has spoken.

3.3.2 Speaker verification

The aim of SV is to *confirm* or *deny* whether the speech record belongs to a particular speaker (to confirm the identity). The system has to infer an identity the speaker claims. An example of SV task is the authentication of a user logging into a system. There are some applications where the speech is the only biometric parameter useable, in a phone conversation for instance. Verification depends on what the speaker says. It can be text-independent (speaker says any word or phrase) or text-dependent (speaker pronounces a pre-specified word or phrase, such as a digit or a code word.)

Since we do not understand bird language and we cannot order the bird to sing (except in cases of trained singers). Thus, the most suitable approach for the automatic recognition system of a bird is the *Speaker recognition text-independent task*.

3.3.3 Speaker recognition methods in ornithology

Kuntoro et al. experimented with both song-type classification and individual identity clustering [KUN10]. The HMM was used for song-type classification with achieved accuracy of the song-type between 50% and 98.8%. The error rate of individual identification was from 2.9% to 50%, which

was evaluated by the author as unusable. They used recording from year 2000 as a training data, while recordings from 2001 was used for evaluation.

Clemins dealt with classification of animal vocalization using MFCC and PLP parameters and HMM classifier implemented in HTK [CLE05]. The first part of the research dealt with a call-type identification, the second with an individual identification. It was recommended to use the Greenwood warping function (GWF). Achieved results for call type recognition were between 51% and 90%. The results were highly dependent on the type of used parameters and on the classifier. Tested species were frogs, elephants, and beluga whales. For all species, a particular GWF was computed.

Fox described a call independent identification in birds [FOX08]. The records were divided, some parts were used for training and some for an identification. The length of the parts varied and the average length was about 10 s. MFCC was used, while the classifier used an ANN MLP implemented in the NN toolbox in Matlab. The network had one hidden layer with 16 neurons. Reached identification accuracies were for willie wagtails 72.9%, for canaries 97.1%; 54.3%; 98.6%, and for singing honeyeaters 75.7%; 96.5%. Accuracy dependence on noise was also tested.

Selin focused on bird sound classification using wavelets [SEL05]. An ANN was used for automated classification of acoustic signals. MLP and self-organizing map (SOM) were used as classifiers. Eight bird species were tested with accuracy 96% and 93.8% for MLP and SOM respectively.

[KOG98] compared both methods of Dynamic Time Warping (DTW) and HMMs for automated recognition of bird song elements. The experiment uses sound of zebra finches and indigo bunting (passerine). The article studied both the DTW and a HMMs methods, and summarizes pros and cons applied for bird song.

Some authors deal with the animals sounds in order to identify (interpret) their meaning. Molnar et al collected more than 6,000 barks in an attempt to recognize the meaning of dog barking [MOL08]. Five kinds of barking were distinguished, named their meaning as: stranger, fight, alone, ball, play. Classification efficiency rises between 43% and 52%.

[GRA10] dealt with optimization of feature extraction module to improve bird species recognition. Improvement was found after optimizing a bandwidth and a number of filter banks. Experiment used bird sounds from commercial Audio CD by Cornell Lab of Ornithology.

[CHU09] proposed a Correlation-Maximization Filter to suppress background noise. They used 2246 songs from five bird-ant species. The both GMM (256 Gaussians) and HMM (6 states, 256 Gaussians) were used. The feature vectors were extracted by the MFCC with dimension 39, and by a de-noise filter output based on the Wiener/Correlation-Maximization. The lower classification error rates were 4.1 for HMM and 4.7% for GMM.

We did not find any Czech researchers dealt with an automatic bird recognition.

3.4 Recognition system overview

This chapter deals with a speaker recognition system. We modified and applied it in order to solve the problem of automatic bird recognition. The research is mainly focused on the GMM-UBM method. Figure 3.21 displays an outline of the GMM-UBM recognition system, for iVectors

recognition system see section 3.4.11. Notice the shadowed boxes are discussed in detail in following sections.

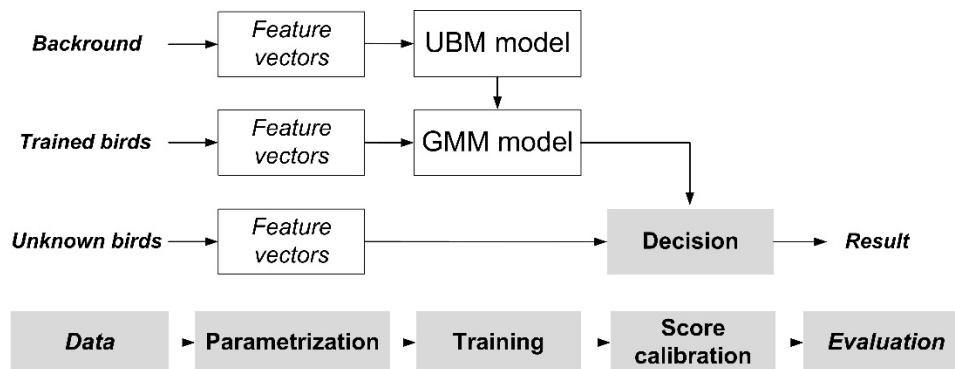


Figure 3.21: General outline of the GMM-UBM recognition system.

The process flow is decomposed into:

- Parametrization: recordings are parametrized to extract features, forming a set of feature vectors.
- Training: UBM model and GMM model estimation.
- Decision: probability comparison of unknown bird and trained models.
- Score calibration: choice of a verification threshold.
- Evaluation: based on EER, DET or any other method.

Following sections give a brief overview of these steps. See sections 3.4.7 and 3.4.7. for more details.

3.4.1 Parametrization

Data are parametrized to extract features, forming a set of feature vectors. In order to extract multiple feature vectors a rectangular sliding window of length l_w given in samples is utilized. The samples

in the window are processed, subsequently the window is shifted to a next position usually by half of its length, and the extraction of feature vectors is repeated, see Figure 3.22.

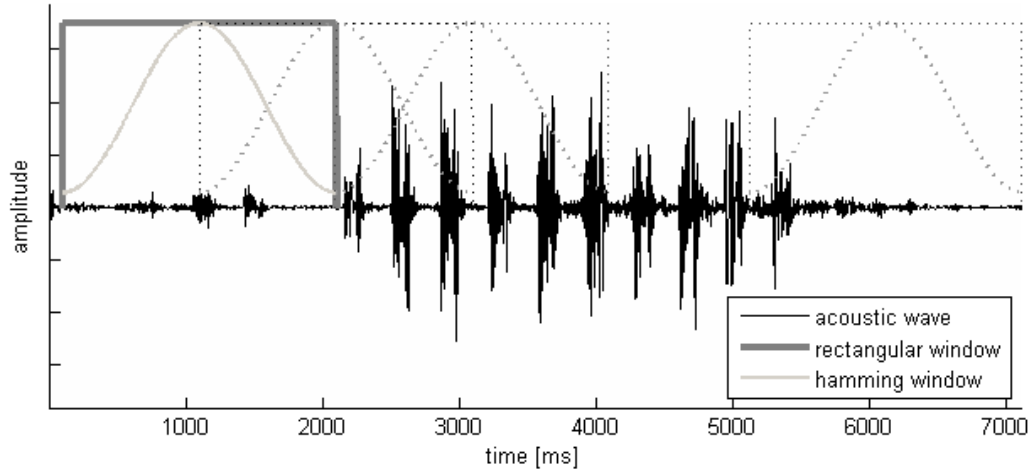


Figure 3.22: Samples in the rectangular window are weighted by the Hamming window, FFT is performed, filtration utilizing triangular filters is carried out, and a cepstral feature vector is extracted. Subsequently, the window is shifted to its new location and the extraction process is repeated.

Samples in the sliding window are at first weighted by a Hamming window to suppress undesirable effects when the Fast Fourier Transform (FFT) is applied right after the windowing. Next, the power spectrum is computed in order to extract frequency characteristics of the signal present in the window, and since it is symmetric only the first half is kept – the interval $[0, f_s/2]$, where f_s is the sampling frequency. To smooth the spectrum a set of triangular shaped Filter Banks (FBs) – bandpass frequency filters – of height one is spread across the frequency domain. Number of filter banks N_{FB} is set empirically by an expert and determines the "smoothness" of the power spectrum – the lower is N_{FB} the smoother is the spectrum, i.e. irregularities in the spectrum are suppressed in a higher extent. Triangular FBs are symmetric and they are defined by the location of their midpoints $m_i, i = 1, \dots, N_{FB}$. The i^{th} FB starts from the midpoint m_{i-1} of the previous FB, reaches its maximum value at m_i , ends at m_{i+1} , and is zero otherwise, see Figure 3.23. Hence the midpoints are located at frequencies

$$m_i = \frac{f_{\max}}{N_{FB} + 1} i, i = 1, \dots, N_{FB} \quad (6)$$

where f_{\max} is the frequency where the last FB ends. After the filtering, cepstral approach is carried out. More precisely, logarithm of the output of FBs denoted as $P_i, i = 1, \dots, N_{FB}$ is computed and a cosine transform is performed in order to decorrelate the features so that the *cepstral coefficients*

$$c_k = \sum_{i=1}^{N_{FB}} P_i \cos\left(\frac{\pi}{N_{FB}} \left(i - \frac{1}{2}\right) k\right), k = 1, \dots, K \quad (7)$$

are extracted, where $K \leq N_{\text{FB}}$ is the number of cepstral coefficients, thus the dimension of extracted feature vector $\mathbf{c} = [c_1, \dots, c_K]^T$ is K .

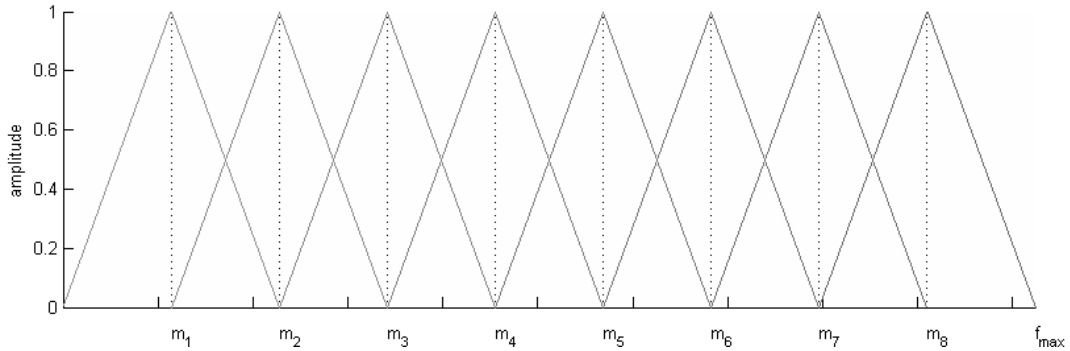


Figure 3.23: Triangular filter banks spread linearly in Hz scale.

Finally, in order to incorporate also some dynamic information on the variation of the signal in time, often numerical approximations of the first derivative of cepstral coefficients are evaluated and added at the end of the cepstral vector [BIM04]. They are called delta coefficients and can be computed as

$$\Delta_t = \frac{\sum_{j=1}^J (\mathbf{c}_{t+j} - \mathbf{c}_{t-j}) j}{2 \sum_{j=1}^J j^2}, \quad (8)$$

where J is the number of neighbours used to compute the numerical approximation of the time derivative, and t denotes the time index of extraction of the feature vector \mathbf{c}_t .

The extracted feature vectors are called Linear Frequency Cepstral Coefficients (LFCCs).

3.4.2 Gaussian Mixture Model (GMM)

Once the feature vectors in the form of LFCCs were extracted, the next step consists in modelling of the probability distribution of the data. We will now focus on GMMs firstly introduced to the speaker recognition by [REY95] and widely used up to now [CAM06], [KEN07], [DEH10]. GMMs are generative statistical models, well suited for description of static (context-independent) data sources, where the time progress of samples is of no interest. The basic assumption is that feature vectors are i.i.d. (independent and identically distributed).

For a D dimensional feature vector \mathbf{x} , the GMM takes the form

$$p(\mathbf{x} | \lambda) = \sum_{m=1}^M \omega_m \mathbf{N}(\mathbf{x}; \mu_m, \mathbf{C}_m), \quad (9)$$

where $\omega_m, \mu_m, \mathbf{C}_m$ denote the m^{th} mixture's component weight, mean and covariance, respectively, $\lambda = \{\omega_m, \mu_m, \mathbf{C}_m\}_{m=1}^M$ is the set of GMM parameters, M is the number of mixture components, and $\mathbf{N}(\mathbf{x}; \mu_m, \mathbf{C}_m)$ is the probability density function of the normal distribution with mean μ_m and covariance matrix \mathbf{C}_m . In order to have a valid probability distribution restrictions of the form

$$\forall m: 0 \leq \omega_m \leq 1 \text{ and } \sum_{m=1}^M \omega_m = 1, \quad (10)$$

have to be laid on GMM weights. An example of a GMM is depicted in Figure 3.24.

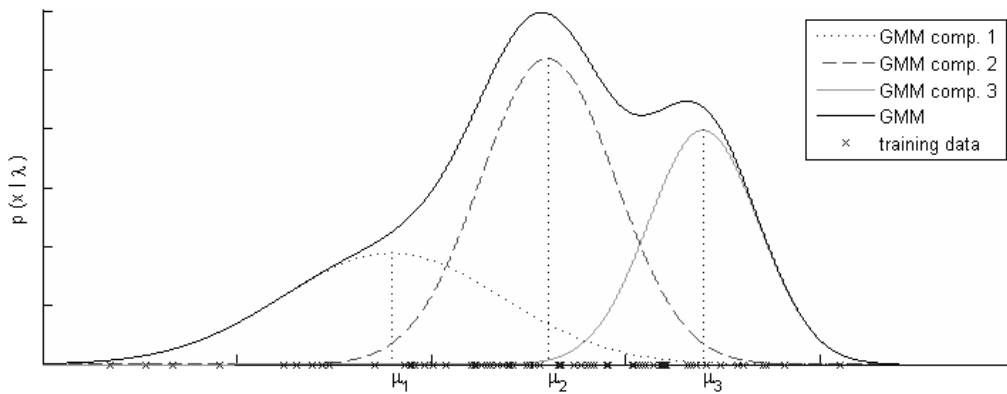


Figure 3.24: Given a set of one dimensional feature vectors (x -axis), the Gaussian Mixture Model with three mixture components which best describes the data set (in the sense of maximal likelihood (4)) is given by the solid line. Note that the GMM is formed from 3 normal distributions each weighted by the relative number of vectors it encloses.

Generally, the covariance matrix \mathbf{C}_m is considered full, nevertheless in most cases diagonal matrices are assumed, especially because of numerical stability reasons and computational costs.

3.4.3 Training

In order to train a GMM an iterative method called Expectation-Maximization (EM) can be exploited [DEM77]. It is based on the Maximum Likelihood (ML) approach and maximizes the probability $p(\mathbf{x}_1, \dots, \mathbf{x}_T | \lambda) = \prod_{t=1}^T p(\mathbf{x}_t | \lambda)$ of submitted training data $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ given the model parameters λ .

Since EM algorithm is iterative it has to be initialized by some suitably chosen parameter λ^0 , which is then update in each iteration until convergence is reached. Also a number of mixture components M has to be set, among others it depends on the dimension of feature vectors D and on the number of input vectors T . Thus it would be convenient to change the number of mixture components M

in the model of each speaker (or bird) in dependence on the value of T . Note that the number of parameters to be trained is equal to $M \times (D + D + 1)$, hence the number of mixture components M multiplied by the number of parameters of the mean vector, number of non-zero elements in the covariance matrix (assuming a diagonal covariance matrix) and one additional parameter for the weight of the mixture component. To estimate the mean of each mixture component we would need at least M feature vectors, and to get a reliable estimate of the covariance matrix the required number of feature vectors would substantially increase. The lack of training data can lead to ill-conditioned models. However, imagine we would have a prior knowledge about the model parameters. This could be used to properly initialize the estimation algorithm (instead of a random initialization often used), moreover it could help suppress the ill-conditioning when only a low amount of training data is available. For this purpose Maximum A-Posteriori (MAP) adaptation of an Universal Background Model (UBM) is utilized [REY00].

UBM is a GMM trained using the EM algorithm on a huge amount of (background) data collected from a lot of speakers. Hence, it reflects actual operating conditions (e.g. channel, noise) presented in the background dataset. Since the amount of data is huge, the number of mixture components M can be set high. Rather than utilizing the EM algorithm to train the model λ_s of a speaker s from scratch, the MAP adaptation of the UBM is used, where

$$\lambda_s = \tau \lambda_{\text{UBM}} + (1 - \tau) \lambda_{\text{ML}}, \quad \tau = \frac{r}{r + T}, \quad (11)$$

λ_{ML} is the maximum likelihood estimate of model parameters given only the data of speaker s (computed utilizing the EM algorithm initialized by λ_{UBM}), T is the amount of data available, and r is a relevance factor set by the user. Thus, if enough training data are available (r goes to zero), the model based only on the s^{th} speaker's data will be preferred, otherwise it will lean toward the universal background model λ_{UBM} .

3.4.4 Decision

Given a set of T_s feature vectors $\mathbf{X}_s = \{\mathbf{x}_1, \dots, \mathbf{x}_{T_s}\}$ and a model λ_q , the score (log-likelihood or their similarity measure) is given as

$$\mathbf{L}(\mathbf{X}_s, \lambda_q) = \sum_{t=1}^{T_s} \log p(\mathbf{x}_t | \lambda_q) \quad (12)$$

where $p(\mathbf{x}_t | \lambda_q)$ is the probability of a feature vector \mathbf{x}_t given in (4), the logarithm is used to ensure numerical robustness. In the closed set identification scenario (see the discussion at the beginning of Section 4) the identity of s would be assigned according to

$$q^* = \arg \max_{q=1, \dots, Q} \mathbf{L}(\mathbf{X}_s, \lambda_q), \quad (13)$$

where Q is the number of speakers in the reference set. Thus, the score (7) is evaluated for each model of each reference speaker ζ , and the identity of the speaker ζ is determined according to the maximal value of the score.

In the case of verification, it is not enough to find the closest speaker given by the maximal value of the score (7), in addition we have to *verify* the identity of the closest speaker. Thus, we have to choose between two hypotheses:

- H_0 : \mathbf{X}_s was spoken/sung by speaker/bird ζ ,
- H_1 : \mathbf{X}_s was not spoken/sung by speaker/bird ζ .

For this purpose the Log-Likelihood Ratio (LLR)

$$\log \frac{p(\mathbf{X}_s | H_0)}{p(\mathbf{X}_s | H_1)} = \log \frac{p(\mathbf{X}_s | \lambda_q)}{p(\mathbf{X}_s | \lambda_{UBM})} = L(\mathbf{X}_s, \lambda_q) - L(\mathbf{X}_s, \lambda_{UBM}) \quad (14)$$

is evaluated [BIM04] and a threshold θ has to be set and compared with LLR. If $LLR > \theta$ then H_0 is accepted, otherwise H_0 is rejected and H_1 is accepted. The threshold is in most cases set by the user and its value reflects the penalization for making errors. Note that hypothesis H_1 was connected with $L(\mathbf{X}_s | \lambda_{UBM})$ – likelihood that \mathbf{X}_s belongs to the background population of speakers or to the environment exposures, which should be also part of the background data set. Therefore it is of importance to strictly distinguish between training, testing and background data. Loosely speaking, the background population must not include any of the training or testing data, otherwise the validity of (9) would be violated.

3.4.5 Score calibration

UBM plays a crucial role in the open set identification, since it defines the operating conditions and simplifies the choice of a verification threshold. The problem related to the choice of a proper verification threshold is often referred to as the score calibration [REY97]. In order to understand the main idea of the score calibration assume that we are given 2 models of 2 different birds and 1 test recording from each of the 2 birds. Next, let v_{12} be the score of the test recording from the 1st bird given the model of the 2nd bird, analogically we can compute v_{11} , v_{21} , v_{22} . Obviously, having a good recognition system yields $v_{11} > v_{12}$ and $v_{22} > v_{21}$. In the closed set identification scenario we would assign the input recording to the bird, which model gave the best score given the input recording (i.e. if $v_{11} > v_{12}$ then recording 1 would be assigned to the 1st bird represented by model 1). However, in a verification scenario we have to compare the score to a unique threshold θ in order to get the final decision (it is not clear whether the most similar bird really is the bird in question). Hence, if $v_{11} > \theta$ then the same identity of the bird represented by the 1st test recording and the bird represented by the 1st model is confirmed, otherwise their identities are assumed to differ. The problem is if the values of v_{11} , v_{12} are significantly higher/lower than v_{22} , v_{21} . E.g. if $v_{11} > v_{12} > v_{22} > v_{21}$ then the value of the threshold θ cannot be set so that simultaneously $v_{11} > \theta$

and $v_{22} > \theta$ while at the same time $v_{21} \leq \theta$ and $v_{12} \leq \theta$. In order to solve the problem an additive constant has to be subtracted from v_{11} , v_{12} and/or added to v_{22} , v_{21} . This is done by the UBM when evaluating LLR given in (10) and it is one of the possibilities how to calibrate the score, for a more detailed description of score calibration techniques see [STU05], [YIN08].

3.4.6 Evaluation

Four different situations may occur during the verification, see Figure 3.25.

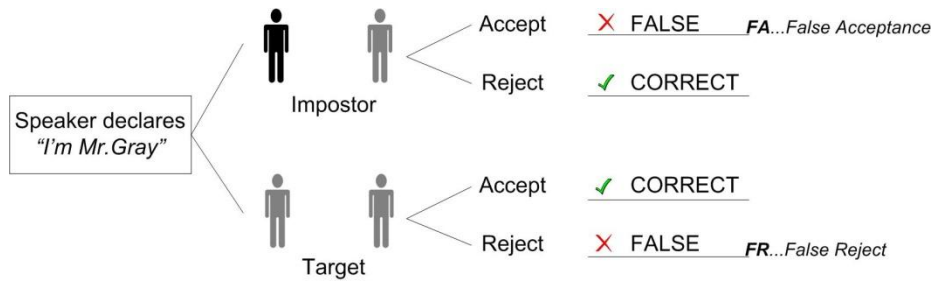


Figure 3.25: Verification: False and correct decision.

Incorrect acceptance error $R_{FA}(\Theta)$ is defined as

$$R_{FA}(\Theta) = \frac{n_{FA}(\Theta)}{n_{IM}}, \quad (15)$$

where Θ is the threshold (see below), n_{FA} is the number of cases when the system incorrectly accepts the impostor, and n_{IM} is the total number of cases where an impostor has been tested.

Incorrect rejection error $R_{FR}(\Theta)$ is defined as

$$R_{FR}(\Theta) = \frac{n_{FR}(\Theta)}{n_{TRGT}}, \quad (16)$$

where n_{FR} is the number of cases when the system incorrectly rejected the Target (right speaker/bird) and n_{TRGT} is the total number of cases where the target has been tested.

Setting the threshold Θ affects the total number of R_{FA} and R_{FR} . Increasing the threshold reduces the false acceptance error rate FA , but it simultaneously increases the false rejection FR error. This happens because the system requires a higher probability of similarity. On the contrary, if the threshold is lower, the FR error decreases, but the FA increases as the system needs lower probability of similarity to accept the speaker. This leverage effect is summarized in Table 2. Both errors are called *operating point* [PSU06].

Highest Θ	$R_{FA}(\Theta)$	decrease
	$R_{FR}(\Theta)$	increase
Lowest Θ	$R_{FA}(\Theta)$	increase
	$R_{FR}(\Theta)$	decrease

Table 2: Level of threshold Θ value and error rates.

The *Equal Error Rate (EER)* is used for single number evaluation of the system, which indicates the threshold value Θ_{EER} at which R_{FA} and R_{FR} are equal. It is defined as

$$R_{EER} = R_{FR}(\Theta_{EER}) = R_{FA}(\Theta_{EER}) \quad (17)$$

In real experiments, however, a threshold Θ must be set first, where after the decisions the R_{FA} and R_{FR} errors can be calculated. Finding the threshold Θ_{EER} can therefore be nontrivial.

To express the system success rate with just one number, the curve DET (*Detection Error Trade-off Curve*) is used. Error rates are plotted as a function of the threshold [BIM04] at the DET. The advantage of the DET curve is a good readability especially for low differences between the errors. Another advantage is a good distinction in case that we plot many curves at the same time. It is particularly useful when the system parameters are fine-tuned.

3.4.7 Probability model methods

The probability model method is based on computing $p(\mathbf{X}|\lambda_i), i = 1, 2, \dots, I$ for each of I speakers, where $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ represents the feature vector of unknown individual composed by N parameters \mathbf{x}_n , and λ_i represents the model of i -th speaker. The method uses speaker models λ_i comparison instead of feature vectors. It radically decreases a dimension of compared data, and make the identification process feasible. We define the unknown individual feature vector \mathbf{X} belongs to the j -th speaker just if the probability $p(\mathbf{X}|\lambda_i)$ gives the highest value for $i = j$.

The *Hidden Markov Models (HMM)* is usually used for *text-dependent* tasks where a content of speech is under research, and a prior knowledge of the speech is unknown [BIM04]. The *Gaussian mixture model (GMM)* is the common choice for *text-independent* tasks [BIM04] where the content of the speech is unimportant and the matter of problem is to identify a speaker.

3.4.8 GMM-UBM Speaker verification system

The definition of the speaker verification task is *to determine if an utterance Y was spoken by speaker S* . If we suppose that Y contains speech of only one speaker then we call the task a single-speaker verification. If not, the task becomes multi-speaker detection. With regard to the objectives of this work the task is defined as automatic bird identification: *to determine if song Y was sung by a bird individual S* .

If we define two possible conclusions of a single-speaker verification

$$\begin{aligned} H_0 \dots Y \text{ was spoken/sung by } S \\ H_1 \dots Y \text{ was not spoken/sung by } S \end{aligned}$$

then the goal of the task is to determine both probabilities

$$p(Y|H_0), \quad (18)$$

$$p(Y|H_1). \quad (19)$$

These are called probability of hypothesis H_0 and H_1 respectively. If $p(Y|H_0)/p(Y|H_1) \geq \theta$ then H_0 is accepted (Y was spoken/sung by S), else if $p(Y|H_0)/p(Y|H_1) < \theta$ then H_0 is rejected (Y was not spoken/sung by S). The θ represents a treshold, see section XXX.

For techniques to compute values for the two probabilities, see [REY00]. Basic stages of the speaker verification system are shown in Figure 3.26.

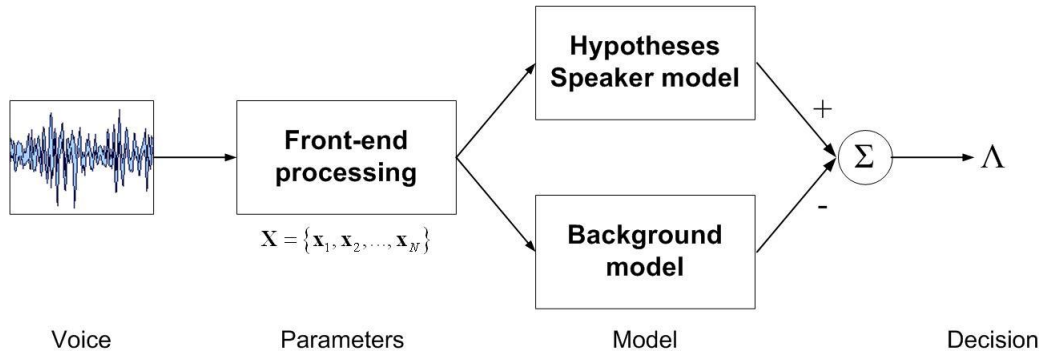


Figure 3.26: Probability ratio-based speaker detection system [REY00].

The input stage should contain a speech (song) of length t . The front-end processing stage extracts feature vectors, which contain speaker-dependent information

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}. \quad (20)$$

where N is the number of feature vectors. It depends on the length of a speech/song and the length of a frame, which the speech is divided into segments. Computing the probabilities H_0 and H_1 follows, based on the feature vectors.

Probability of H_0 is by a GMM with parameters model λ of the speaker. These probabilities represent the hypotheses

Hypothesis	Model	Denotation
H_0	λ_{hyp}	<i>Hypothesized speaker S is in the feature space of \mathbf{X}</i>
H_1	$\lambda_{\overline{hyp}}$	<i>Alternative to H_0</i>

Table 3: Two models of speakers.

Probability ratio statistics of both models can be expressed as

$$\frac{p(\mathbf{x}|\lambda_{hyp})}{p(\mathbf{x}|\lambda_{\overline{hyp}})}. \quad (21)$$

The logarithm of this quantity gives the log-probability ratio

$$\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\lambda_{hyp}) - \log p(\mathbf{X}|\lambda_{\overline{hyp}}). \quad (22)$$

The model λ_{hyp} of the H_0 hypothesis will be well-established using the training data of S . On the contrary, model $\lambda_{\overline{hyp}}$ must include all of the possible alternative hypotheses covered by H_1 . To estimate model $\lambda_{\overline{hyp}}$ two approaches come into consideration:

- Use models of all others speakers in order to cover all alternative hypotheses. The approach is known as *background speakers (BS)*.
- Merging speeches from several speakers to train a single model. The approach is known as the *universal background model (UBM)*.

GMM-UBM system for Speaker verification task is described in [REY00], [REY95], [DEH09]. The used system is referred to as the *Gaussian Mixture Model-Universal Background Model* speaker verification system (GMM-UBM) [REY95].

3.4.9 Expectation-maximization EM

Estimation of model parameters (i.e. $\lambda = \{w_y, \mu_y, \Sigma_y\}$ for GMM) is commonly based on Maximum Likelihood (ML) criteria. The ML supposes we get a likelihood model $p(\mathbf{x}|\lambda)$ with unknown parameter λ ; then we need to set up parameter values based on training data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$.

To find the estimate of parameter $\hat{\lambda}$ by ML maximizes the log-likelihood function

$$\hat{\lambda} = \arg \max_{\lambda} \prod_{i=1}^N p(\mathbf{x}_i|\lambda) = \arg \max_{\lambda} \sum_{i=1}^N \log p(\mathbf{x}_i|\lambda). \quad (23)$$

Notice that estimation of the UBM parameters is similar to the estimate of the Speaker model. For GMM training it is not necessary to use a high number of Gaussian mixtures at the beginning, but the number can be increased stepwise. The EM algorithm uses a log-likelihood function instead as an auxiliary function Q . Let us state a function Q which is defined as

$$Q(\lambda, \lambda) = \sum_{n=1}^N \sum_y p(y|\mathbf{x}_n, \lambda) \cdot \log p(\mathbf{x}_n, y|\lambda), \quad (24)$$

for (λ, λ) and for the case of $Q(\lambda, \bar{\lambda})$ it is defined as

$$Q(\lambda, \bar{\lambda}) = \sum_{n=1}^N \sum_y p(y|\mathbf{x}_n, \lambda) \cdot \log p(\mathbf{x}_n, y|\bar{\lambda}). \quad (25)$$

Then we can express

$$\sum_{n=1}^N \log \frac{p(\mathbf{x}_n|\bar{\lambda})}{p(\mathbf{x}_n|\lambda)} \geq Q(\lambda, \bar{\lambda}) - Q(\lambda, \lambda). \quad (26)$$

The equation implies that if we select new parameters $\bar{\lambda}$ by replacing the previous parameters λ , and a function Q increases simultaneously, then we obtain an increasing logarithm of the probability of log-likelihood function too. In other words, if we use “better” parameters for Q we obtain better parameters for model $p(\mathbf{x}|\lambda)$. The solution of the EM algorithm then follows equations 2.44 - 2.45. A derivation of the function can be found in [MUL99] or in [PSU06].

General solution. Based on previous equations 2.41 - 2.43 we can derive general solution of statistical model parameter estimation using EM algorithm:

1. Initiate start parameter values

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_0.$$

2. Calculate the expectation over all values of y and over all observations \mathbf{x}_n

$$Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) = \sum_{n=1}^N \sum_y P(y|\mathbf{x}_n, \boldsymbol{\lambda}) \cdot \log p(\mathbf{x}_n, y|\bar{\boldsymbol{\lambda}}). \quad (27)$$

3. Select the parameters set $\boldsymbol{\lambda}^*$ from all of possible values of parameters sets $\bar{\boldsymbol{\lambda}}$, for which a maximization of a function Q occurs

$$\boldsymbol{\lambda}^* = \arg \max_{\bar{\boldsymbol{\lambda}}} Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}). \quad (28)$$

4. Set up new parameter values

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$$

5. Go to the step 2 and repeat the algorithm.

In case of **GMM** we can express

$$P(y|\mathbf{x}_n, \boldsymbol{\lambda}) = \frac{w_y N(\mathbf{x}_n, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)}{\sum_{y=1}^M w_y N(\mathbf{x}_n, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)}, \quad (29)$$

and

$$p(\mathbf{x}_n, y|\bar{\boldsymbol{\lambda}}) = w_y N(\mathbf{x}_n, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y). \quad (30)$$

Note that the variable y is an index of particular Gaussian.

Particular solution, GMM. Based on equations (27) - (30) the EM algorithm for GMM parameters estimation can be derived. The aim of the training step in GMM-UBM modelling is to estimate the parameters of the GMM $\boldsymbol{\lambda}$, which in some sense best matches the distribution of the training feature vectors [DEH09]. The parameters can be obtained iteratively using EM algorithm so that

$$P(\mathbf{X}|\bar{\boldsymbol{\lambda}}) \geq P(\mathbf{X}|\boldsymbol{\lambda}). \quad (31)$$

Let us suppose we have parameters of the UBM model. Now we have to estimate parameters of a new speaker. First, we initiate the start parameter values by known parameters of UBM model

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_{UBM}.$$

Then the equation (29) is used and new parameters are calculated as follows [REY95]. For each updated GMM $\bar{\lambda}$ of the y -th Gaussian component the mixture weights \bar{w}_y are defined as

$$\bar{w}_y = \frac{1}{N} \sum_{i=1}^N P(y|\bar{\mathbf{x}}_i, \lambda), \quad (32)$$

means $\overrightarrow{\bar{\boldsymbol{\mu}}_y}$ as

$$\overrightarrow{\bar{\boldsymbol{\mu}}_y} = \frac{\sum_{i=1}^N P(y|\bar{\mathbf{x}}_i, \lambda) \bar{\mathbf{x}}_i}{\sum_{i=1}^N P(y|\bar{\mathbf{x}}_i, \lambda)}, \quad (33)$$

and variances $\overrightarrow{\bar{\boldsymbol{\Sigma}}_y}$ as

$$\overrightarrow{\bar{\boldsymbol{\Sigma}}_y} = \frac{\sum_{i=1}^N P(y|\bar{\mathbf{x}}_i, \lambda) \mathbf{x}_i^2}{\sum_{i=1}^N P(y|\bar{\mathbf{x}}_i, \lambda)} - \overrightarrow{\bar{\boldsymbol{\mu}}_y}^2. \quad (34)$$

The mixture index y varies from 1 to C , $\overrightarrow{\bar{\boldsymbol{\Sigma}}_y}$, $\bar{\mathbf{x}}_i$, and $\overrightarrow{\bar{\boldsymbol{\mu}}_y}$ refer to elements of the particular vectors $\vec{\Sigma}_y$, $\vec{\mathbf{x}}_i$, and $\vec{\boldsymbol{\mu}}_y$. The iteration of EM algorithm is repeated until some convergence threshold is reached, i.e. a significant change in parameters occurs. Notice an article [CHU12] describes the use of the so-called FBEM algorithm based on the EM algorithm used for bird classification.

3.4.10 JFA

The Joint Factor Analysis (JFA) was introduced by [KEN07]. It operates with a so-called GMM *supervector*, which is defined by concatenating the mean vectors associated with individual Gaussians in the GMM of the particular speaker S . The supervector distribution is assumed to be Gaussian [KEN07]. The JFA reduces the dimension of used supervector, which leads to a decrease of training data sources needed for modelling. In the GMM approach each speaker is represented by a model $\boldsymbol{\lambda}$ composed by M Gaussians, each consists of weight w , mean vector $\boldsymbol{\mu}$ and diagonal matrix $\boldsymbol{\Sigma}$. Contrary to GMM-UBM, which uses data with no differences between speaker and channel, JFA decomposes data in order to discover channel effects and the speaker identity component. It combines both eigenvoice adaptation and eigenchannel adaptation for modelling speaker- and channel-variability, respectively. The JFA assumes that GMM supervector can be decomposed into a sum of supervectors

$$\mathbf{M} = \mathbf{S} + \mathbf{C} \quad (35)$$

where \mathbf{S} is a supervector dependent on the speaker and \mathbf{C} is supervector dependent on the channel. In context of GMM the supervector \mathbf{S} can be expressed as

$$\mathbf{S} = \mathbf{m} + \mathbf{V}\mathbf{y}$$

where \mathbf{m} corresponds to supervector constructed by concatenating the UBM means, \mathbf{V} matrix represents the eigenvectors of between-speaker covariance matrix, and vector \mathbf{y} is the channel independent speaker component. The supervector \mathbf{C} depends on a channel and can be expressed as

$$\mathbf{C} = \mathbf{U}\mathbf{x}$$

where \mathbf{U} matrix represents the eigenvectors of (channel) within-speaker covariance matrix, and vector \mathbf{x} describes the channel component of given supervector. The result equation can be written as

$$\mathbf{M} = \mathbf{V}\mathbf{y}^S + \mathbf{m} + \mathbf{U}\mathbf{x}^{H,S} + \mathbf{D}\mathbf{z} \quad (36)$$

where the upper indexes “ S ” and “ H ” denote speaker and record identifier, respectively, and matrix \mathbf{D} and vector \mathbf{z} represent characteristics of some additional noise.

The idea is based on finding speaker- and channel- dependent correlations among the data. JFA significantly reduces the supervector dimension, see [KEN07] or [DEH09].

3.4.11 i-Vector

Another advanced technique reducing the dimension of the supervector is the so called *i-vector* described in [DEH09], [SEN10]. Dehak found that separation of channel and speaker dependent data is partially successful. He proved that a large amount of data with high channel and speaker variability can give similar results as by using JFA. He modelled the GMM supervector as follows

$$\mathbf{M} = \mathbf{m} + \mathbf{T}\mathbf{w}, \quad (37)$$

where \mathbf{m} is a high-dimensional speaker- and channel-independent supervector which can be estimated using UBM. The rectangular matrix \mathbf{T} is called the *Total Variability Matrix*. It is estimated by the EM algorithm using a high amount of speech data containing both speaker- and channel-variabilities. The low-dimensional vector \mathbf{w} is denoted the identity vector (so called *i-vector*) that depends on the speaker as well as on the channel of given speech recording. The matrix \mathbf{T} is estimated on a large population of development data and once estimated it remains unchanged. We only have to calculate the \mathbf{w} vector. However, computation of the \mathbf{T} matrix is difficult and sources demanding, so simplification approaches are still under research; see for example [GLE11] or [DEH11].

For detail description of both the PLDA, and a Tool used in our experiments, see section 4.3.

3.5 Feature extraction

Feature vectors for particular segments are described as

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}. \quad (38)$$

The variable T is a number of samples of particular speech. Clearly, T depends on length of the speech and number of windows (which depend on the window width and the overlap). Every feature vector consists of D parameters

$$\begin{aligned} \mathbf{x}_1 &= \{x_{11}, x_{12}, \dots, x_{1D}\} \\ &\dots \\ \mathbf{x}_T &= \{x_{T1}, x_{T2}, \dots, x_{TD}\} \end{aligned} \quad (39)$$

where D denotes their total number. Figure 3.27 illustrates basic framework for feature vector extraction and feature vector origination.

Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Coding (LPC) [DEH09], and Perceptual Linear Predictive (PLP) analyses are the most important methods used for speech parameterization [BIM04].

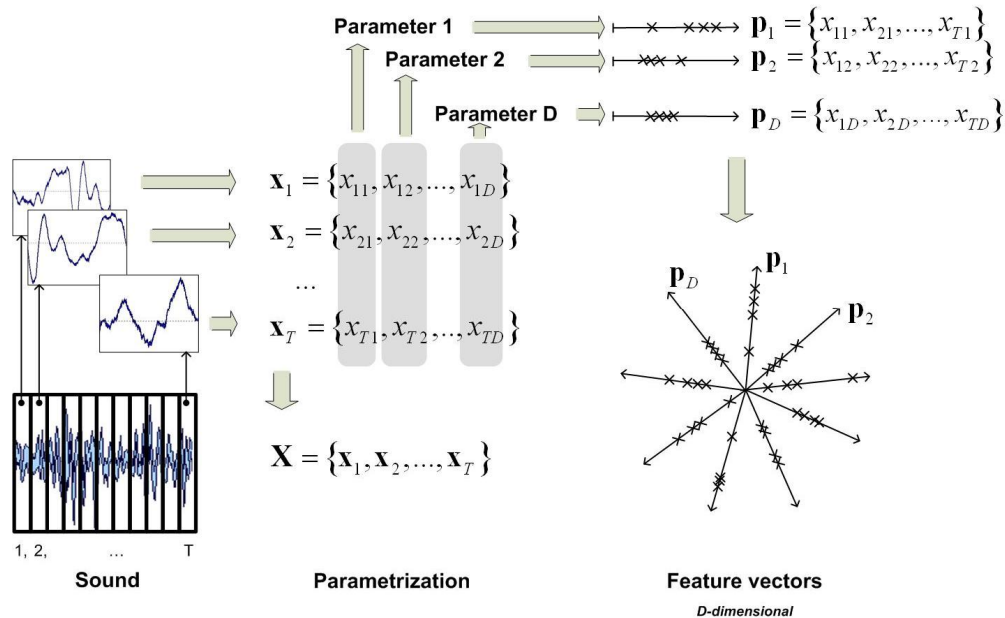


Figure 3.27: Parameter extraction and feature vectors origination.

3.5.1 Vocal tract model, cepstral coefficients

From the anatomical point of view, the vocal tract of a passerine is similar to humans, see section 3.1.3. Human vocal tract is shown in Figure 3.28. When air flows from the lungs the glottis stays open during breathing, and during speech production it is being opened and closed while it vibrates. The air flows through the vocal cord causing oscillation and producing sound. The vocal fundamental frequency F_0 is based on these vibrations. When creating the *voiced* vowels the glottis is nearly closed. When the voiced consonants are produced the glottis is not closed so tight causing the sound of non-periodical (tonal, pure) character. When creating *unvoiced* sounds the vocal cords are almost open and the sound is created by modification of the air stream in the cavities.

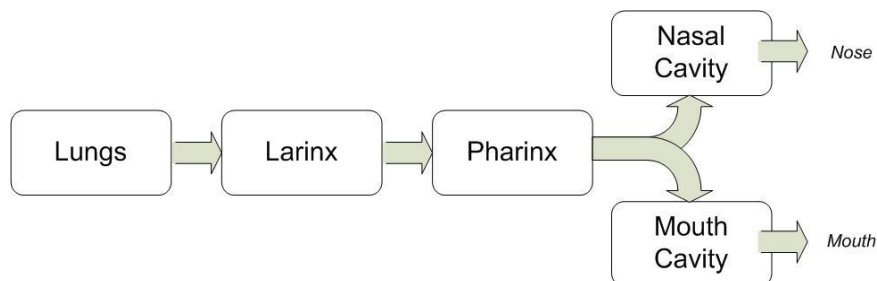


Figure 3.28: Human vocal tract.

For a detailed description of the vocal tract a *voice model* was created using an equivalent circuit diagram. Since the real process of vocalization is very complicated, a so called *stationary model* was used for both human and bird voices. The voice producing process is divided into *time frames*. If the length of the frames is short enough we may presume that the vocal tract is in a stationary state within

the frame. Each of the states is described in detail, then the voice producing process can be described as the sequence of these stationary states and the process transforms into *quasi-stationary*.

The aim of the modelling is to define the length of these time frames. For human voice it is usually about 30 ms. It is difficult to set a standard values for birds, see chapter 5.4.

A circuit diagram of the vocal tract is shown in Figure 3.29. Two generators are connected to the circuit to process sound. *Pulse generator* is dedicated for voiced sounds and *white noise source* for unvoiced.

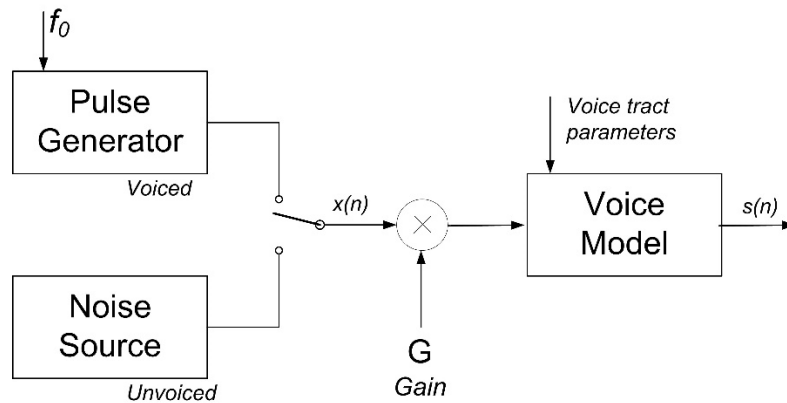


Figure 3.29: Vocal tract, equivalent circuit diagram.

After simplification, the whole process can be replaced by the source signal $x(t)$ passing through the system with impulse response $h(t)$, as shown in Figure 3.30

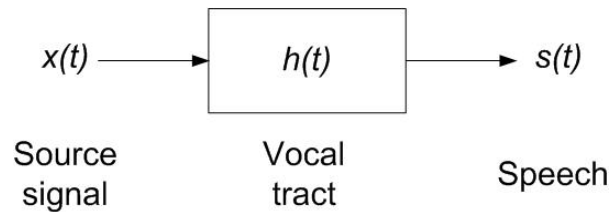


Figure 3.30: Vocal tract, simplification.

Human speech is then modelled by convolution of the excitation signal $x(t)$ and the vocal tract with the impulse response $h(t)$. This is used for speech synthesis purposes as well as for finding the speaker voice characteristics. In this thesis, the same approach is used to obtain bird vocal characteristics. Speech dependence on time can be described as a convolution of two signals

$$x(t) * h(t) = s(t), \quad (40)$$

and for a discrete signal

$$x(n) * h(n) = s(n). \quad (41)$$

In a frequency domain the convolution transforms into multiplication and we obtain

$$X(f) \cdot H(f) = S(f), \quad (42)$$

and if Z-transform is used then

$$X(z) \cdot H(z) = S(z). \quad (43)$$

To obtain vocal characteristics it is necessary to perform deconvolution, Figure 3.31.

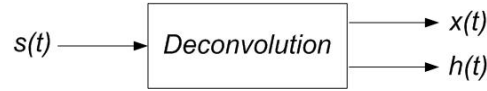


Figure 3.31: Deconvolution.

The real cepstrum is defined as the Inverse Discrete Fourier transform (*IDFT*) of the logarithm of the amplitude signal spectrum. For the n -th cepstral coefficient $c(n)$ is applicable

$$c(n) = IDFT \{ \ln |S[k]|, n \} = \sum_{k=0}^{N-1} \ln |S[k]| \exp \left\{ j \frac{2\pi kn}{N} \right\} \quad (44)$$

where $S[k]$ is a signal spectrum of the discrete signal. We use the real cepstrum because $c(n) = c(-n)$ (even function). So the real particular cepstral coefficients (speech, excitation and impulse response) can be expressed as

$$c_s(n) = \text{Re} \left\{ IDFT \left\{ \ln [S(k)], n \right\} \right\}, \quad (45)$$

$$c_x(n) = \text{Re} \left\{ IDFT \left\{ \ln [X(k)], n \right\} \right\}, \quad (46)$$

$$c_h(n) = \text{Re} \left\{ IDFT \left\{ \ln [H(k)], n \right\} \right\}. \quad (47)$$

For automatic recognition it is necessary to obtain parameters $x(n)$ and $h(n)$ separately. It is assumed that the speech signal is given by the convolution of two input signals, then for the cepstral coefficient applies:

$$c_s(n) = \text{Re} \left\{ IDFT \left\{ \ln [S(k)], n \right\} \right\} = \text{Re} \left\{ IDFT \left\{ \ln [X(k) \cdot H(k)], n \right\} \right\} \quad (48)$$

$$c_s(n) = c_x(n) + c_h(n). \quad (49)$$

Equation 2.13 transforms into a sum of coefficients. In practice, separation is achieved by a so-called liftering. The lower coefficients represent the spectral envelope, i.e. the vocal tract, the higher are the excitation coefficients. Typically, for SR tasks about 20 cepstral coefficients are used. Notice that the MFCC algorithm uses the discrete cosine transform (DCT) instead of the Inverse Fourier transformation, see section 3.5.4.

3.5.2 Hamming window

The most commonly used window is Hamming or Hanning window. Both windows taper the original signal on the sides and thus reduce the side-effects [BIM04]. The Hamming window reduces leakage in the spectrum due to its side-effects. It is defined as

$$w[n] = 0.54 - 0.46 \cos \frac{2\pi n}{N}, \quad (50)$$

where $0 \leq n \leq (N - 1)$, and N indicates the number of samples in length windows.

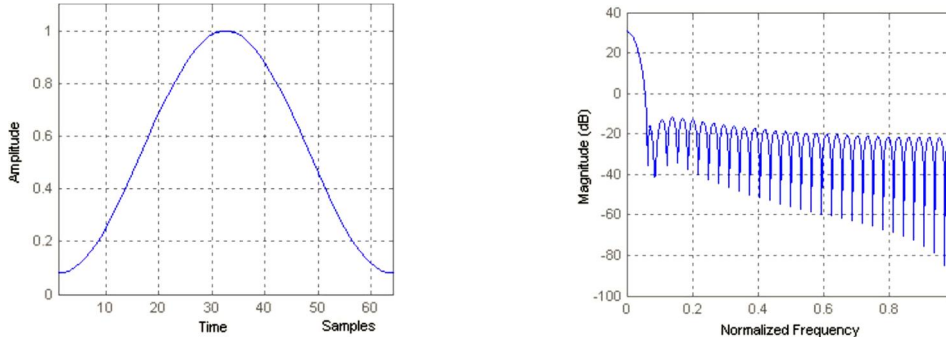


Figure 3.32: Hamming window a) Time domain, b) Frequency.

The length of the window is chosen so that the signal can be considered quasi-stationary. For the speech signal a Hamming window of length 30 ms with steps of 15 ms is usually used.

3.5.3 Pre-emphasis

During the progression of sound through the articulatory organs higher frequencies are normally suppressed. This suppression is compensated by the application of a first-order filter, which amplifies the higher frequency components. For modulating the filter shape

$$x_p(t) = x(t) - a \cdot x(t-1), \quad (51)$$

where a is a predefined pre-emphasis coefficient. In discrete domain then

$$s_p[n] = s[n] - a \cdot s[n-1]. \quad (52)$$

Pre-emphasis coefficients are usually chosen in the interval from 0.95 to 0.99 [BIM04].

3.5.4 Mel frequency cepstral coefficients

Mel Frequency Cepstral Coefficients (MFCC) are often used for speaker recognition systems. In this case, a sliding window is used to divide the speech into short segments. Each signal segment is then pre-emphasised. Next, the Mel-frequency filter is applied to better adapt signal to human hearing. Due to logarithmic calculation the multiplication of spectrum changes to addition. Finally, by application of the Discrete Cosine Transform (DCT), cepstral coefficients are obtained. The cepstrum is the so-called *Mel-frequency cepstrum* because a Mel-frequency filter is used within the process.

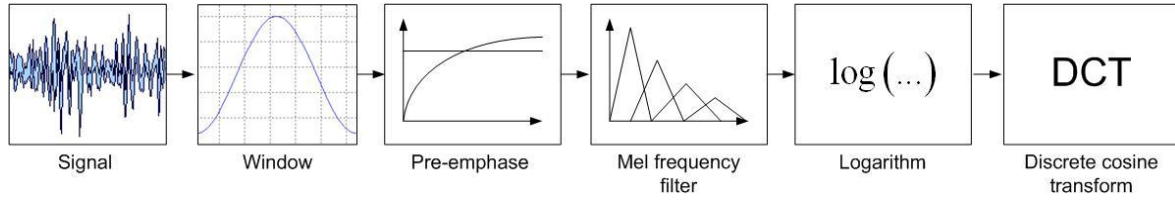


Figure 3.33: Mel-frequency cepstral coefficients computing, data diagram.

The Mel-filter bank is adapted to human hearing. We designed a new so-called Bird Adapted Filter distribution (BAF) tailored to the bird's song. In our experiments, we also used linearly distributed filters. So we compared results from the three filtering structure: BAF, Mel, and Linear. See section 8. for details.

3.5.4.1 Mel filter bank.

It was empirically found that the human ear perceives frequency sound intensity signals with dependence on the frequency. Therefore, in applications of automatic speech recognition it is desirable to adjust the signal so that its distribution is near to the hearing. Mel filter banks are used for this correction. They convert the frequency f [Hz] into a so-called frequency f_{MEL} [mel] which is based on human hearing. The conversion between f and f_{MEL} is defined by the relationship

$$f_{MEL} = 2595 \log_{10} \left(1 + \frac{f}{700} \right). \quad (53)$$

Figure 3.34 shows the behaviour of the function.

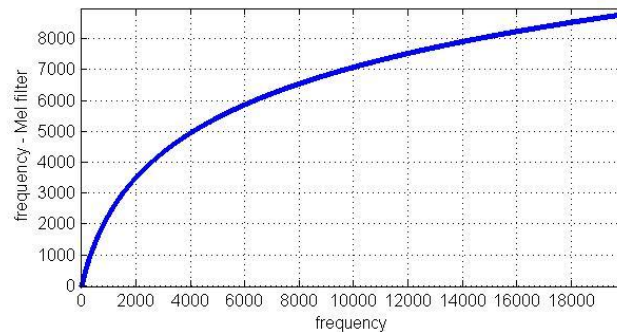


Figure 3.34: Characteristic Mel-frequency [mel] and frequency [f] domains.

For the reversed conversion the following relationship is valid

$$f = 700 \left(10^{\frac{f_{MEL}}{2595}} - 1 \right). \quad (54)$$

Mel filter banks are realized by a set of M bands. For instance, if the bandwidth is 4 kHz 20 banks are usually used. These filters have triangular shapes with bands overlapping by half, and they are distributed non-linearly in the frequency domain as depicted in Figure 3.35.

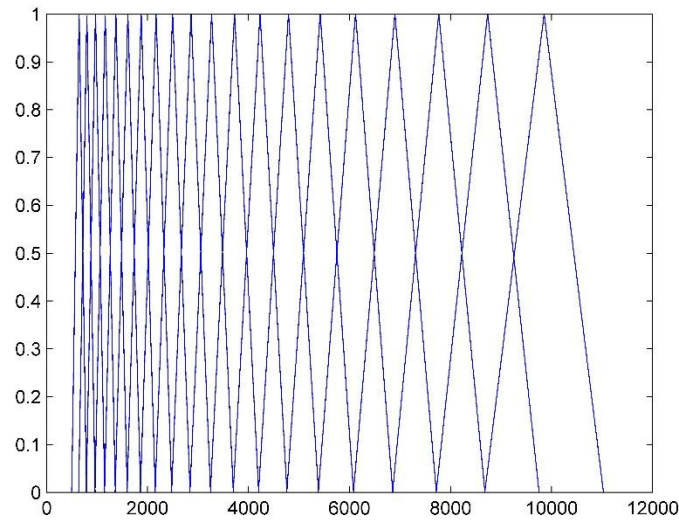


Figure 3.35: Mel filter bank.

The use of triangular overlapping filters helps to modify the magnitude of the spectrum with respect to human hearing.

3.5.4.2 *Linear filter bank.*

A linear filter bank distribution is demonstrated on Figure 3.36.

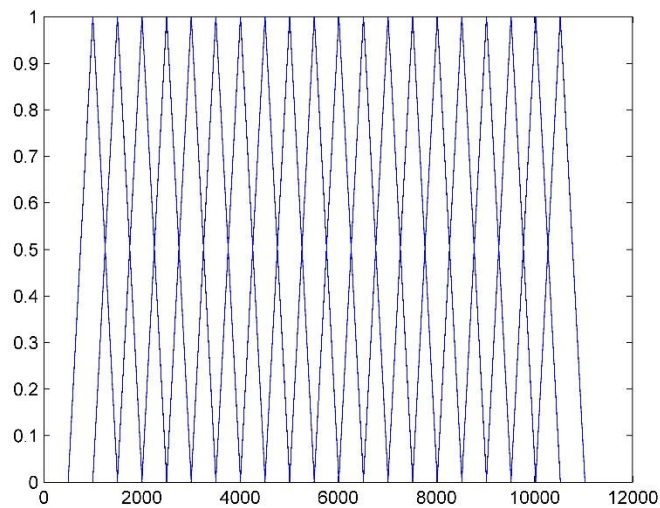


Figure 3.36.: Linear filter bank.

3.5.5 **Perceptual Linear Predictive analysis**

Perceptual Linear Predictive (PLP) method was introduced by [HER90]. It implicates three concepts based on the psychophysics of hearing to derive to an estimate of the auditory spectrum. The critical-band spectral resolution, the equal-loudness curve, and the intensity-loudness power law [HER90].

Unlike the Linear predictive analysis (LP), PLP analysis takes basic characteristics of human hearing into consideration, namely non-linear hearing, sound masking, and aural range.

The block diagram of PLP is shown in Figure 3.37. Now will be described each step involved in PLP speech analysis.

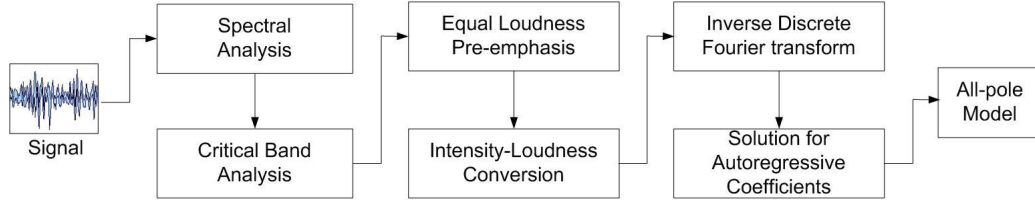


Figure 3.37: Block diagram of PLP speech analysis.

Spectral analysis. The signal is weighted by Hamming window. The Discrete Fourier Transform is then used to transform the signal (speech, song) segment into the frequency domain. Finally, a short-term power spectrum is calculated from the spectrum $S(\omega)$ of the segment

$$P(\omega) = |S(\omega)|^2 = [\text{Re } S(\omega)]^2 + [\text{Im } S(\omega)]^2. \quad (55)$$

Critical Band. The PLP warps the spectrum $P(\omega)$ from frequency ω rad/s into barks $\Omega(\omega)$ by

$$\Omega(\omega) = 6 \ln \left(\frac{\omega}{1200\pi} + \sqrt{\left(\frac{\omega}{1200\pi} \right)^2 + 1} \right), \quad (56)$$

where the angular frequency $\omega = 2\pi f$ rad/s. The resulting warped power spectrum is then convolved with the power spectrum of the simulated critical-band masking curve $\Psi(\Omega)$. The critical-band curve is given by

$$\Psi(\Omega) = \begin{cases} 0 & \text{for } \Omega < 1.3, \\ 10^{2.5(\Omega+0.5)} & \text{for } -1.3 \leq \Omega \leq 0.5, \\ 1 & \text{for } -0.5 < \Omega < 0.5, \\ 10^{-1.0(\Omega-0.5)} & \text{for } 0.5 \leq \Omega \leq 2.5, \\ 0 & \text{for } \Omega > 2.5. \end{cases}$$

which is an approximation to the asymmetric masking curve. The filter band is cut off at -40 dB. Following discrete convolution of $\Psi(\Omega)$ and $P(\omega)$ gives

$$\Theta(\Omega_i) = \sum_{\Omega=-1.3}^{2.5} P(\Omega - \Omega_i) \Psi(\Omega). \quad (57)$$

The convolution significantly reduces the spectral resolution of $\theta(\Omega)$ in comparison with the original $P(\omega)$.

Equal-loudness pre-emphasis. The sampled $\theta\Omega(\omega)$ is pre-emphasized by the simulated equal-loudness curve

$$\Xi[\Theta(\omega)] = E(\omega)\Theta[\Omega]. \quad (58)$$

Function $E(\omega)$ is an approximation of the ear hearing different frequencies at different levels. Such approximations are known as the equal-loudness contours (Fletcher–Munson curves, Robinson–

Dadson ISO 226). An example of $E(\omega)$ for the level 40 Phon derived from equal-loudness contours [PSU06] is

$$E(\omega) = K \frac{\omega^4 (\omega^2 + 56.9 \cdot 10^6)}{(\omega^2 + 6.3 \cdot 10^6)^2 (\omega^2 + 379.4 \cdot 10^6) (\omega^6 + 9.6 \cdot 10^{26})}, \quad (59)$$

where K is a parameter which normalizes to an equal loudness of 0 dB.

Intensity-loudness power law. This power law is an approximation of the hearing law, which defines relationship between the intensity of sound and the perceived loudness and is the cubic-root compression of the pre-emphasized signal

$$\Phi(\Omega_i) = \Xi(\Omega)^{0.33}. \quad (60)$$

This operation also reduces the spectral-amplitude variation of the critical-band spectrum so that the following all-pole modelling can be done with a relatively low model order [HER90].

Autoregressive modelling. The function $\Phi(\Omega)$ is approximated by spectrum of an all-pole model. It uses autocorrelation method of all-pole spectral modelling using the Inverse Fourier Transform (IDFT) [HER90].

Weighted spectral summation. The passage of short-term spectrum through the m -th critical band filter can be formulated as

$$\Xi[\Omega(\omega_i)] = \sum_{\omega=\omega_{il}}^{\omega_{ih}} \omega_i(\omega) P(\omega). \quad (61)$$

The sum limits ω_{il} and ω_{ih} are calculated as $\omega_{il} = 1200\pi \sinh((\Omega_i - 2.5)/6)$ and $\omega_{ih} = 1200\pi \sinh((\Omega_i + 1.3)/6)$ respectively.

For detail derivation of auto-correlation function see [PSU06]. It is obvious that using PLP to analyse bird songs is limited by the impossibility to set up accurate parameters of a bird's hearing. In contrary to research on human hearing it is not possible to conduct an experiment where the test object (i.e. a bird individual) cooperates with the inquirer. Some sources bring basic knowledge of bird masking and hearing, see [CAT08], [MAR04]. That information is based on research of bird hearing concerning anatomy, impedance measurement, autopsy, and laboratory experiments. Other facts are based on reasonable presumptions. For example, [MAR04] says that bird has to hear what he sings.

3.5.6 Linear prediction cepstral coefficients

Linear prediction coding (LPC) predicts speaker parameters directly from a speech signal. The main stages of LPC calculation are shown in the Figure 3.38.

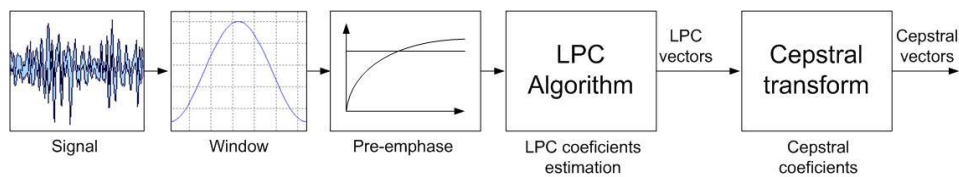


Figure 3.38: LPC coefficients calculation stages [BIM04].

The principle of LPC is computation of an $s(n)$ sample of a voice as a linear combination of previous samples with excitation $u(k)$ enhanced with the gain G , so

$$s(k) = -\sum_{i=1}^Q a_i s(k-1) + Gu(k) \quad (62)$$

where G is the gain coefficient and Q is the order of the model. Transfer function $H(z)$ can then be written as

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{A(z)} \quad (63)$$

where $H(z)$ is defined as

$$H(z) = \frac{G}{1 + \sum_{i=1}^Q a_i z^{-1}} \quad (64)$$

Figure 3.39 shows the block diagram of cepstral coefficient extraction. The signal has to be weighted by a window short enough to be considered approximately stationary. This can be used to determine the parameters a_i and G using the method of least squares.

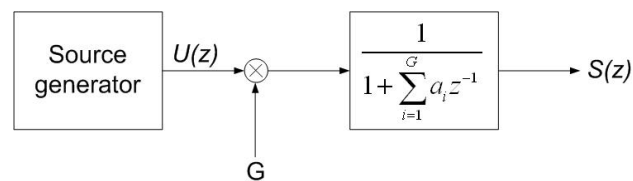


Figure 3.39: LPC, cepstral coefficients.

4 Development Framework

The experimental framework was developed from the scratch, and it consists of Matlab procedures and functions (Mathworks Inc. 2010) as well as C++ modules. The author programmed the Matlab work as well as the Experiment Manager module see detail in section 4.1. My colleagues from Faculty of applied science implemented a *SV tool* programmed in C++ see detail in section 4.2.

Main framework modules are:

1. Experiment manager. The experiment process is fully automated. The user can start any number of experiments in one sitting. Parameters are set up individually for each experiment and is implemented in Matlab.
2. Recording classification. The manager creates lists of files: gmm training, ubm training, and testing. It also generates the list of trials. The lists are clustered into so-called *data sets*. The process is fully automated with advanced features, i.e. file sequential or random sorting, file selecting, etc., all implemented in Matlab.
3. Feature extraction. Extraction algorithms are implemented in Matlab as well as in SV tool.
4. Support modules. New modules implemented in this work are VAD, BAF, and Data merging. All are implemented in Matlab, and the VAD is implemented in SV tool.
5. Model estimation. GMM/EM module, the MAP adaptation module, and score calibration algorithms are implemented in SV tool.
6. Verification. Decision algorithms are implemented in SV tool.
7. Experiment evaluation. Resulting statistics and EER calculation are obtained. Evaluation is automatically processed and linked with data sets along with experimental parameters, all implemented in Matlab.

All experiments described in this thesis were processed using these modules.

4.1 Matlab

The experimenter set up both the data and parameters for the required experiments, and he or she starts up the *experiment manager* module. Then the framework operates automatically. Compared to the common methods (requiring data manual handling, visual check, software manual control, manual parameter setting, etc.) the system is very effective.

4.1.1 Experiment manager

See Figure 4.5 for the block diagram of the experiment modules flow. The experiments are controlled by the so-called experiment manager module which is able to automatically perform up to 999

sequential experiments. Notice we do not use parallel experiment performance because we do not have a particular Matlab license.

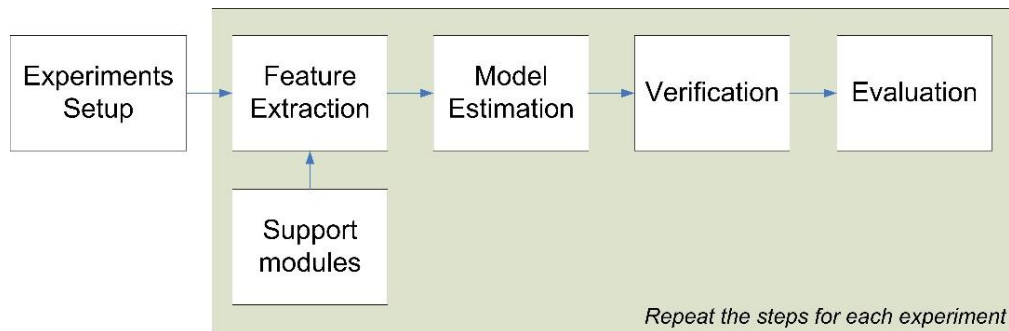


Figure 4.1: Experiment manager, block diagram.

Both the parameters and input data are defined independently of each experiment. Moreover, an experiment can use the data parametrized by a previous experiment to save computing time. For the relationships between experiment set up, result, parameters, and data see Figure 4.2.

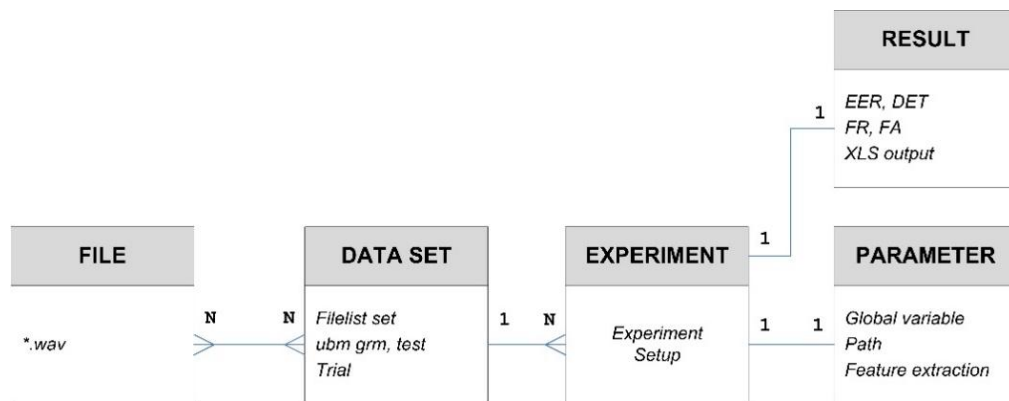


Figure 4.2: Experiment entities relationship.

The experiment setup is defined by an experiment parameters definition Excel sheet, see Figure 4.3. Each row contains specific parameters for an experiment as well as a link to the particular data set. The Excel data are uploaded automatically into the Matlab during experiments initialization.

	A	B	E	F	J	K	L	M	N	O	Q	R	S	T	W	X	Y	Z
1	param_01	param_04	param_05	param_06	param_10	param_11	param_12	param_13	param_14	param_15	param_17	param_18	param_19	param_20	param_23	param_24	param_25	param_2
2	Nrxpermt	ShodnZnak	SetFile	Trials	Wdelka	Wshift	Wtype	MinFreq	MaxFreq	Koefficient	NumFilter	Shape	IvarFiltru	TypFiltru	NumCept	Delta	DeltaDelta	Numt
40	38	3	53	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
41	39	3	53	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
42	40	3	53	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
43	41	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
44	42	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
45	43	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
46	44	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
47	45	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
48	46	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
49	47	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
50	48	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
51	49	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
52	50	3	57	1	0,02	0,01	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
53	51	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
54	52	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
55	53	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
56	54	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
57	55	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
58	56	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
59	57	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
60	58	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
61	59	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
62	60	3	57	1	0,02	0,02	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
63	61	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
64	62	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
65	63	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
66	64	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
67	65	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
68	66	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
69	67	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4
70	68	3	57	1	0,03	0,015	2	1000	0	0,97	20	0,0	0	1	20	1	0	4

Figure 4.3: Experiment parameters definition.

4.1.2 Recording classification, data set

First, it is necessary to sort the data (i.e. recordings). An ornithologist usually provides the data with specific names due the bird identification. For instance, one of the origin file names may be: *PC1107-110613-MZ000011.wav*. The name, thus, consists of the following tags:

- PC... Phylloscopus collybita (chiffchaff)
- 1107...bird id,
- 110613...ring id,
- MZ000011...recording number, ornithology's internal counting of recordings.

However, it is not necessary to keep information about the *bird_id* and *ring_id* in our experiments. Therefore, we usually use simplified names. For instance, the file mentioned above was renamed to *G09_005.wav*, which implies:

- G...bird id,
- 09...recording id,
- 005...song id.

Notice, not all experiments need a *song_id* information; therefore, we use a simplified name *G09.wav*.

Four experiment file lists are prepared after data sorting and recording name unification. The list consists of the names of the files; thus, it defines which recording belong to the particular step (training, test, etc.). Each experiment needs a file list for the following steps:

- *Ubm*
UBM model estimation
- *Training*
GMM model(s) estimation. Sometimes we call the bird(s), to whom the recordings belong, target bird(s).
- *Testing*
Identification/Verification process (for differences between identification and verification

see Figure 5.4). Includes an unknown bird(s) recordings to be identified. It also usually includes the recordings of a target bird to calibrate the system, however different recordings has to be used than for the GMM model estimation.

- *Trials*
Pair of GMM and Testing recordings.

In summary four lists of files have to be prepared: *filelist_ubm*, *filelist_test*, *filelist_train*, *trials*. These lists are clustered into one so-called *data set*. The classifying of records into the *file lists* as well as into the *data set* is fully automated with advanced features, i.e. file sequential or random choice, selection based on the song or recording number, sorting, validation check etc. The functionality is implemented in Matlab. A user defines the file matching conditions in an Excel sheet (see Figure 4.4) and then the *recording classification* module performs the files assignment automatically.

ID	ASCII	RecsCelk	Typ	1..KolikUBM	2..KolikGMM	3..KolikTST	RecsOd	RecsDo	SongOd	SongDo	Legenda	Vysledky	VysledkyKrat	Kontrola
A	65	488	1	0	40	40	1	1000	16	26	celkemUBM	200	200	././.
B	66	329	1	40	0	0	1	1000	16	26	celkemGMM	40	40	././.
C	67	332	1	0	0	40	1	1000	16	26	celkemTST	320	440	././.
D	68	226	1	40	0	0	1	1000	16	26	TST false	280	400	././.
E	69	240	1	0	0	40	1	1000	16	26	TST true	40	40	././.
F	70	788	1	40	0	0	1	1000	16	26	CelkemFiles	560	680	././.
G	71	545	1	0	0	40	1	1000	16	26				././.
H	72	598	1	40	0	0	1	1000	16	26	trialuFalse	11 200	16 000	././.
I	73	713	1	0	0	40	1	1000	16	26	trialuTrue	1 600	1 600	././.
J	74	304	1	0	0	40	1	1000	16	26	trialu	12 800	17 600	././.
K	75	98	1	0	0	40	1	1000	1	20	Pomer False/ True	7,00	10,0	././.
L	76	293	1	40	0	0	1	1000	3	14	Navic trialu	9 600	4 800	././.
M	77	221	1	0	0	40	1	1000	3	14				././.

Figure 4.4: File lists definition parameters.

Finally, the experiment setup is linked to the data sets with relationship 1:N, see Figure 4.2, so we can use the same data repeatedly.

4.1.3 Feature extraction

This Matlab module provides the feature extraction just as is described into the chapter 2. The operator defines feature extraction parameters (windows length, shift, type, etc.), chooses from many optional functions (VAD, BAF, etc.), and selects settings (for instance FFT length) by the parameters defined in the Excel sheet (see Figure 4.3.).

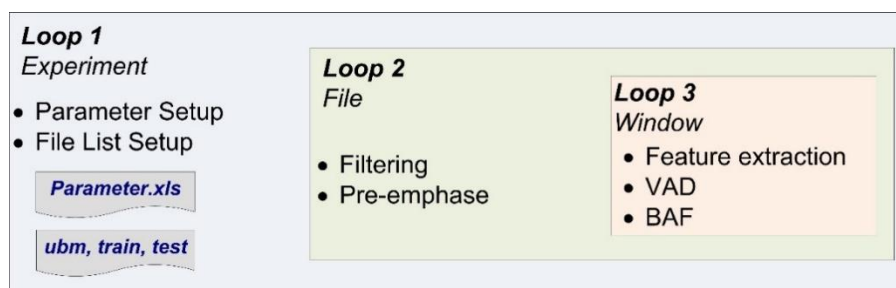


Figure 4.5: Feature extraction, block diagram.

From the global point of view the experiment process consists of the three loops: Experiment, File, and Feature extraction. The relations among loops demonstrates Figure 4.5.

4.1.4 Support modules

An additional functionality was implemented (VAD, BAF, Data merging, MEL filters overlap, etc.). Its setting and activation is defined by the parameters definition Excel sheet (see Figure 4.3.).

4.1.5 Model estimation

The module provides the GMM/EM, the MAP adaptation, and score calibration algorithms just as is described in chapter 2.

4.1.6 Verification and identification

The module provides the verification just as is described in section 5.3. and the second chapter.

4.1.7 Experiment evaluation

The experiment(s) result is saved into an Excel file. The result gives EER, FR, FA, and other values like number of true and false trials, see Figure 4.6. The DET curve draw is optional based on the experiment setting. Based on our experiences we also save the crucial parameters into the Excel because of easiest evaluation.

	A	B	C	D	E	F	G	H	I	J	K	O	P	T	U
1	param_01	param_30	param_31	param_32	param_33	param_34	param_35	param_36	param_37	param_38	param_39	param_05	param_06	param_10	param_11
2	NrExpermt	EER	TotalTrials	TotalError	TotalRate	TrueTrials	FR	TrueRate	FalseTrials:FA	FalseRate	SetFile	Trials	WdelKa	Wshift	
3	1	9,73%	16830	1788	89,38%	8415	474	0,943672014	8415	1314	0,84385	3	4	0,03	0,015
4	2	11,71%	16830	1934	88,51%	11550	965	0,916450216	5280	969	0,816477	10	4	0,03	0,015
5	3	13,90%	30000	4031	86,56%	15000	2172	0,8552	15000	1859	0,876067	94	4	0,03	0,015
6	4	13,71%	30000	4452	85,16%	19800	1894	0,904343434	10200	2558	0,749216	129	4	0,03	0,015
7	5	16,80%	30000	5062	83,13%	19800	2854	0,855858586	10200	2208	0,783529	135	4	0,03	0,015
8	6	18,19%	30000	4564	84,79%	21400	1930	0,909813084	8600	2634	0,693721	141	4	0,03	0,015
9	7	23,75%	30000	7073	76,42%	19800	3460	0,825252525	10200	3613	0,645784	149	4	0,03	0,015
10	8	13,98%	30000	4184	86,05%	20985	3055	0,854419824	9015	1129	0,847664	3	4	0,03	0,015
11	9	15,25%	30000	4539	84,87%	24120	3625	0,849709784	5880	914	0,844558	10	4	0,03	0,015
12	10	17,78%	30000	4555	84,82%	15000	3228	0,7848	15000	1327	0,911533	94	4	0,03	0,015
13	11	14,51%	30000	4410	85,30%	19800	2828	0,857171717	10200	1582	0,844902	129	4	0,03	0,015
14	12	17,58%	30000	5323	82,26%	19800	3110	0,842929293	10200	2213	0,783039	135	4	0,03	0,015
15	13	18,77%	30000	5268	82,44%	21400	2494	0,883457944	8600	2774	0,677442	141	4	0,03	0,015
16	14	22,74%	30000	6912	76,96%	19800	3416	0,827474747	10200	3496	0,657255	149	4	0,03	0,015
17	15	13,62%	30000	4040	86,53%	20985	3194	0,847796045	9015	846	0,906156	3	4	0,03	0,015
18	16	15,11%	30000	4552	84,83%	24120	3669	0,847885572	5880	883	0,84983	10	4	0,03	0,015
19	17	16,26%	30000	4193	86,02%	15000	3135	0,791	15000	1058	0,929467	94	4	0,03	0,015
20	18	15,15%	30000	4557	84,81%	19800	2766	0,86030303	10200	1791	0,824412	129	4	0,03	0,015
21	19	21,55%	30000	6402	78,66%	19800	3804	0,807878788	10200	2598	0,745294	135	4	0,03	0,015
22	20	17,92%	30000	4771	84,10%	21400	2680	0,874766355	8600	2091	0,75686	141	4	0,03	0,015
23	21	19,44%	30000	5811	80,63%	19800	3726	0,811818182	10200	2085	0,795588	149	4	0,03	0,015

Figure 4.6: An Excel file result.

4.2 Speaker verification tool

Aleš Padrta, Jan Vaněk, and Lukáš Machlica from the Department of Cybernetics, Faculty of Applied Sciences in Pilsen developed a *Speaker Verification tool* written in C++. The tool can perform whole SV process. The next chapters describe the whole tool's functionality; however, we use just some modules in our experiments, see chapter 0 *Introduction*.

4.2.1 Flow diagram

The SV task is divided into four stages:

1. PRM. Parameterization of all input files.
2. UBM. Creation of an UBM model.
3. GMM. Adaptation of UBM/GMM models.
4. VERIFY. Test phase of a speaker verification task.

The following Figure 4.7. demonstrates the SV tool flow diagram.

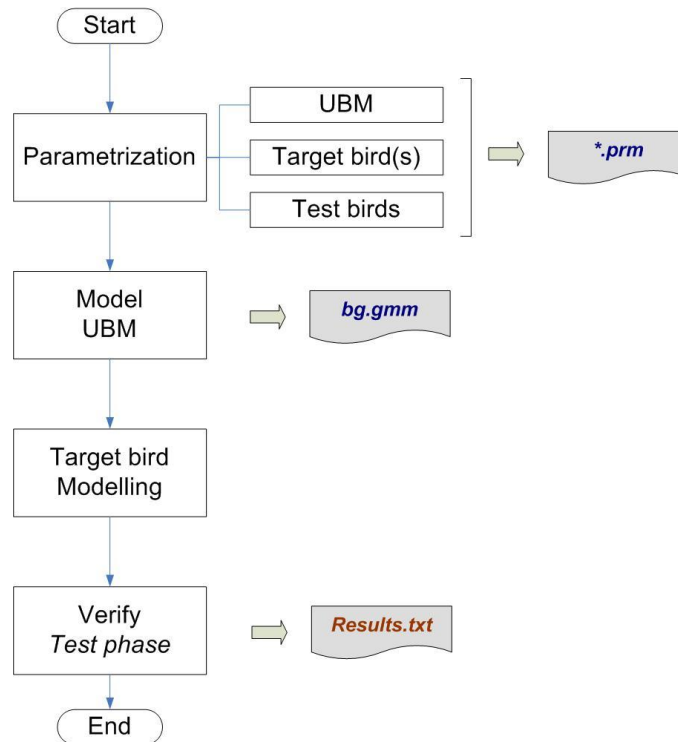


Figure 4.7: Function of SV tool, flow diagram.

In the first stage the corresponding parameters are extracted for every input **.wav* file. These parameters are saved into **.prm* files. In the next stage, a model of UBM (*ModelUBM*) is computed and saved as a *bg.gmm* file. Then follows an UBM/GMM model adaptation based on the incoming data. At the last stage, named *verification*, defined pairs of songs are tested. The matching probability is computed for every couple and the results are written into *result.txt*.

4.2.2 Input and output data

Table 4 summarizes the input and output data of all process stages. The recorded **.wav* files, listed in configuration files, are entered into the *parameterization* PRM stage. Further stages use data created here, and finally the last stage, *verify*, writes results into the *results.txt* file.

Process	Input directory	Input files	Output directory	Output files
PRM	WAV\	*.wav	UBM_DIR\ GMM_DIR\ TEST_DIR\	*.prm
UBM	PRM\UBM_DIR\	*.prm	UBM\	bg.gmm
GMM	PRM\GMM_DIR\	*.prm	models\	*.gmm

VERIFY	PRM\TEST_DIR\	*.prm	VERIFY\	result.txt
--------	---------------	-------	---------	-------------------

Table 4: Speaker verification tool, inputs and outputs

The described directory structure was designed to make data storage and manipulation easiest possible.

4.2.3 Results

The SV tool computes the probability of song/speech pair similarity. The probabilities are written into the file **results.txt**. Table 5 shows the SV tool output data. In the first column, *result.txt* probabilities are copied from the result.txt file. In the second column, couples of tested songs are placed, where the letter represents a bird individual while the number labels particular bird record. The last column describes the decision made by the supposed threshold $\Theta=0$. If the result value is lower than threshold Θ , the result is *rejected* and vice versa. In this case, the threshold is $\Theta=0$, however, the threshold value may differ for every particular task.

Trial	Calculated score	Result action	Correct result?
A01-B01	-2.177158	reject	yes
A01-A21	0.49978	accept	yes
A02-A07	2.836717	accept	yes
A02-A22	1.461095	accept	yes
A01-B06	-2.14189	reject	yes
A01-A22	-0.012654	reject	<i>Error: False reject</i>
A02-B08	-3.909118	reject	yes
A02-D03	0.328327	accept	<i>Error: False accept</i>
A01-C07	-4.295674	reject	yes
A01-A04	4.007644	accept	yes

Table 5: Output file *results.txt* example.

4.2.4 Using the SV tool

The SV tool does not require an installation. It is ready for both 32-bit and 64-bit OS. The system is running on Windows Vista and Windows 7. Before starting the application, it is necessary to set up configuration parameters and to determine the input files. For details see Table 6. The SV tool runs in CMD (Window Command Line). The program continuously prompts its status as well as the progress of the current operation.

File	Assignment
Param_KW.ini	Set up the parameters of feature vectors extraction. <i>For instance length of window, overlapping, number of MFCC parameters, switch on/off the pre-emphasis, etc.</i>
filelist_test	List of *.wav files for testing. <i>The files will be compared with trained speaker/bird (Target) during Verify/Test stage.</i>
filelist_train	List of *.wav files for GMM model. <i>The files will be compared with trained speaker/bird (Target) during Verify/Test stage.</i>
filelist_ubm	List of *.wav files for UBM model.
Model_KWGMM.ini	Set up the parameters for an UBM model creating process. <i>For instance Number of Gaussians, etc.</i>
Model_ADAPT.ini	Set up the parameters for a GMM model creating process. <i>For instance type of adapting (MAP, MLLR,...) , etc.</i>
Verify.ini	Set up the parameters for verification (test). <i>For instance Threshold, format of results written in results.txt, etc.</i>
Trials.ndx	List of testing pairs. <i>The final probability is computed for each trial and saved into results.txt file.</i>
Results.txt	List of computed probabilities for each trial.

Table 6: SV tool, configuration of the bird verification.

4.3 iVector tool

We used an iVector tool in our experiments developed by Jan Vaněk, Lukáš Machlica, and Zbyněk Zajíc from the Department of Cybernetics, Faculty of Applied Sciences in Pilsen. The tool was written in C++, similarly as the SV tool.

A speaker is represented by a supervector of accumulated statistics of speaker's data with respect to the Universal Background Model (UBM). The Factor Analysis (FA) decomposition is used to reduce the huge dimensionality of the supervector to a low dimensionality space vector – iVector. An iVector could be a final representation of the speaker; otherwise, it is further processed by the Probabilistic Linear Discriminant Analysis (PLDA) model to maximize the ratio of between- to within-class covariance in order to increase separability of given classes. For a block diagram of the iVectors process, see Figure 4.8.

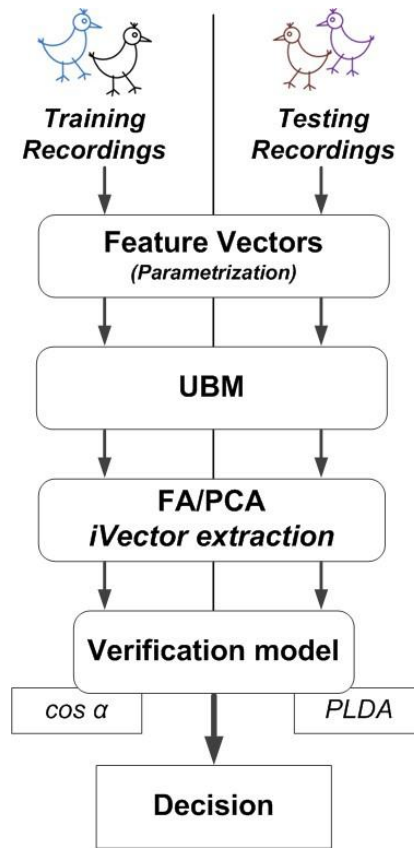


Figure 4.8: A block diagram of the Identity Vectors process.

The training of both the FA decomposition and PLDA model requires a huge amount of data. Although ornithologists usually record a few hundreds or thousands of songs, FA and PLDA training requires millions or much more recordings. Moreover, the records have to be precisely annotated (species, individuals). Due to the lack of bird recordings the system was trained by the speech recordings.

5 Bird individual identification using as-is recordings

5.1 Introduction

A presented research differs from mentioned works, where identification run on the close set [CLE05], [FOX08], songs or syllables have been extracted, or records have been pre-processed [TRA05], [CHE10], [BUD14], or even as-is recordings are used but for the species recognition [POT14], [VEN15]. [XIA11] deals with bird individual identification based on ANN and spectrographic cross-correlation. They generated spectrograms and measured some variables in Avisoft-SASLab software. Petruskova et. al [PET16] provided acoustic monitoring based on syllable repertoire. They proved it could be more efficient for individual recognition than colour ringing (for some species). They distinct elements of particular syllable types by visually checking of spectrograms in SW Avisoft. By contrast in the current work an advanced adaptive VAD is introduced. By this as-is recordings can be used independently on both its quality and length. Further, the system works on the open set (the number of birds is not known beforehand, a new individual may appear anytime). We realize that the introduced method cannot fully replace standard methods for bird identification (ringing, DNA) nowadays. However, it can be used alone if an absolute identification is not required, or as a support tool for these methods.

Although the GMM method was used in some previous works [CHE10], [GRA11] it meets the real condition requirements in this work thanks to the implementation of both an UBM and the VAD. To our knowledge this is the first application of individual identification of birds on the open set by processing raw recordings, fully-automated and without pre-processing.

The aims of this research are as follows:

- To demonstrate the feasibility of using the GMM-UBM with an advanced VAD algorithm for bird individuals identification in real conditions on the open set.
- To determine whether using as-is records without any pre-processing can give a reasonable accuracy.

We begin by introducing the recorded data and a description of the goal of the experiment followed by a description of the system from both the theoretical point of view and the system implemented. Then the experiment evaluation is described including error type and accuracy calculation. We end with a discourse on the results and open questions.

5.2 Bird song data

5.2.1 Chiffchaff

With estimated population of 90-180 mil. Individuals, chiffchaff (*Phylloscopus collybita*) belongs to most common European songbird species, see . It is a migrating bird wintering in Mediterranean and North Africa. It is small, (c. a. 8 g), inconspicuous but very vocal species with a distinct song. Males are territorial and defend their territory vigorously. They start to advertise their territories by singing soon after the spring migration at the end of March and beginning of April and continue to sing over the breeding season. The song of the chiffchaff is simple consisting of varying number of rhythmically repeated syllables transcribed as “chiff” and “chaff” (see Figure 5.2). Nevertheless,

each male can have over 10 different syllables in its repertoire that can be organized into “song types” (specific groups of syllables occurring together within a song).



Figure 5.1.: Chiffchaff (*Phylloscopus collybita*). © Kristyna Felendova.

According to our recordings the Chiffchaff sings, on average, 7.3 songs per minute. The average length of the song is approximately 12 syllables, equal to approximately 4 s [LIN12b]. The average syllable duration is 117 ms and the average inter-syllable interval is 234 ms [LIN13]. The band-width of the chiffchaff song lies between 2.5 and 7 kHz with most of sound energy concentrated around 4350 Hz [LIN12b].

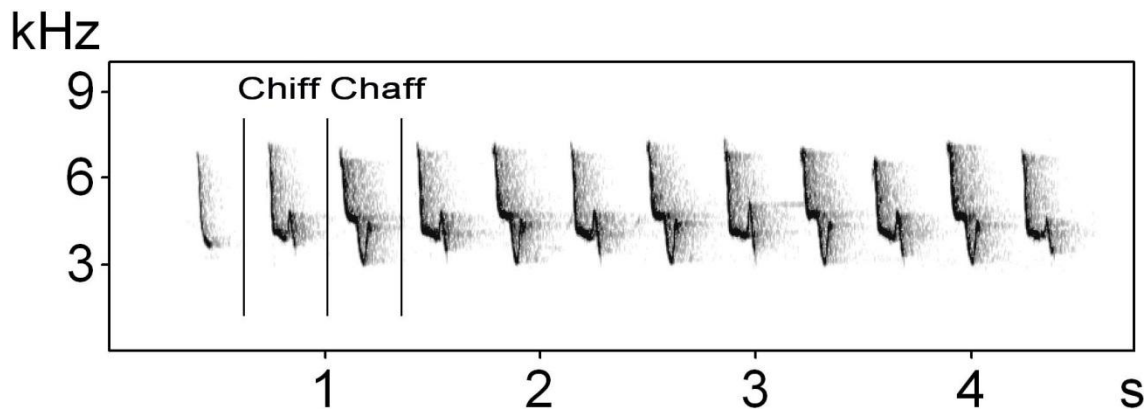


Figure 5.2: Single song of a chiffchaff male with the “chiff” and “chaff” syllable type examples highlighted.

5.2.2 Recording

We recorded males at a former military training area in close proximity to Ceske Budejovice (100,000 inhabitants), South Bohemia, Czech Republic. The area stretches over 1 km² and consists of wooded marshland with ponds and stands of willow (*Salix* spp), birch (*Betula* spp), and aspen (*Populus* spp) trees, and some old oak (*Quercus* spp) avenues. At the edges of the area considerable traffic noise comes from the two busy roads which pass the territory. Chiffchaffs have optimal conditions here resulting in relatively high breeding densities. Each year, about 70-90 breeding pairs can be found in this area [LIN12b]. Rather than aiming for high quality recordings, we followed the recording procedure that could be easily applied at large scales by professional or amateur

ornithologists. We recorded 13 males between the 3rd and 30th of June, 2011. Songs were recorded over a considerable, two-day time span in order to have the individuals recorded with various background soundscapes (e.g. traffic noise, other birds singing, leaf rustling, etc.). We followed each focal male over four hours from 6:00 am to 10:00 am on one day and also for half an hour from 5:30 am to 6:00 am on the next day. We always tried to get as close as possible to the singing male, but the distance varied depending on the males' boldness and the habitat structure (e.g. tree height).

Recordings were made using Marantz PMD 660 solid state recorder and Sennheiser ME67 directional microphone equipped with the Rycote Softie windshield and sampled at 16 bit and 44.1 kHz. The recordings were down-sampled to 22.05 kHz before processing. We usually recorded within a distance of 5-15 meters from the singing male with no obstacles between the male and the microphone. Recorded material is summarized in Table 7.

Bird ID	Recording number	Recording quality (Poor, Average, Good)	Total length [min]	Min number of songs in one recording	Max number of songs in one recording	Total number of songs	Average number of songs per one minute of recording
A	12	G	50	13	89	487	10
B	4	P	26	37	125	329	13
C	12	A	43	4	48	332	8
D	10	A	38	3	39	226	6
E	9	P	39	10	70	170	4
F	22	G	108	3	90	788	7
G	11	G	80	22	103	546	7
H	13	P	68	7	99	599	9
I	8	A	99	6	200	713	7
J	9	P	48	7	63	304	6
K	4	G	22	3	52	98	5
L	9	A	43	6	88	293	7
M	4	G	37	24	115	221	6
Sum	127		698	-	-	5096	-
Avg	41860		53.7	41681	90.8	392.0	41705

Table 7: The Chiffchaff recording. The rating of quality is an aggregate value based on the subjective opinion of the operators based on coefficients: noise, masking by other birds, distance, and song clearness.

5.2.3 Recording quality

Besides the singing of a target bird, the recordings used contain many other unwanted sounds, as commonly found in nature, for instance: anthropogenic traffic noise: coming from neighbouring roads and urban areas; sounds of animals and other bird species: e.g. barking, meowing, calling and singing of other bird species; different chiffchaff individuals: can mask the song of the target bird; variable volume of singing: the level varies with the distance to the targeted male, with position (e.g. head turning) and with barriers; background noise: wood cracking when the ornithologist moves, leaf rustling, ornithologist's spoken commentaries, etc.

Because our recordings were obtained during a long period (almost one month) the amount of the abovementioned unwanted sounds and noise naturally differs record by record.

Previously described unwanted sounds contained in the recorded songs are common to the majority of field recordings. Therefore, the song and syllables are usually cut off from the recordings before use in experiments to eliminate most of the mentioned disturbances, and/or visual check is required [KOG98], [POT14], [VEN15].

The Figure 5.3 shows a spectrogram of a recording used in the experiment. It demonstrates the as-is quality, just as they were recorded by the ornithologists.

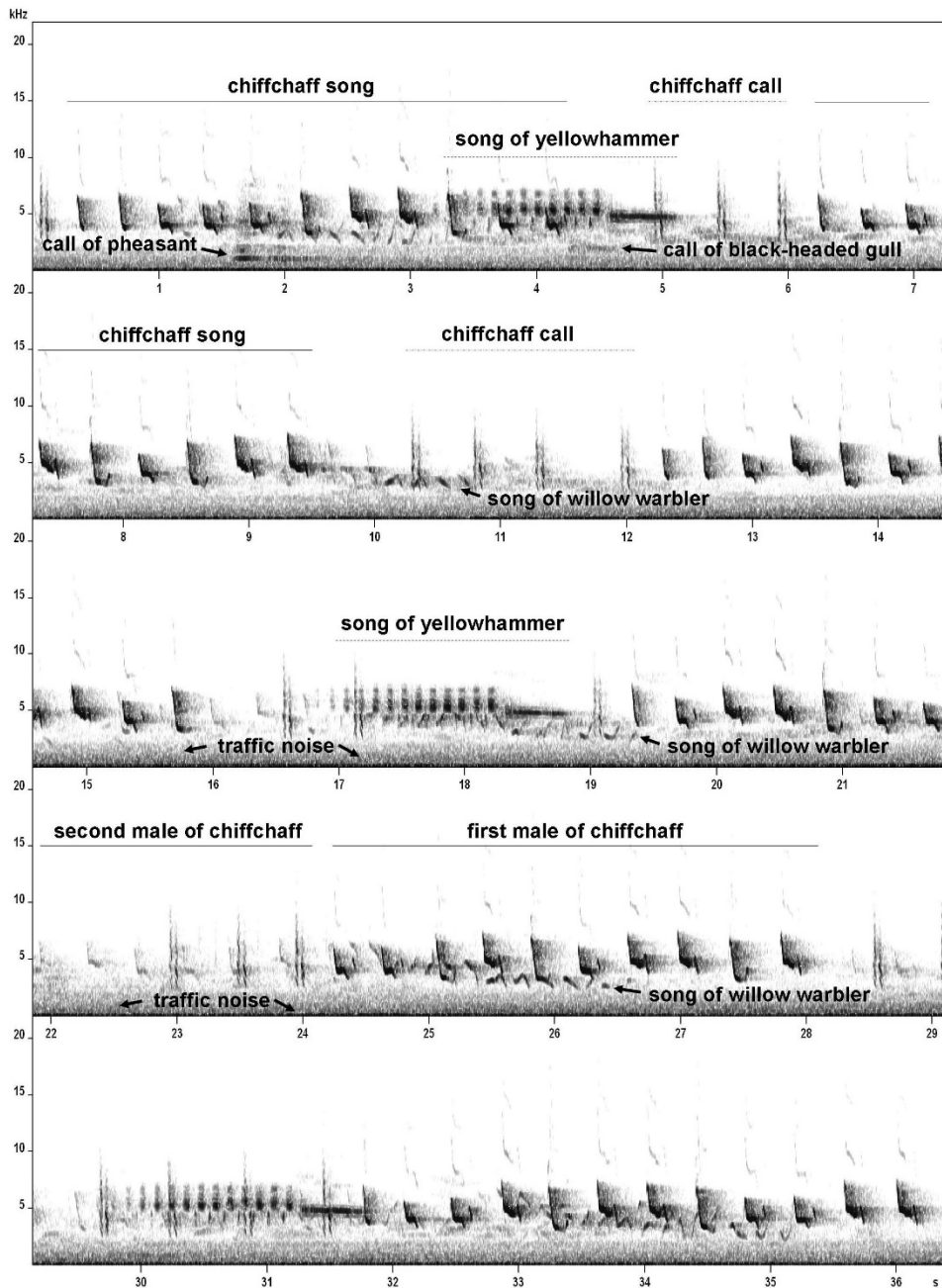


Figure 5.3: Spectrogram of real recording used for the experiments without any pre-processing (cut off songs, de-noising, etc.). The Chiffchaff song is masked by another male, different species, wind blowing noise, continuous traffic noise, etc.

5.3 Task definition

Based on the [BIM04] the task of bird (or speaker) recognition can be split into bird verification and identification. The technical solution depends whether the set of the individuals is open or closed. In the case of the **Closed set**, the unknown bird is assigned just to one of the trained GMM models where each represents a known bird. In the **Open set**, any number of new birds (without

corresponding trained GMM models) may appear anytime during the process. So, there is possibility that the unknown bird is a new one for which a GMM model does not exist.

Identification. In the identification scenario a set of reference birds with known identities is given. In the **closed set**, only one from the set of known individuals is selected based on the highest likelihood. The trained model with highest score is chosen. In the **open set**, either one of the trained birds is selected or a decision is made to investigate a new bird. This occurs if the likelihood does not exceed a threshold for any of known models. In most cases, the threshold is set by the user, and its value reflects the penalization for making errors.

Verification. The verification scenario is a one-to-one matching. Here, just one of the two decisions can be made: the bird either belongs to the compared model or not at all. As seen in Figure 5.4, the verification can be considered as a special case of identification on the open set where just one known bird is trained. Relating the identification on the open set, the unknown individual belongs to the one of known birds (models GMM 1 ÷ GMM 4) or to a new identity (UBM model). The model with highest likelihood is selected: any of the GMMs or UBM respectively. Verification, a one-to-one comparison, represented by just the one trial. Notice the similarity with identification which can be split into N trials. Basically $N = n + 1$ where n is a number of a trained birds, and one more trial is needed to calculate a new identity. In real experiment the N is much higher than n .

Essentially the identification task consists of n verifications, so called *trials* where n is a number of trained birds. Total number of trials N is required to accuracy score calculation, see equation (67).

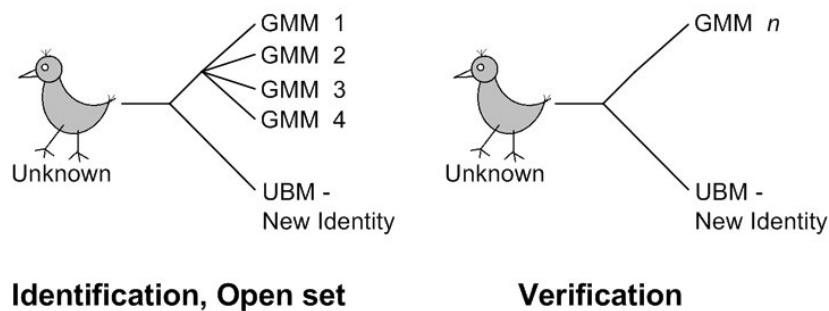


Figure 5.4: Identification and verification.

From the technical point of view, the processing of the closed set is easier than an open set, and it usually also gives better results. Using the closed set is usual for laboratory tests or in cases when the ornithologist can be sure of the number of singing birds. On the contrary, the open set reflects situations common in the nature. Obviously, it gives lower accuracy because it includes the uncertainty of a new individual appearance.

The main problem of individual identification on the open set is recognizing a new singer. Although, GMM models of all singers are known and trained, a GMM model of a new individual is evidently unknown. Thus, it is essential to use the UBM because it represents a model of the background: unknown singers, background noise, channel influence, etc.

5.4 System description

The GMM-UBM method described here is adopted from the well-known Speaker Recognition task, which is ordinarily used in human speech, as well as in animal recognition research. The introduced

system is tailored for individual identification on the open set, even when using non pre-processed recordings. The system performs identification with a sequence of particular verification trials. It is decomposed into: Parametrization, Voice Activity Detection (VAD), Model estimation, UBM model training, GMM model training, and Identification/Verification. For whole process see Figure 3.21.

An advanced VAD algorithm was designed and evaluated to ensure only the frames contain a song are using for the parametrization. As-is bird song recordings vary enormously in quality and therefore both short- and long-term energy parameters are estimated. A VAD's decision (signal/noise) follows from adapting VAD to particular record quality. Figure 5.5 shows an outline of the VAD, and Figure 5.6 demonstrates detection result on an as-is record. The global (whole recording) Signal-to-Noise Ratio (SNR) is firstly adapted from *i*-th frame SNR which follows from particular filters SNR estimations (so-called *local estimations*). The decision process is then adapted from both the frames- and the global- SNR estimations. This is crucial to balancing the global SNR to recording quality. A frame is labelled as a song if it's Signal-to-Noise Ratio (SNR) is higher than an overall SNR.

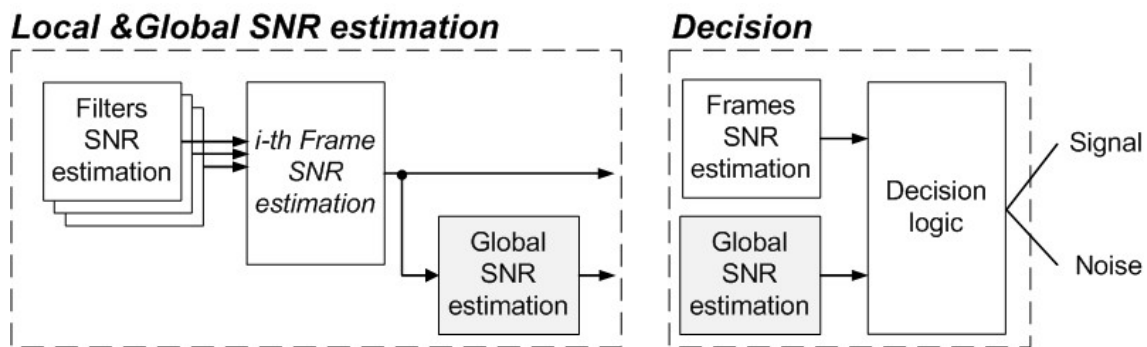


Figure 5.5: Outline of the VAD detector.

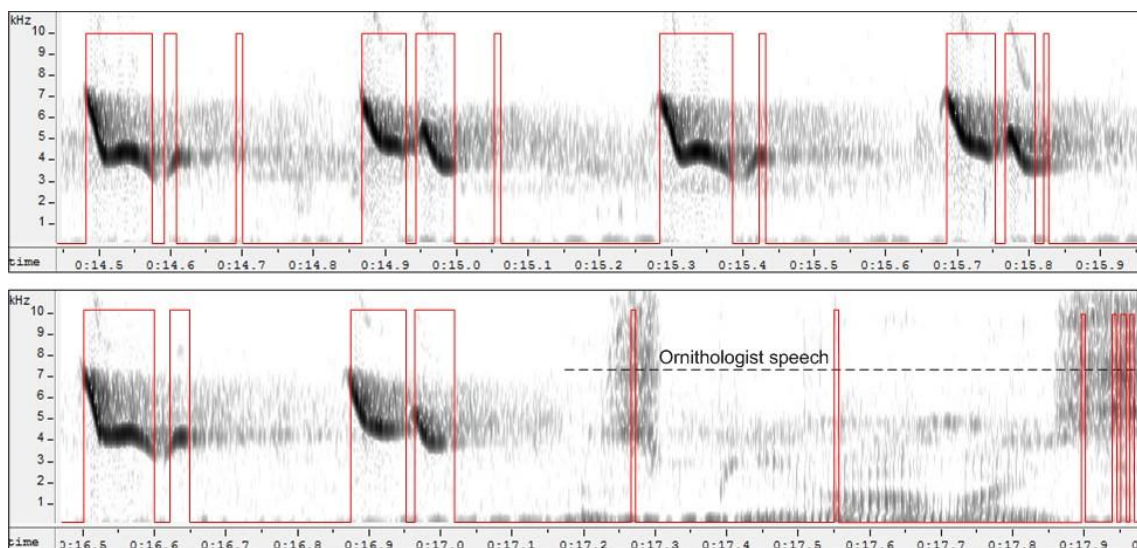


Figure 5.6: Spectrogram of a recording and a result of VAD. See false songs detections at 0:14.7, 0:15.05 and a segment containing an ornithologist speech (from 0:17.2 to 0:18.0).

VAD is based on detection of energies in both time (frame-by-frame) and frequency (filter outputs) domains while the result is given by their merging. At first a noise level estimate a_{ij} for each frequency filter is included where i represents frame index and j is a frequency filter. This estimate

a_{ij} is independently adapted for each frame just when an actual signal energy x_{ij} is not higher than a relative threshold an empirical constant β . The adaptation is then realized by so-called exponential forgetting function as

$$a_j = \alpha \cdot a_j + (1 - \alpha) \cdot x_{ij} \quad (65)$$

where α is an empirical constant which controls dynamic of the adaptation. We discovered that $\alpha = 0.94$ and $\beta = (2 \cdot a_{ij})$ gives the best results for the bird records.

Then the local SNRs (Signal to Noise Ratio) are estimated for i -th frame as a mean value of SNRs in each of the filter-banks as

$$SNR_i = \frac{1}{M} \sum_{j=1}^M 20 \cdot \log_{10} \frac{x_{ij}}{a_j} \quad (66)$$

where M is the number of filters.

Finally the VAD decision process compares the SNR of i -th frame to a global SNR. The global SNR is represented by a mean value of local SNRs computed across the entire recording. A frame is marked as non-song if the SNR of j -th frame is lower than the global SNR and vice versa. For feature processing with inbuilt VAD see Figure 5.7.

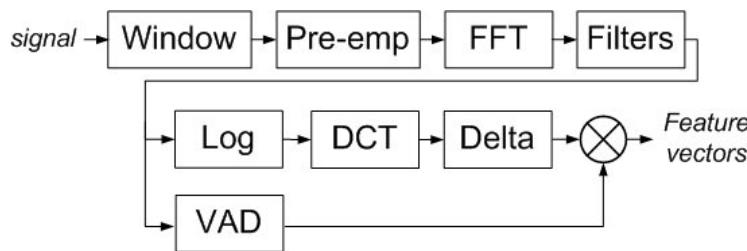


Figure 5.7: Parametrization of the recordings. The output parameters are formed into feature vectors.

5.5 Experiment evaluation

As explained above the experiment was conducted under the *content independent bird identification on the open set* scenario. It was performed by a sequence of particular verification trials. The main advantage of this approach is that we get not only the overall accuracy but also the particular accuracy for each single model. Thus, the results can be better understood and explored in detail.

For testing the system the three experiments were prepared. The recordings were randomly sorted into halves. The first experiment utilized the first half, the second experiment used the rest. Finally, the third experiment used all of the records.

It is crucial for each experiment to separate recordings into the three strictly distinguished sets:

1. **Training.** The set is used for computing GMM model(s) of known bird(s).
2. **UBM.** Used for computing the UBM model. Contains as many bird recordings as possible.

3. **Testing.** Contains unknown birds which validation is supposed to be tested. During the system design phase the set contains both unknown and trained bird records (notice, the particular recordings used in the Training set must not be used).

After the three sets of recordings are prepared the testing trials must be matched. They consist of recording pairs where the first element is selected from the Training set, and the second from the Testing set. During trial performances the testing pairs are compared and the results are obtained.

If there are, for instance, 20 records for the Training set, and 50 records for the Testing set (say, 40 of them belong to unknown birds and 10 to trained), then $N=1,000$ trials are preformed ($=20*50$) where 200 are true trials ($=20*10$) and 800 are impostor trials ($=20*40$). For accuracy calculation see equation (5).

We distinguish two types of error: False Acceptance (FA) and False Rejection (FR). The first type occurs when an unknown bird is evaluated as identical to the tested bird when, actually, they are different. The FR occurs when the identical birds are not recognized, see Figure 5.8.

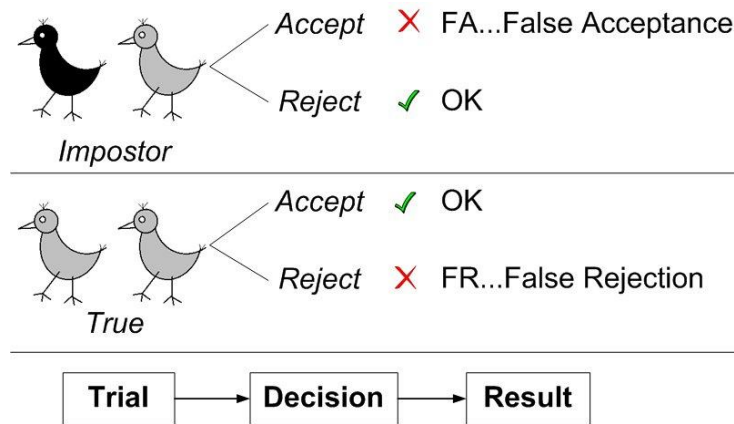


Figure 5.8: Two types of errors: False Acceptance and False Rejection.

The total number of errors is the sum of both FR and FA. The accuracy ranges from 0% to 100% and it is calculated as

$$Accuracy = \left(1 - \frac{N_{FA} + N_{FR}}{N}\right) \cdot 100 \quad (67)$$

where N is the total number of trials, N_{FA} is the number of the FA, and N_{FR} is the number of FR. These errors depend on the choice of the verification threshold.

5.5.1 Parameters

At first, the suitable parameters were analysed. For example, Table 8 shows two of the parameters, *Hamming window length* and *Shift Window iteration*, used in this analysis. Each combination of *Hamming window length* and *Shift Window* parameters was tested by seven different randomly selected data-sets. Because of high among of results rather than state concrete numbers the accuracy was categorized into groups: high (90-100%), mid (80-90%), low (70-80%), very low (<70%). The fifth parameter settings up was finally selected for its suitable ratio between accuracy and computer performance.

Other parameters were explored similarly. Table 9 gives the final parameters set up.

Iteration	1	2	3	4	5	6	7	8	9	10	11	12
Window length [ms]	20	20	10	10	30	30	40	50	50	80	200	400
Shift Window [ms]	20	10	5	10	15	30	20	25	50	40	100	200
Success rate	high	high	low	mid	high	mid	high	mid	low	low	low	very low

Table 8: Parameter iteration example.

Parameter	Value
Window type	Hamming
Window length	30 ms
Window overlap	15 ms
Scale of triangular filters	linear
Number of filters	25
Number of cepstral coefficients	20
Compute zero coefficient logE	yes
High pass filter	2.5 kHz
Low pass filter	not used
Delta coefficients	yes
VAD detector	yes
Preemphasis	0.97
Linear/MelFilter scale	Linear

Table 9: Parametrization set up values.

The low-pass filter was set to 2.5 kHz because of the bandwidth of the chiffchaff song. The low-pass filter was not used. Since the delta coefficients were also extracted, the number of dimensions of feature vector was 40 (2 times number of cepstral coefficients). A linear distribution of filters was chosen as a basic approach. This is in accordance with experiments conducted in [GRA11], where no significant differences in the performance of an automatic bird recognition system were observed when utilizing linear and Mel scale.

5.6 Results

The main result of the experiments is given in Table 10 where 16,480 trials were performed. First two experiments used a different half of the data, in the third experiment, all available records were used. General accuracy of identification across all experiments is 78.5%.

	Round 1	Round 2	Round 3	Overall
# trials	3975	4265	8240	16480

# errors	754	1011	1782	3547
general accuracy	81.0%	76.3%	78.4%	78.5%

Table 10: General experiment result for all three experiments.

The detailed result of the experiments is given in Table 11 and visualised in the graph, see Figure 5.9. Each column of the table represents the identification accuracy which belongs to a particular bird. The highest variance between maximum and minimum accuracy belongs to the birds B, D, and E with 17.5%, 18.1%, and 20.8%, respectively. On the contrary, the birds A, C, F, G, I, J, and M have lowest variance from 0.3% (bird C) to 3.8% (bird G). The birds H, K, and L have variance 11.3%, 7.7%, and 9.8%.

Bird	A	B	C	D	E	F	G	H	I	J	K	L	M	Overall
Round 1														
# trials	324	124	324	325	215	781	324	462	224	320	116	320	116	3975
# errors	76	35	66	70	35	147	14	81	48	65	19	81	17	754
Accuracy	76.5%	71.8%	79.6%	78.5%	83.7%	81.2%	95.7%	82.5%	78.6%	79.7%	83.6%	74.7%	85.3%	81.0%
Round 2														
# trials	438	130	324	310	415	616	432	420	284	310	138	310	138	4265
# errors	100	14	65	123	154	138	35	121	64	56	12	109	20	1011
Accuracy	77.2%	89.2%	79.9%	60.3%	62.9%	77.6%	91.9%	71.2%	77.5%	81.9%	91.3%	64.8%	85.5%	76.3%
Round 3														
# trials	762	254	648	635	630	1397	756	882	508	630	254	630	254	8240
# errors	178	50	132	196	185	292	49	207	111	119	31	191	41	1782
Accuracy	76.6%	80.3%	79.6%	69.1%	70.6%	79.1%	93.5%	76.5%	78.1%	81.1%	87.8%	69.7%	83.9%	78.4%

Table 11: Detailed result. Accuracy for particular bird. The lowest value is 60.3% (round 2, bird D) the highest 95.7% (round 1, bird G).

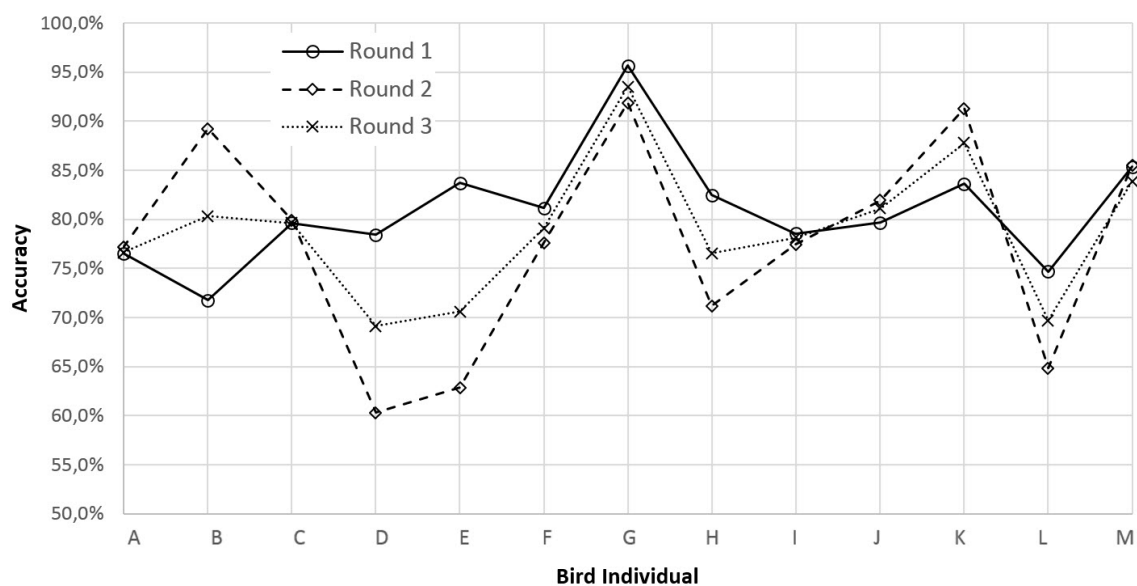


Figure 5.9: Detail result. Accuracy for each particular bird, see Table 5.

The Table 12 gives another view to the experiments. It summarized results from all three rounds, moreover it reveals both the FA, and FR errors. See the true trials accuracy of the bird “I” which gives 46.9%; it is the lowest result. In opposite the false trials accuracy of the bird “G” reaches 96.1% which is the higher obtained result.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Over all
# true trials	146	16	144	100	80	484	120	168	64	80	16	80	16	1514
# FR errors	30	4	14	24	24	101	44	41	34	32	2	16	4	369
accuracy true trials	79.5 %	75.0 %	90.3 %	76.0 %	70.0 %	79.1 %	63.3 %	75.6 %	46.9 %	60.0 %	87.5 %	80.0 %	75.0 %	75.6 %
# false trials	1378	492	1152	1170	1180	2310	1392	1596	952	1180	492	1180	492	1496 6
# FA errors	324	95	249	365	350	476	54	368	189	208	60	365	74	3178
accuracy false trials	76.5 %	80.7 %	78.4 %	68.8 %	70.3 %	79.4 %	96.1 %	76.9 %	80.1 %	82.4 %	87.8 %	69.1 %	85.0 %	78.8 %
# sum trials	1524	508	1296	1270	1260	2794	1512	1764	1016	1260	508	1260	508	1648 0
# sum errors	354	99	263	389	374	577	98	409	223	240	62	381	78	3547
general accuracy	76.8 %	80.5 %	79.7 %	69.4 %	70.3 %	79.3 %	93.5 %	76.8 %	78.1 %	81.0 %	87.8 %	69.8 %	84.6 %	78.5 %

Table 12: Results summary. Includes a data from all experiments, and reveal the FA and FR errors in detail.

Distribution of results into accuracy levels is given in Table 13. Thus, it follows that 90% of experiments have accuracy higher than 70%, and 51% experiments even higher than 80%.

Result category	Number of results	Percentage
60-70%	5	13%
70-80%	18	46%
80-90%	12	31%
90-100%	4	10%
Total	39	

Table 13: Distribution of the experiment results. The results were first rounded and then assigned to a particular level.

Finally, Table 14 shows the dependence of the accuracy on the both number and quality of songs of a bird. It combines the data from Table 7 (recording description) and experimental results.

Success rate	60% - 74%	75% - 84%	>85%
Number of songs	170 to 293	304 to 788	98 to 546
Quality of recordings (Poor, Average, Good)	P,A	P,A,G	G

Table 14: Number of songs in dependence on the accuracy and the recording's quality.

5.7 Contribution

A GMM-UBM based Automatic System for Recognition of Bird Individuals (ASRBI) was described with the added VAD algorithm. The aim of our work was to show that the individual identification on the open set using real recordings without pre-processing is feasible, and could be used by ornithologists under real conditions. The use of the UBM is of great importance since it identifies the environmental conditions of the recognition task, calibrates the verification score, and facilitates the choice of the verification threshold yielding a superior performance of the ASRBI. To our knowledge, this is the first experiment dealing with bird individual identification in real, open set conditions and, moreover, on real recordings without any pre-processing. The research is described in more detail in [PTA15a].

5.8 Summary

We report a method for bird individual identification, content-independent, working on the open set, processing as-is (raw, long real-field) recordings. The method can be used in real conditions, e.g. situations when ornithologists have targeted recordings of individuals (i.e. they recorded one individual for some time) from a population and they would like to know if the new recordings belong to birds previously recorded (and what is the birds identity in that case) or if it is a new bird. This approach modifies the traditional capture-mark-recapture approach but without the necessity of capturing and marking the subjects.

Our work was motivated primarily by the need of ornithologists, from the *Institute of Animal Science and University of South Bohemia, Department of Zoology in Czech Republic*, who were looking for a system that could replace the ringing of chiffchaffs by some other less invasive technique. Achieved accuracy is 78.6% and is understandable due to the more realistic recording setup. Accuracy improves substantially if only the songs with the best recording quality are used.

We are aware of the fact that the chiffchaff song represents rather ordinary singing style and the methods should be evaluated in species with more varied songs as well. However, chiffchaffs apparently change more call types within one recording. Therefore, several song types can be present in the training set as well as in the test set. Our approach mimicked the task expected in real content-independent situations, without any prior knowledge about the number of song types (and their prior classification).

The question of high importance is how many recordings are needed for reasonable accuracy. The result presented here reveals that recognition, even for birds with lower number of recorded songs (the case of the birds K and M), can be made with very good accuracy. From this question originates the idea to split recording into the halves to have the possibility of measuring accuracy independently for each half as well as for all the data in the third experiment. The results cue the amount of data should not be a crucial problem.

The identification accuracy not just depends on data quality and information content. The successful UBM has to cover a wide background not just contain as much records is possible. Upon evaluation of our results, it is our opinion that the accuracy of the experiment depends also upon the UBM set. It is also confirmed by our current experiments which are not described in this paper as they are still in progress.

6 Identification Vectors

6.1 Introduction

As stated in section 3.4.11 iVectors are State-of-the-Art method in Speaker Recognition. Because we had not enough of bird recordings to train the system, we used a system trained on the human-speech data provided by National Institute of Standards and Technology (NIST) in the experiment. The main scope of the evaluation was to prove iVectors can be used for Bird Individual Identification on the Closed Set.

6.2 Experiment evaluation

We used 5,176 bird song records from thirteen chiffchaff individuals. The songs were cut from the raw recordings, described in section 0; for examples, see the figures below. We decided just for basic parametrization optimization to consider the main experimental scope. Then the recording preparation involved meaning sorting and labelling. Moreover, we provided a visual check and listening of the training data.

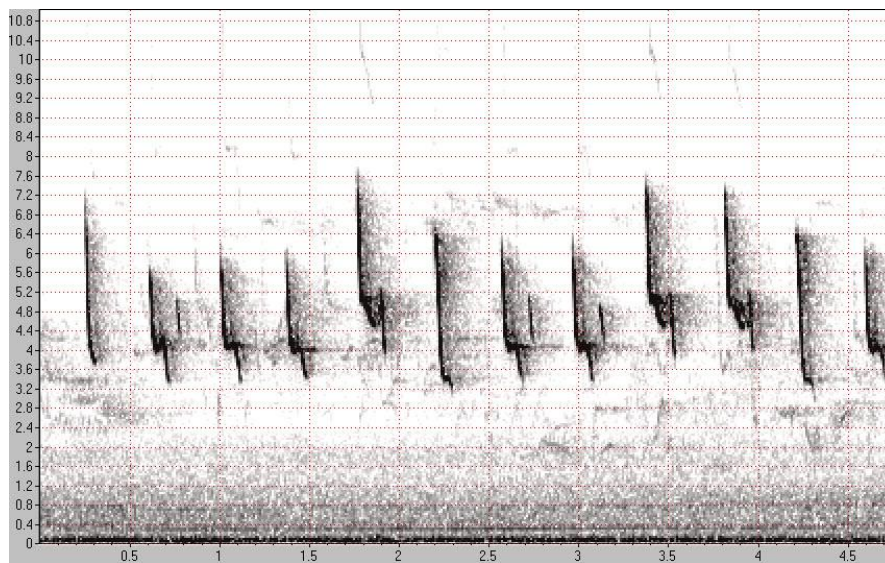


Figure 6.1: An example of a song extracted from the raw recordings. Low overlap level, standard noise, duration 4.8 sec.

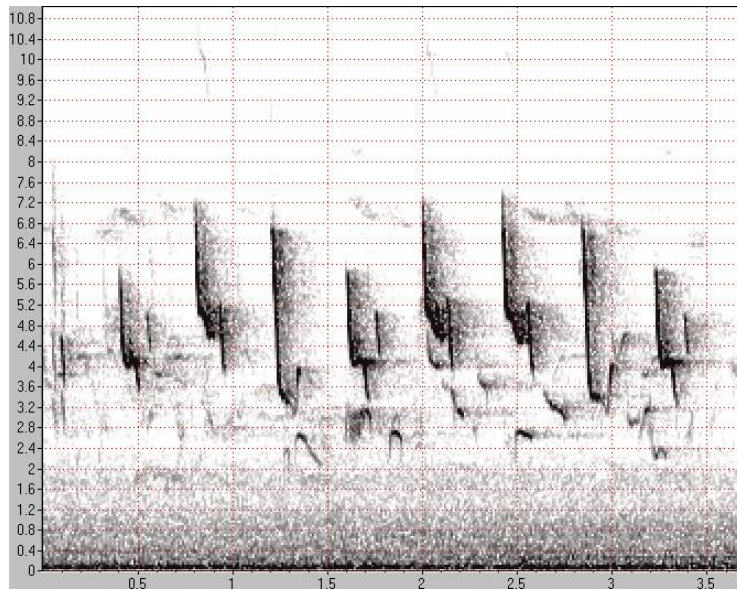


Figure 6.2: An example of a song extracted from the raw recordings. High overlap level, high noise, duration 3.7 sec.

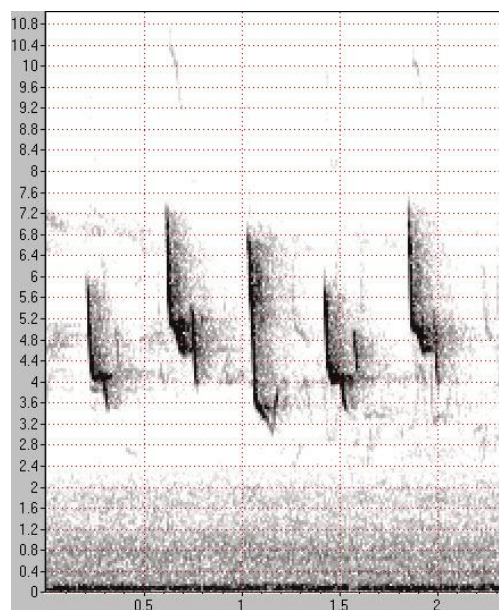


Figure 6.3: An example of a song extracted from the raw recordings. Low overlap level, low noise, duration 2.5 sec.

6.3 Results

Table 15 gives the experiment results. The birds in the first line (column heading) represent those for which the system was trained. Columns contain the corresponding probability of a particular bird. The grey coloured cells have the expected highest score (intersection of the column bird and the line bird). An orange colour highlights the highest score of incorrectly identified birds; for instance, bird F was identified as a bird D with highest score 0.745, whereas the true cell (grey coloured intersection cell of Bird F column and Bird F line) gives just a second highest score (0.591). The identification accuracy varied between 61.9% and 85.8%.

	Bird A	Bird B	Bird C	Bird D	Bird E	Bird F	Bird G	Bird H	Bird I	Bird J	Bird K	Bird L	Bird M
Bird A	0.802	0.388	0.482	0.632	0.524	0.482	0.434	0.516	0.371	0.337	0.571	0.465	0.481
Bird B	0.364	0.795	0.318	0.543	0.421	0.440	0.286	0.369	0.329	0.185	0.308	0.294	0.386
Bird C	0.446	0.400	0.633	0.507	0.548	0.495	0.474	0.307	0.363	0.396	0.496	0.599	0.385
Bird D	0.704	0.644	0.598	0.858	0.734	0.745	0.574	0.632	0.476	0.408	0.622	0.699	0.576
Bird E	0.459	0.320	0.504	0.394	0.797	0.422	0.350	0.478	0.263	0.296	0.639	0.501	0.318
Bird F	0.335	0.238	0.284	0.452	0.275	0.591	0.374	0.267	0.236	0.215	0.212	0.301	0.271
Bird G	0.230	0.064	0.164	0.113	0.176	0.168	0.476	0.147	0.146	0.182	0.257	0.224	0.221
Bird H	0.568	0.369	0.381	0.645	0.399	0.456	0.398	0.564	0.336	0.218	0.316	0.432	0.400
Bird I	0.275	0.152	0.205	0.192	0.244	0.180	0.215	0.183	0.216	0.206	0.320	0.212	0.219
Bird J	0.517	0.271	0.496	0.422	0.557	0.497	0.513	0.344	0.341	0.644	0.560	0.550	0.525
Bird K	0.548	0.139	0.452	0.200	0.620	0.205	0.369	0.333	0.180	0.317	0.848	0.444	0.287
Bird L	0.556	0.346	0.609	0.571	0.697	0.529	0.588	0.440	0.328	0.497	0.646	0.856	0.536
Bird M	0.605	0.293	0.491	0.418	0.479	0.392	0.554	0.403	0.407	0.448	0.579	0.535	0.619

Table 15: iVectors confusion matrix.

6.4 Contribution

Based on our knowledge these are the first experiments regard bird individual identification by iVectors. Though the scope of the experiments was small, we proved our ability to identify iVectors by T (Total Variability Matrix), estimated not by animal data but speech (human speech and channels matrices). We introduced and discussed the research in [PTA15b].

6.5 Summary

The experiment tested the identification of thirteen bird individuals. The system correctly identified nine birds of the thirteen bird (i.e. 69.2%), and for four birds an error occurred (30.8%). However, three of these four errors occurred with the bird D data (line four, birds F, H, and I). If, hypothetically, we do not involve the bird D into experiments, just two birds would be identified incorrectly (Bird G, and Bird I).

Because iVectors promise a high potential for individual identification, we are considering to use iVectors in the future in the next stages of our research.

7 Bird Audiogram Unified Equation

7.1 Introduction

While working with an automated system for an individual or species recognition one should use a feature extraction matching the bird's vocalization, particularly the signal filtering which precede the parametrization. There are two basic types of filter arrangements. The MFCC is used in many papers focused on this topic ([TRA05], [CLE05], [FOX08], [TRI08], [CHE10]). Another possibility is to use the linear distribution. MFCC corresponds with the human hearing properties [BIM04], and comparison of the linear with MFCC arrangements reveals the similar results for bird [GRA11], [PTA15a].

The original plan of our work was to create a bank of filters adapted to bird songs. During the design, we took into consideration the research of available bird audiograms. After preliminary tasks, we had decided to divide our research into two parts because of amount of work for each part. The first part, described in this chapter, deals with bird audiogram. The second part, described in chapter 8, deals just with a bank of filters adapted to bird songs based on bird audiograms.

Currently, there are audiograms available for about 60 bird species ([HEF98], [DOO02b], [KON70], [OKA85], [CAT08], [LAU07], [MAR04]). The audiograms' common characteristic is that they were measured only for a small amount of frequencies with a small sample of individuals. For clarity, these audiograms are usually illustrated in a graphic form, with lines connecting the measured points. However, a chart containing measured values is often missing.

7.2 Audiogram equation definition

We propose five types of functions, each one of which may be selected based upon how well it fits available audiograms [PTA16]. Only a few articles directly address bird audiograms, e.g. [HEF98]. Unfortunately, these papers often reference other articles, which are not available (origin from seventies, sixties, even fifties). Finally, [DOO02b] was selected as a principal article not only because it collects many audiograms in a unified form, but the author, Prof. Dooling, is known as a pioneer and one of the highest authorities in bird hearing research. The article [DOO02b] contains 47 audiograms². Notice we use prof. Dooling's short cuts (B-04, B05, etc.) when specifying an audiogram from [DOO02b].

The goal was to discover just the one function (i.e. equation) for known bird audiograms in order to allow its implementation in automated systems. Notice we call this equation as *Audiogram Unified Equation* (AUE). Then, only setup of parameters belonging to a particular bird is necessary for use of the AUE in ARSBI or ARSBS.

First, we extracted from the original paper [DOO02b] the datasets describing individual curves. Notice, all data was available in PDF graphical format no source text is available. See an example of original data at the Figure 7.1, audiogram of Emu (*Dromaius novaehollandiae*).

² We discovered that two pairs are similar: B-05 (*Pedionomus torquatus*) and B-06 (*Columbia livia*); B-08 (*Accipiter nisus*) and B-09 (*Colinus virginianus*). We sent an email to prof. Dooling to be assured the pairs similarity is a mistake or not, but we have received no answer until publishing of this work.

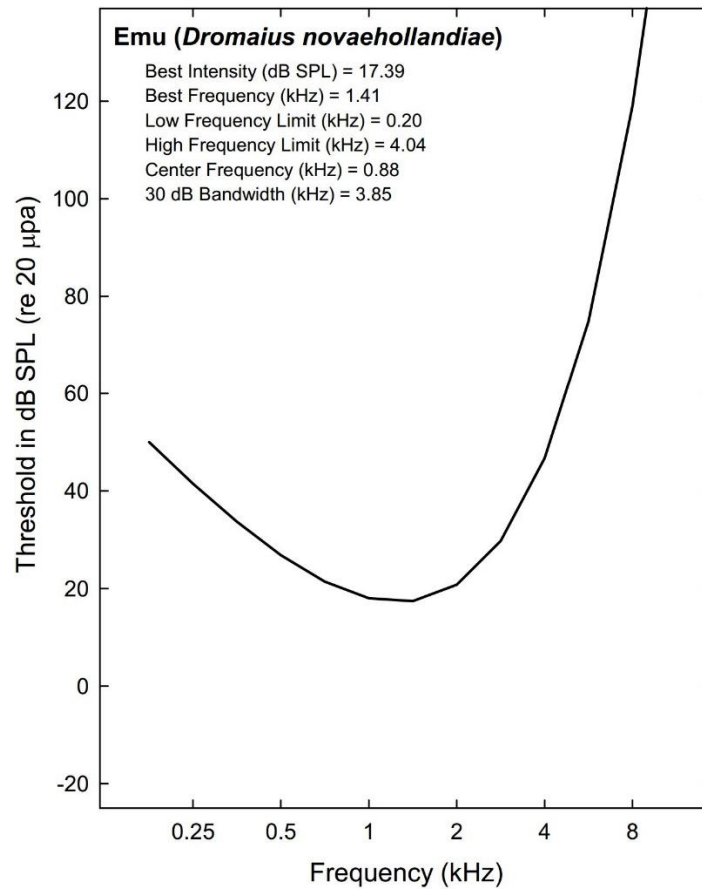


Figure 7.1: Original data example: Audiogram, B-04, Emu (*Dromaius novaehollandiae*) [DOO02b].

Table 16 summarizes basic parameters of the audiograms. For an explanation of these parameters, see Figure 7.2.

#	Name	Genus and Species	BI [dB SPL]	BF [kHz]	LF [kHz]	HF [kHz]	CF [kHz]	30dB [kHz]
1	Mallard Duck	<i>Anas platyrhynchos</i>	15.57	2.00	0.32	5.22	1.28	4.91
2	Australian Grey Swiftlet	<i>Collocalia spodiopygia</i>	20.31	2.00	0.49	5.71	1.66	5.23
3	Oilbird	<i>Steatornis caripensis</i>	20.31	2.00	0.49	5.71	1.66	5.23
4	Emu	<i>Dromaius novaehollandiae</i>	17.39	1.41	0.20	4.04	0.88	3.85
5	Plains Wanderer	<i>Pedionomus torquatus</i>	33.80	0.71	0.05	3.56	0.44	3.50
6	Pigeon	<i>Columbia livia</i>	16.90	1.41	5.67	0.13	5.80	5.67
7	American Kestrel	<i>Falco sparverius</i>	2.42	2.00	0.36	5.25	1.37	4.89
8	European Sparrowhawk	<i>Accipiter nisus</i>	4.27	2.00	0.35	5.39	1.37	5.04
9	Bobwhite Quail	<i>Colinus virginianus</i>	13.15	2.00	2.13	8.70	1.35	6.57
10	Chicken	<i>Gallus gallus</i>	7.37	1.41	0.20	4.10	0.91	3.90
11	Japanese Quail	<i>Coturnix coturnix japonica</i>	1.40	2.00	0.47	5.90	1.66	5.43
12	Turkey	<i>Meleagris gallopavo</i>	15.43	2.00	0.29	5.25	1.22	4.96
13	American Robin	<i>Turdus migratorius</i>	7.49	2.83	0.34	8.73	1.72	8.39
14	Blue Jay	<i>Cyanocitta cristata</i>	14.46	2.00	0.28	6.31	1.33	6.03

15	Brown-headed Cowbird	<i>Molothrus ater</i>	11.50	2.83	0.35	8.50	1.72	8.15
16	Bullfinch	<i>Pyrrhula pyrrhula</i>	-0.50	2.83	0.48	10.20	2.21	9.72
17	Chipping Sparrow	<i>Spizella passerina</i>	2.06	4.00	0.59	12.90	2.75	12.31
18	Common Canary	<i>Serinus canarius</i>	15.98	2.83	0.47	9.37	2.08	8.90
19	Common Crow	<i>Corvus brachyrhynchos</i>	-16.41	2.00	0.47	4.57	1.46	4.10
20	European Starling	<i>Sturnus vulgaris</i>	8.00	2.00	0.23	6.43	1.20	6.20
21	Field Sparrow	<i>Spizella pusilla</i>	9.61	2.83	0.32	8.65	1.65	8.33
22	Fire finch	<i>Lagonosticta senegala</i>	10.89	2.00	0.50	6.49	1.79	5.99
23	Great tit	<i>Parus major</i>	3.07	2.00	0.32	8.17	1.60	5.02
24	House finch	<i>Carpodacus mexicanus</i>	14.55	2.00	0.44	6.00	1.61	5.56
25	House Sparrow	<i>Passer domesticus</i>	-8.31	1.41	0.29	4.55	1.13	4.27
26	Pied Flycatcher	<i>Ficedula hypoleuca</i>	11.70	2.83	0.44	7.34	1.79	6.90
27	Red-winged Blackbird	<i>Agelaius phoeniceus</i>	11.85	2.83	0.33	8.20	1.64	7.87
28	Slate-colored Junco	<i>Junco hyemalis</i>	-5.29	2.83	0.68	8.25	2.36	7.57
29	Song Sparrow	<i>Melospiza melodia</i>	4.98	2.83	0.33	8.76	1.69	8.43
30	Swamp Sparrow	<i>Melospiza georgiana</i>	6.05	2.83	0.37	9.00	1.82	8.63
31	Western Meadowlark	<i>Sturnella neglecta</i>	-3.06	2.00	0.37	5.55	1.42	5.18
32	Zebra Finch	<i>Taeniopygia guttata</i>	17.98	2.83	0.44	8.24	1.89	7.81
33	Bourke's Parrot	<i>Neophema bourkii</i>	17.36	2.00	0.23	6.50	1.22	6.27
34	Budgerigar	<i>Melopsittacus undulatus</i>	0.80	2.00	0.36	5.97	1.45	5.62
35	Cockatiel	<i>Nymphicus hollandicus</i>	8.60	1.41	0.22	5.22	1.08	5.00
36	African Wood Owl	<i>Strix woodfordii</i>	-13.40	2.40	0.33	7.90	1.60	7.58
37	Barn Owl	<i>Tyto alba</i>	-16.20	2.83	0.32	12.00	1.95	11.68
38	Brown Fish Owl	<i>Ketupa zeylonensis</i>	-1.60	1.00	0.08	4.00	0.57	3.92
39	Eagle Owl	<i>Bubo bubo</i>	-23.48	2.00	0.21	6.52	1.18	6.31
40	Great Horned Owl	<i>Bubo virginianus</i>	4.31	0.71	0.03	4.15	0.35	4.12
41	Long Eared Owl	<i>Asio otus</i>	-25.05	2.83	0.41	8.06	1.81	7.65
42	Mottled Owl	<i>Strix virgata</i>	-9.54	1.41	0.06	8.20	0.72	8.14
43	Scops Owl	<i>Otus scops</i>	-14.29	2.00	0.34	6.65	1.50	6.31
44	Snowy Owl	<i>Nyctea scandiaca</i>	-25.25	2.00	0.63	5.88	1.91	5.26
45	Spotted Wood Owl	<i>Strix seloputo</i>	-17.89	2.00	0.21	6.55	1.17	6.34
46	Tawny Owl	<i>Strix aluco</i>	-24.62	2.00	0.22	6.62	1.19	6.41
47	White-faced Scops Owl	<i>Otus leucotis</i>	-23.26	2.00	0.28	6.04	1.29	5.76

Table 16: Main parameters of 42 audiograms. Legend: BF (Best frequency) is the frequency with the best sensitivity BI (Best Intensity). LF (Low frequency) and HF (High frequency) define the bandwidth of an audiogram. CF (Center frequency) is the frequency in the middle of an audiogram. 30 dB defines the frequency an audiogram reaches 30 dB SPL sensitivity. For graphical legend, see Figure 7.2.

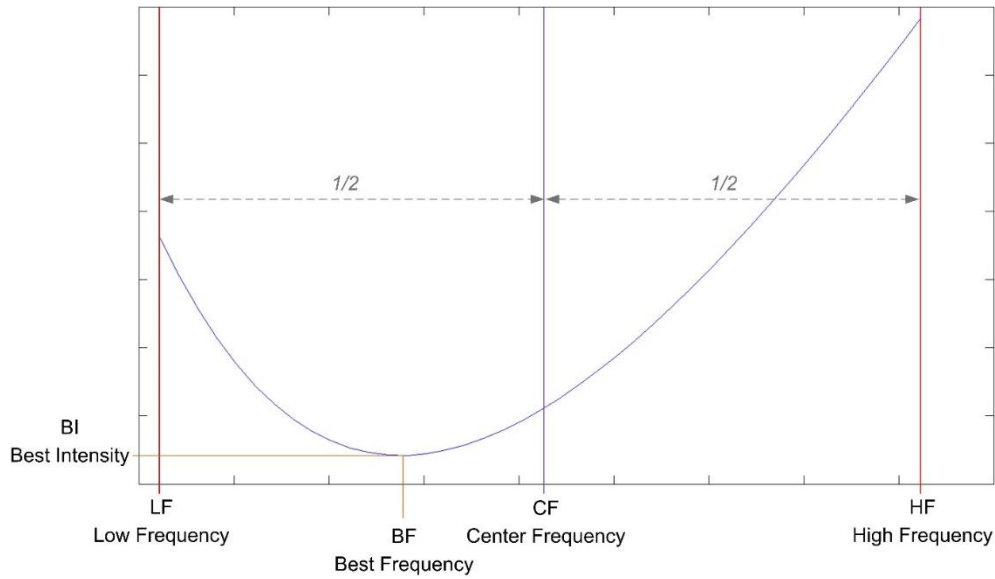


Figure 7.2: Table 16 graphical legend.

Each dataset was then fitted by the nonlinear least squares Marquardt-Levenberg algorithm (MLA) [MAR63] implemented in *GNUPlot* software (© 1986 Thomas Williams, Colin Kelley) via one of the five proposed functions. The goal was to select the one, which gives the best results. We proposed these five functions for the fitting:

$$f_1(x) = a \cdot e^{b \cdot x} + c \cdot e^{d \cdot x} + k \quad (68)$$

$$f_2(x) = a \cdot e^{b \cdot x} + k \quad (69)$$

$$f_3(x) = a \cdot x^3 + b \cdot x^2 + c \cdot x + k \quad (70)$$

$$f_4(x) = a \cdot x^4 + b \cdot x^3 + c \cdot x^2 + d \cdot x + k \quad (71)$$

$$f_5(x) = a \cdot x^5 + b \cdot x^4 + c \cdot x^3 + d \cdot x^2 + f \cdot x + k \quad (72)$$

Notice the preliminary functions, $\sin x$ and $\sin^2 x$, were also tested, but the fitting error was too high. Therefore, we decided do not involve them into the final pool.

Then we searched for the parameters a , b , c , d , f , and k by the fitting procedure. These steps were repeated for each function; for an example see Figure 7.3. The purple crosses represent localization points we put on the original audiogram. The colored line is the graph of f_i .

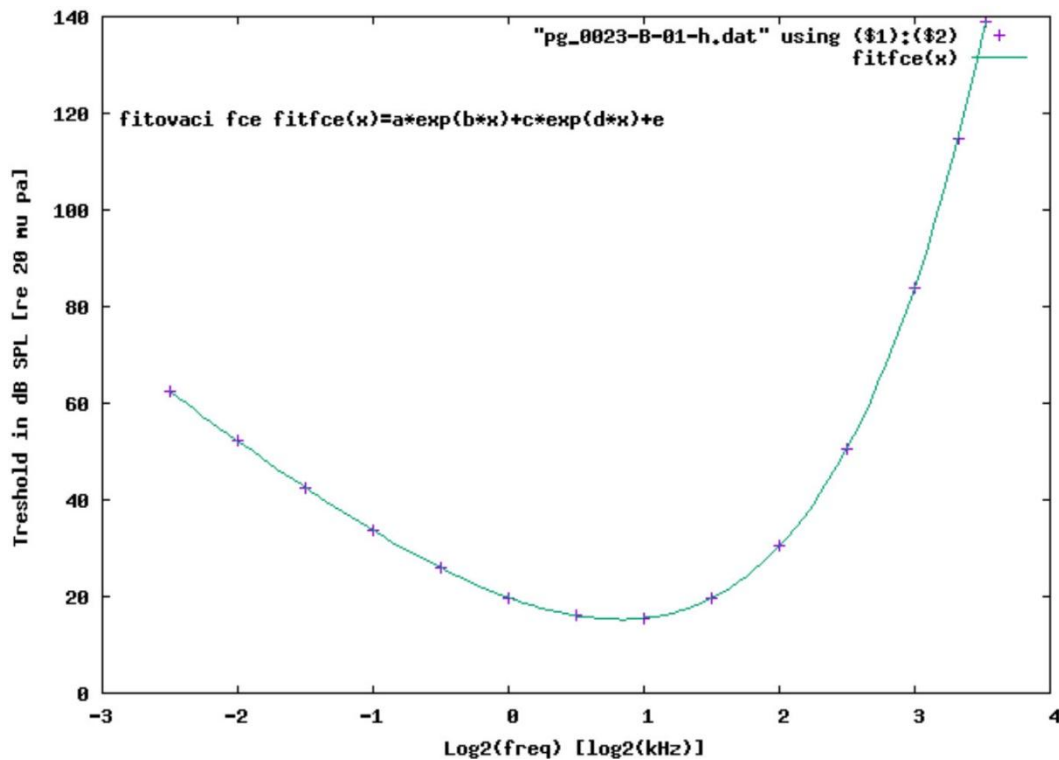


Figure 7.3: An example of the fitting (Mallard Duck).

The final sum of squares of residuals was minimal for the f_1 (exp-exp) function for all datasets (see Figure 7.4) with the exception of five ones (B-05, B-18, B-30, B-35 and B-41) for which the function f_5 was the best. On the other hand, the approximations via f_1 function consumed the longest computational time, as the number of approximation steps was the highest among all kinds of selected fitting functions. Approximation via f_1 led not only to the least sum of residuals after reaching the stop-criterion, which was the minimum relative change of the sum of residuals, but it also showed the largest relative change during the last iteration, which indicates that there is still some possibility to improve the final approximation of parameters we were looking for. On the contrary, functions f_3 - f_5 showed the relative change during the last iteration to be considerably small leading to the conclusion that there is not so much space for further improvement.

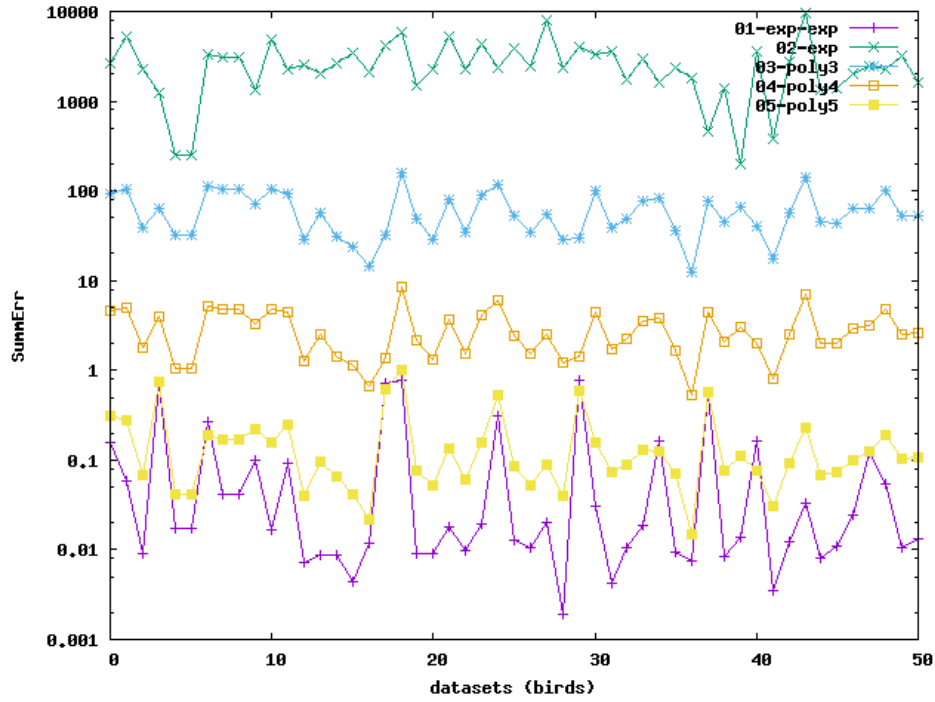


Figure 7.4: Final sum of squares of residuals for all five functions. Legend: f_1 Purple, f_2 Green, f_3 Blue, f_4 Orange, f_5 Yellow. Notice the yellow is even lower than purple for just five points.

Furthermore, we divided the datasets into three groups:

- Non-Passeriformes (B-01 up to B-12 and B-33 up to B-35),
- Passeriformes (B-13 up to B-32), and
- Strigiformes (B-36 up to B-47).

For each group we calculated the average coefficients, which was later fitted via the same functions f_1 to f_5 for the graphs of individual species. Moreover, we also calculated the averaging data for the data set, i.e. for all birds.

Let us remark, that the doubled values B-06, as well as B-09, were removed before we calculated the average threshold-values. Most of the original datasets were drawn to match the same values of frequency on the x-axis. For those 6 exceptional datasets, we interpolated and extrapolated the graphs (the inter- and extrapolation was performed via function f_1) in order to ultimately obtain the average threshold-values for the same frequency-values.

7.3 Result

We propose f_1 as the final, universal equation for bird audiograms (AUE) after considering the fitting errors:

$$f_1(x) = a \cdot e^{b \cdot x} + c \cdot e^{d \cdot x} + k. \quad (73)$$

Let us recall that the fitting procedure tries to minimize a sum of squares of residuals, i.e. the function

$$WSSR(f) = \sum_{n=1}^N (f(x_n) - y_n)^2 \quad (74)$$

by finding “optimal values” of all parameters appearing in the given function f . Here, (x_n, y_n) , where $n=1$ up to N , are the points being fitted, and the weights of all points are equal to one. Since the data show a non-linear dependence, we used a non-linear fitting method. Therefore, we obtained only an approximation of the optimal parameter values via a step-by-step iteration process. The stop-criterion remained the implemented the default, i.e. when the sum of squared residuals changes between two successive iteration steps by a factor less than $1e-5$, the fit is considered to have 'converged'.

The fitting errors are defined

$$rms = \sqrt{\frac{WSSR}{ndf}}, \quad (75)$$

and

$$rms_{VAR} = \frac{WSSR}{ndf}, \quad (76)$$

where rms represents the remains of residuals, rms_{VAR} is a variance of residuals, and ndf represents the number of degrees of freedom. The final statistical values for f_i over all birds are:

$$rms = 0.00985,$$

$$rms_{VAR} = 9.7023e - 05,$$

$$\sum rms = 0.016355,$$

$$L_{app} = -7.10079e - 06,$$

where $\sum rms$ is the final sum of squares of the residuals, and L_{app} is the relative change during last iteration.

After 5 or 6 iterations, the fit converged for polynomial functions (f_3, f_4 , and f_5). For all birds $L_{app} = -1.15631e - 14$ and $\sum rsm = 0.123618$. The exponential functions (f_1 and f_2) needed 1442 iterations, for all birds, in order for the fit to converge. Although their final sum of residuals is similar to polynomials $\sum rsm = 0.1188$, the relative change is much higher: $L_{app} = -6.85045e - 06$.

In summary, computing time is much longer for exponential functions than for polynomials, but the fit accuracy can still increase upon addition of iterations, if required. Table 17 contains the final parameters for all 47 birds, using fit function f_i .

#	Name	Order	a	b	c	d	k
1	Mallard Duck	Anseriformes	4.31	0.74801	166.40	-0.03923	-155.87
2	Australian Grey Swiftlet	Apodiformes	5.12	0.72541	722.61	-0.01213	-709.73
3	Oilbird	Caprimulgiformes	3.03	0.72096	555.64	-0.00995	-544.39
4	Emu	Struthioniformes	4.93	0.76381	94.71	-0.05690	-85.47
5	Plains Wanderer	Charadriformes	5.28	0.71786	269.61	-0.01264	-256.23

6	Pigeon	Columbiformes	5.28	0.71786	269.61	-0.01264	-256.23
7	American Kestrel	Falconiformes	5.20	0.71901	773.27	-0.00976	-767.12
8	European Sparrowhawk	Falconiformes	4.98	0.72022	686.06	-0.01059	-679.18
9	Bobwhite Quail	Galliformes	4.98	0.72022	686.06	-0.01059	-679.18
10	Chicken	Galliformes	5.24	0.74625	129.75	-0.04339	-123.75
11	Japanese Quail	Galliformes	4.88	0.72212	808.10	-0.01038	-800.67
12	Turkey	Galliformes	4.22	0.74520	162.55	-0.03794	-152.20
13	American Robin	Passeriformes	2.57	0.72342	560.77	-0.00961	-549.50
14	Blue Jay	Passeriformes	3.65	0.72056	564.62	-0.01004	-553.53
15	Brown-headed Cowbird	Passeriformes	2.68	0.72258	541.25	-0.01020	-528.91
16	Bullfinch	Passeriformes	2.29	0.72620	612.80	-0.00973	-602.20
17	Chipping Sparrow	Passeriformes	1.77	0.72844	626.83	-0.00938	-613.52
18	Common Canary	Passeriformes	1.93	0.77461	131.32	-0.04369	-115.68
19	Common Crow	Passeriformes	3.65	0.84982	55.85	-0.14035	-53.08
20	European Starling	Passeriformes	3.40	0.71989	495.43	-0.01019	-486.38
21	Field Sparrow	Passeriformes	2.55	0.72329	525.87	-0.00988	-514.20
22	Fire finch	Passeriformes	4.35	0.72216	851.44	-0.00967	-840.20
23	Great tit	Passeriformes	2.84	0.72198	533.27	-0.01014	-524.10
24	House finch	Passeriformes	4.64	0.72086	821.00	-0.00957	-809.68
25	House Sparrow	Passeriformes	4.97	0.74783	157.05	-0.04175	-154.77
26	Pied Flycatcher	Passeriformes	3.63	0.71599	1611.10	-0.00436	-1599.17
27	Red-winged Blackbird	Passeriformes	2.83	0.72217	603.29	-0.00920	-591.26
28	Slate-colored Junco	Passeriformes	3.47	0.72581	868.48	-0.00993	-858.83
29	Song Sparrow	Passeriformes	2.63	0.71710	1055.87	-0.00503	-1045.56
30	Swamp Sparrow	Passeriformes	2.39	0.73464	589.90	-0.00929	-578.57
31	Western Meadowlark	Passeriformes	4.91	0.71937	752.96	-0.00985	-747.94
32	Zebra Finch	Passeriformes	3.00	0.72377	638.74	-0.01001	-624.11
33	Bourke's Parrot	Psittaciformes	3.42	0.71947	525.21	-0.00971	-513.36
34	Budgerigar	Psittaciformes	4.30	0.71979	712.44	-0.00959	-705.55
35	Cockatiel	Psittaciformes	4.51	0.71872	546.44	-0.01033	-538.82
36	African Wood Owl	Strigiformes	2.93	0.72179	567.49	-0.00977	-563.23
37	Barn Owl	Strigiformes	1.69	0.72489	494.89	-0.00895	-489.41
38	Brown Fish Owl	Strigiformes	5.15	0.71383	-570.65	0.00701	573.70
39	Eagle Owl	Strigiformes	3.31	0.71952	508.69	-0.00956	-509.03
40	Great Horned Owl	Strigiformes	4.27	0.70748	-353.80	0.00769	356.86
41	Long Eared Owl	Strigiformes	3.03	0.72271	650.37	-0.00959	-648.88
42	Mottled Owl	Strigiformes	2.03	0.72161	201.92	-0.01329	-197.65
43	Scops Owl	Strigiformes	1.84	0.84928	37.69	-0.13088	-32.77
44	Snowy Owl	Strigiformes	5.65	0.72275	1029.44	-0.01038	-1029.17
45	Spotted Wood Owl	Strigiformes	2.86	0.74366	124.33	-0.03647	-122.58
46	Tawny Owl	Strigiformes	3.26	0.71946	537.69	-0.00902	-538.25
47	White-faced Scops Owl	Strigiformes	3.91	0.72062	568.18	-0.01025	-568.72

Table 17: Function f_i coefficients for 47 species.

As mentioned above, we aggregated the birds into three groups based on the orders. Final groups and overall coefficients are presented in Table 18.

#	Group	a	b	c	d	k
1	Non-Passeriformes	18.7657	0.692153	-2884.21	0.0076418	2881.98
2	Passeriformes	12.0993	0.712241	1904.88	-0.0110408	-1901.79
3	Strigiformes	12.2142	0.710617	1379.51	-0.012282	-1402.81
4	All Birds	11.8718	0.734457	524.5	-0.0360242	-527.81

Table 18: Species group aggregate. Final f_i coefficients for four group, based on the order.

Figure 7.5 contains the final group audiograms from Table 18. All particular species audiograms are addressed in the Attachment section together with an error value calculation.

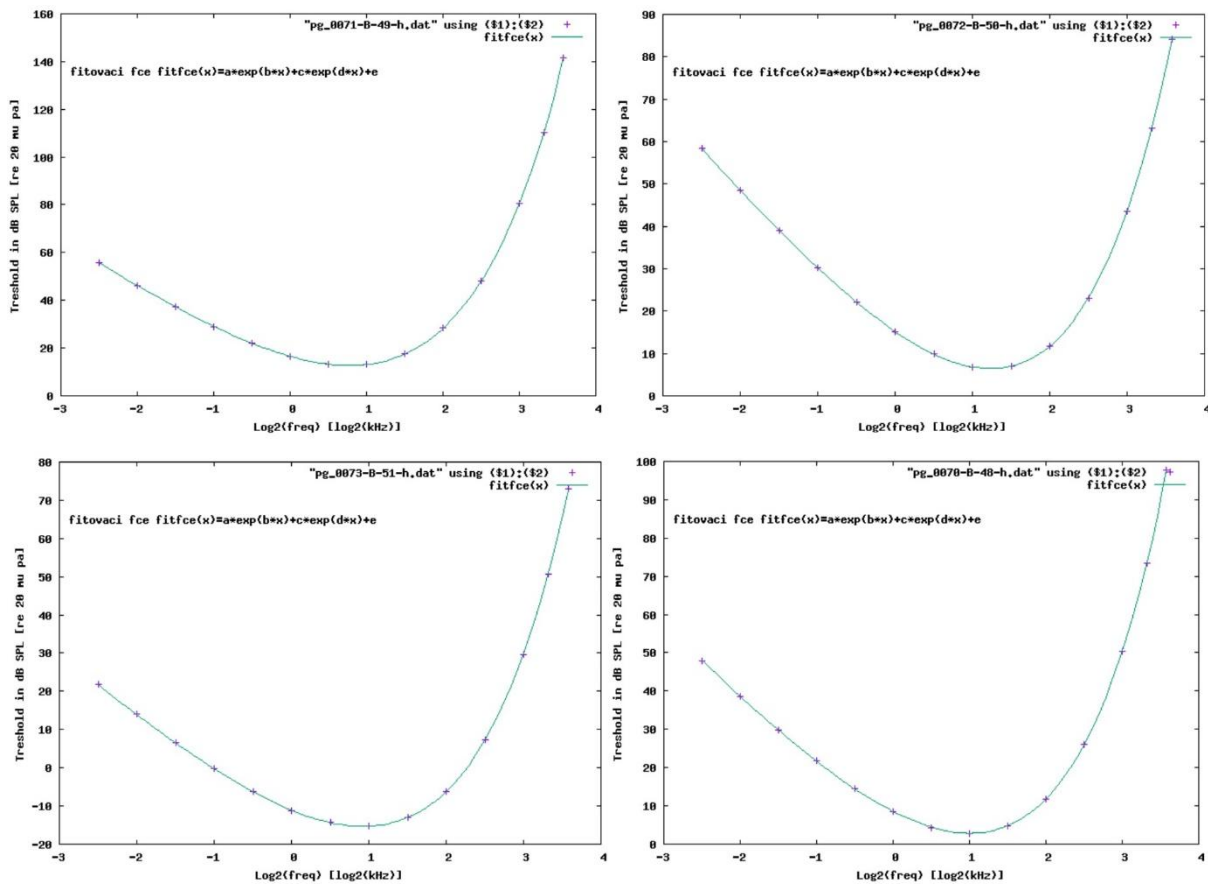


Figure 7.5: Final bird audiograms. The graphs display audiograms aggregated by order. Non Passeriformes (up left), Passeriformes (up right), Strigiformes (down left), and all birds (down right). All species audiograms see in Attachment.

7.4 Contribution

Both automated systems as for an individual identification as well as for the species recognition should use the introduced unified audiogram equation together with particular coefficients setup. One could automate and tailor the signal processing toward, for example, a filter bank distribution, pre-emphasis, or band weighting respecting species hearing sensitivity.

We could also use AUE when we are handling with bird audiograms (researches, comparative studies, etc.). Until now, we depend on scanning of the audiogram figures. Similarly, the audiogram values reading is more accurate thanks the AUE equation. However, as we stated above, we should keep in mind the bird audiogram source data is more rare than as for the human data.

We plan to introduce the bird audiogram unified equation in [PTA16a].

7.5 Summary

From a technical point of view, the audiograms are *just pictures*. As audiograms are not determined by mathematical expressions, it is impossible to work with them any further (e.g. in the Matlab, C++, ARSBI, ARSBS, etc.).

The presented Bird AUE describes the bird audiograms by one equation. We alternate audiogram types just by substituting particular coefficients of specific bird. We discovered coefficients for forty-seven species and for four aggregated audiograms (all birds, Non-Passeriformes, Passeriformes, and Strigiformes).

A new coefficient's setting can be discovered for a new species in the future to extend the equation's range of use, that is, not just for the birds but theoretically for any animals including exotic ones (meaning in hearing range) such as a bat or a whale.

8 Bird Adapted Filter

8.1 Introduction

The Mel filter bank is the common filter distribution (see sections 3.5.4.1. and 7.1), but it is optimized for human speech. A bird song differs in both the time domain and frequency features compare to speech. We aimed to discover a feature extraction setup optimized for bird song.

At first, we determined the basic parameter setup. Then we developed and tested a new so-called *Bird Adapted Filter* (BAF) filter distribution based on the results of the chapter 7. In the last section, we compare the common filter distributions, namely the linear, and MEL with BAF for a bird individual identification.

8.2 Parameters optimization

There is not any recommended parameters set up for a bird song unlike for human speech (for instance, *window length* 30 ms, *overleap* 15 ms, *pre-emphasis* 0.97, etc.). Actually, the setup has to be evaluated from the scratch for each research. It follows from the differences among

- species (frequency, dynamic, duration),
- bird individuals (syllables rate, lazy/diligent singers),
- recordings (noise level, background type and level, weather conditions).

At first, we determined the optimal parameters setup. The Framework cannot perform a multi-dimensional optimization. Therefore, this process was complicated because there are many crucial parameters, which have to be tuned one at a time.

In total, 267 experiments were executed, consisting of 165.4 million trials. See Table 19 for the optimized values of the main parameters.

	Test range		Optimized values
	From	To	
Min frequency	0 Hz	2 kHz	800 Hz
Max frequency	5 kHz	11.025 kHz	10 kHz
Window length	10 ms	50 ms	20 ms
Window shift	10 ms	50 ms	10 ms
Window type	Triangle, Hamming, Blackmann, Hanning		Hamming
Number of band filters	10	25	20
Band filters distribution	Linear, MEL		Linear
Filter shape	Triangle, Hamming, Blackmann, Rectangle		Triangle
Pre-emphasis	0.80	1	0.97 (no effect)
Number of cepstral coefficients	10	30	20
Number of delta coefficients	0	30	20
Number of delta-delta coefficients	0	30	0

Table 19: List of optimized parameters.

8.3 Filter distribution definition

The audiogram equation (73) is taken as the initial function. Each audiogram is defined from a LF (low frequency) to HF (high frequency). Both frequencies HF and LF depend on the particular bird see Table 16. We adapt the equation (73) to calculate discrete variables as

$$y_i = a \cdot e^{b \cdot x_i} + c \cdot e^{d \cdot x_i} + k. \quad (77)$$

It is also valid that

$$LF \leq x_i \leq HF, \quad (78)$$

for $i = \{1, 2, \dots, F\}$, where F is the total number of frequencies. We now define a cumulative sum z as

$$z_i = \sum_{j=1}^i [\max(y) - y_j], \quad (79)$$

where z_i is the i -th coefficient of the cumulative sum. Finally, we use z as the distribution function of BAF triangle filters fitting a bird's hearing sensitivity. Figure 8.1 displays an example of cumulative sum z function.

We can also normalize z_i , if an application required, as

$$\tilde{z}_i = \frac{HF - LF}{z_i} (z_i - z_1) + LF, \quad (80)$$

where \tilde{z}_i represents the normalized cumulative sum of z .

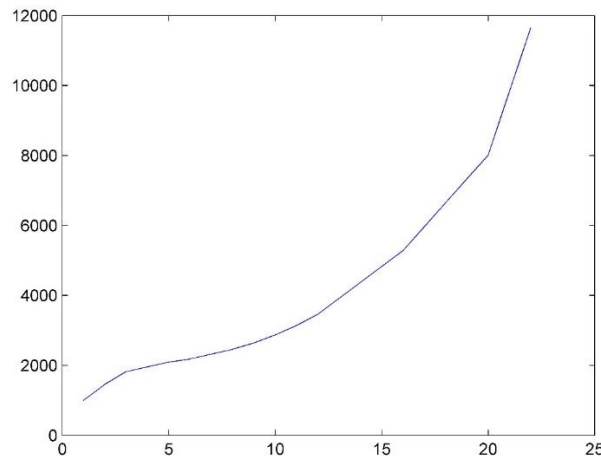


Figure 8.1: Example of cumulative sum z function.

Further, we can change the filters' overlap to increase a particular band sensitivity. Let us consider set of N triangle filters with linear distribution. We could compute a half of the filter length a as

$$a = \frac{HF - LF}{N + 1}, \quad (81)$$

where N is the filter's number. We state an *overlap ratio* b that is a simple difference between overlap and a . Then $L_2 = (a + b)$, where L_2 is the starting point of a second filter. Based on both a and b there are three variants of the overlapping, see Figure 8.2:

- $b=0$ Overlap and a are the same.
- $b<0$ Overlap is higher than a half.
- $b>0$ Overlap is lower than a half.

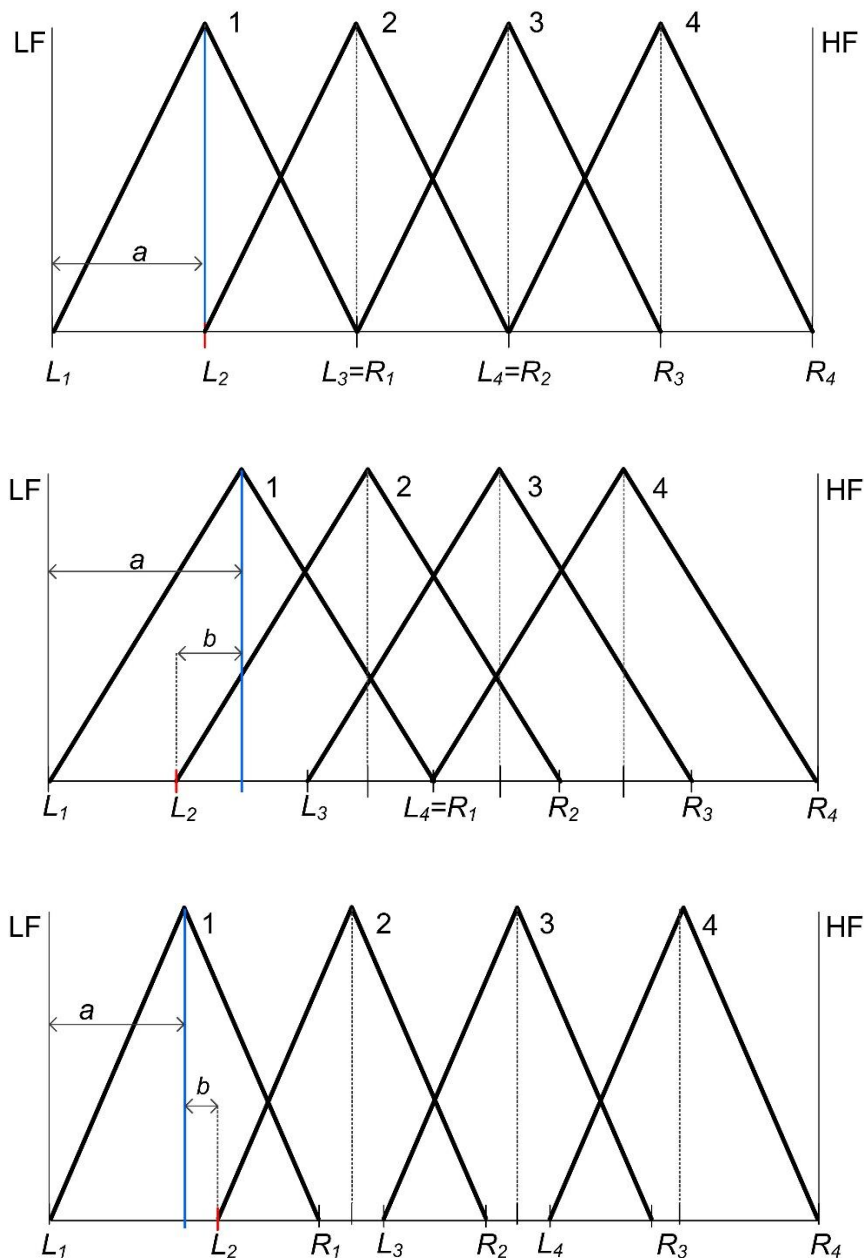


Figure 8.2.: Linear filter distribution with different overlap. A) Overlap is just a half of triangle length a . The overlap ratio $b=0$. B) Overlap is higher than a half. The overlap ratio $b<0$. C) Overlap is lower than a half.

The overlap ratio $b>0$. Legend: $L_x...$ Triangle x left point, $R_x...$ Triangle x right point.

Finally, we can express the filter's starting and ending points L_x, R_x

$$L_x = LF + a(x - 1)(1 + b), \quad (82)$$

$$R_x = HF - a(N - x)(1 - b). \quad (83)$$

Notice, one variant typically defines the ratio between *overlap* and *filter length* in percent; for example, if $b=0$ then the overlap is 50%.

8.4 Experiment evaluation

We experimented with the BAF distribution to reveal any improvements by changing b value:

$$b = \left\{ -\frac{1}{2}, -\frac{1}{3}, -\frac{1}{4}, -\frac{1}{5}, 0, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2} \right\}.$$

For the $b = -\frac{1}{2}$, the accuracy did not rise significantly, but the computation time was affected. Higher values $b > -\frac{1}{4}$ did not improve the accuracy; even $b > 0$ worsened it. Therefore, $b = -\frac{1}{3}$ was chosen after these preliminary experiments as a balanced value.

We chose four types of filter distributions for our experiments:

- Linear ($b = 0$)
- MEL ($b = 0$)
- BAF ($b = 0$)
- BAF 1/3 ($b = -\frac{1}{3}$).

We use the terms *BAF* and *BAF 1/3* to distinguish between BAF with overlap $b = 0$ and $b = -\frac{1}{3}$. For BAF *Passerine* distribution see Figure 8.3. Figure 8.4 displays the *Passerine* BAF 1/3.

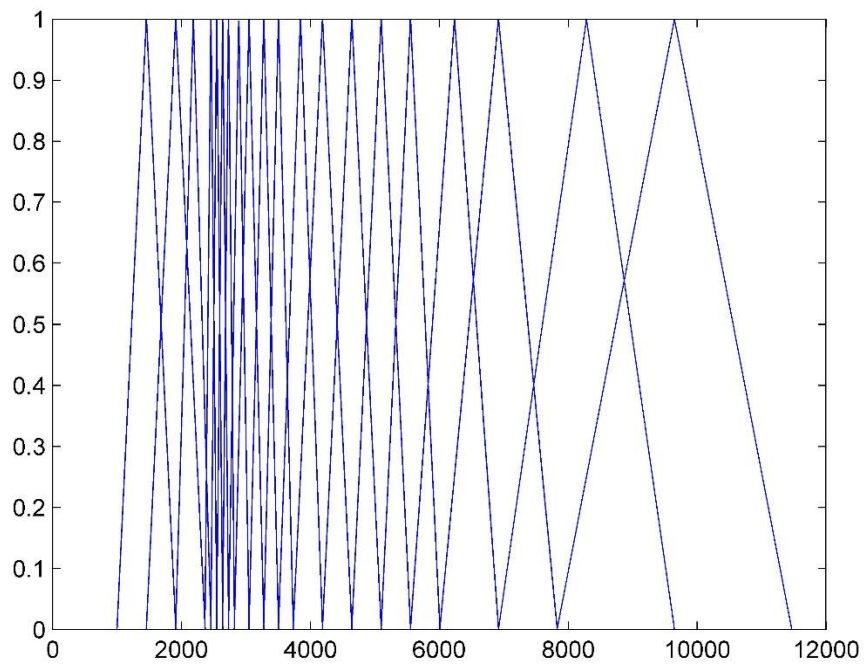


Figure 8.3.: BAF distribution for Passerine. Standard overlap $b = 0$.
Number of filters $N=20$.

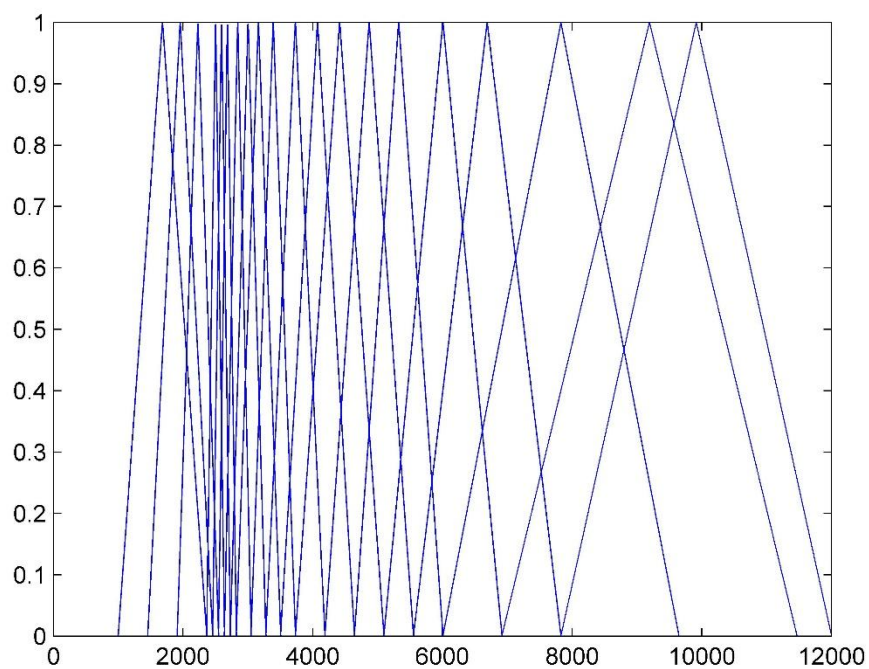


Figure 8.4: BAF 1/3 filter distribution for Passerine. Overlap $b = -\frac{1}{3}$. Number of filters $N=20$.

8.5 Results

Table 19 contains the system parameters setup. We chose 15 testing sets for the experiments. Each set consists from 1400 to 3150 files. In total, 405 experiments were performed, which required 637.2 million trials. Table 20 gives the result of comparing Linear, MEL, BAF, and BAF 1/3 filter distributions. Graphical results are shown in Figure 8.5. We used EER (see section 3.4.6) as the main decision value.

	Lin	Mel	BAF	BAF 1/3
DataSet 01	13.06%	15.38%	13.13%	10.56%
DataSet 02	12.38%	16.11%	15.55%	11.61%
DataSet 03	13.54%	18.44%	13.00%	9.26%
DataSet 04	13.13%	16.31%	12.00%	9.36%
DataSet 05	13.53%	21.38%	12.63%	9.69%
DataSet 06	23.44%	24.26%	23.25%	20.23%
DataSet 07	25.63%	23.86%	21.92%	18.31%
DataSet 08	14.85%	15.09%	14.55%	13.47%
DataSet 09	12.50%	14.70%	12.70%	12.42%
DataSet 10	13.25%	14.48%	12.29%	11.29%
DataSet 11	15.56%	16.88%	12.19%	11.54%
DataSet 12	16.96%	18.85%	14.99%	12.13%
DataSet 13	16.25%	22.52%	15.13%	12.14%
DataSet 14	18.88%	22.56%	15.31%	13.05%
DataSet 15	20.56%	19.88%	17.94%	15.13%
Average	16.23%	18.71%	14.29%	12.68%
Min	12.38%	14.48%	10.38%	9.26%
Max	25.63%	24.26%	23.18%	20.23%

Table 20: EER for different bank filter distribution: Linear, Mel, BAF ($b = 0$), and BAF 1/3 ($b = -\frac{1}{3}$).

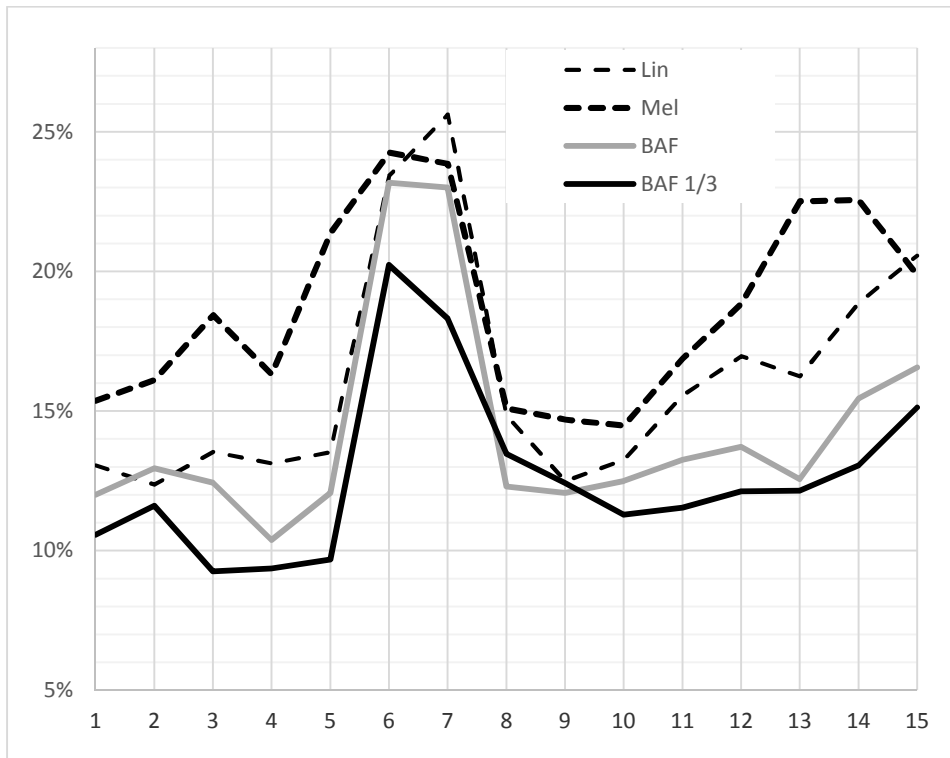


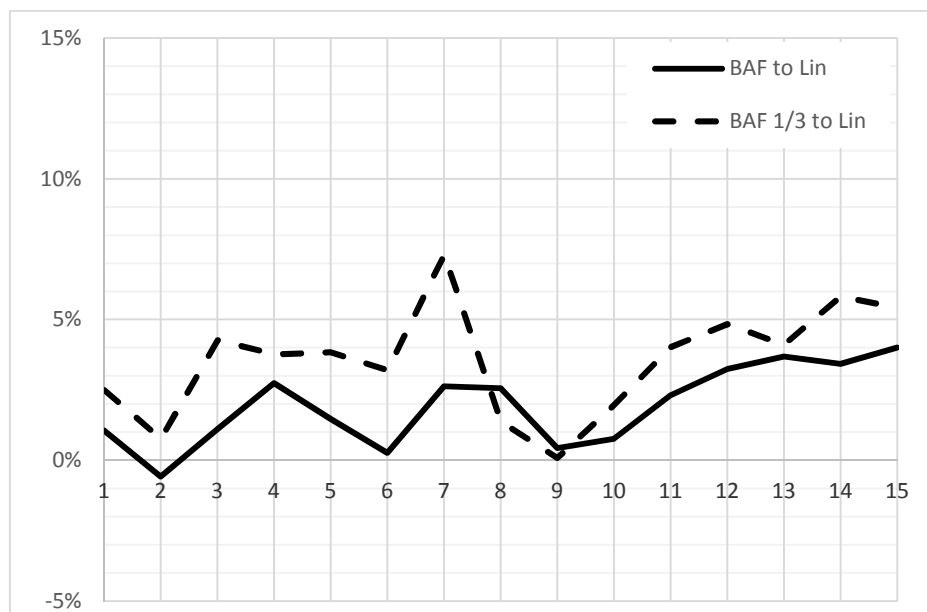
Figure 8.5: EER of different bank filter distribution: Linear, Mel, BAF ($b = 0$), and BAF 1/3 ($b = -\frac{1}{3}$).

Table 23 contains the accuracy improvement in detail. A positive improvement was reached in all cases with just one exception (BAF to Lin, Data set 02). For graphs, see Figure 8.6.

	BAF to Lin	BAF 1/3 to Lin	BAF to MEL	BAF 1/3 to MEL
DataSet 01	1.06%	2.50%	3.38%	4.81%
DataSet 02	-0.58%	0.77%	3.15%	4.50%
DataSet 03	1.11%	4.28%	6.01%	9.18%
DataSet 04	2.74%	3.77%	5.93%	6.96%
DataSet 05	1.46%	3.84%	9.31%	11.69%
DataSet 06	0.26%	3.21%	1.08%	4.03%
DataSet 07	2.63%	7.31%	0.86%	5.54%
DataSet 08	2.56%	1.38%	2.80%	1.63%
DataSet 09	0.44%	0.08%	2.63%	2.28%
DataSet 10	0.76%	1.96%	1.99%	3.20%
DataSet 11	2.31%	4.03%	3.63%	5.34%

DataSet 12	3.24%	4.84%	5.13%	6.72%
DataSet 13	3.69%	4.11%	9.96%	10.38%
DataSet 14	3.43%	5.82%	7.12%	9.51%
DataSet 15	4.00%	5.44%	3.31%	4.75%
Average	1.94%	3.56%	4.42%	6.03%
Min	-0.58%	0.08%	0.86%	1.63%
Max	4.00%	7.31%	9.96%	11.69%

Table 21: Accuracy comparing. Positive value represents improvement and negative, worsening. The table gives EER differences among particular filter distributions. For source data, see Table 20.



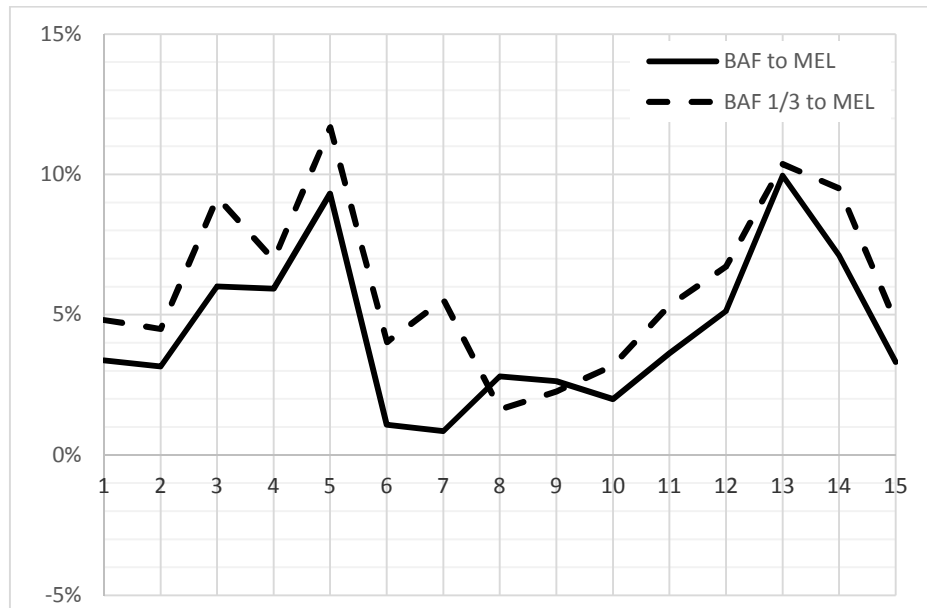


Figure 8.6: Different bank filter distribution accuracy improvement. For source data, see Table 23.

8.6 Contribution

Experiments revealed a BAF distribution improved an accuracy of bird individual identification for the tested data. A BAF filter distribution was based on the *passerine* hearing sensitivity, see details in chapter 7, and Table 18 respectively. Because the BAF distribution respects a hearing sensitivity of a particular species, it can be also used by an automation system for species identification. We plan to introduce the BAF in [PTA16b].

8.7 Summary

A BAF filter distribution improved the accuracy in all experiments except for one (*BAF to Lin, Data set 02*). The average EER value for Linear was 16.23%, and for MEL 18.71%. The BAF gave 14.29%, and BAF 1/3 12.95%. The lower EER improvement occurred between BAF and Linear with 1.94% on average and the greater improvement of 6.03% between BAF 1/3 and MEL. The highest improvement was 11.69% (Data set 5, BAF 1/3 to MEL).

A statistical analysis was performed to compare the results of the Linear and BAF distribution methods while the results were obtained by the standard (baseline) MEL method. The results of the statistical tests revealed that both the Linear and the BAF method yield a statistically significant ($p < 4.9e-04$ and $p < 3.1e-05$ respectively) improvement over the baseline. All comparisons were performed using the one-sided Wilcoxon signed rank test; the results with p values less than 0.05 were considered statistically significant. The computations were done with the MATLAB Statistics Toolbox.

We plan to use a BAF for different bird data to verify an identification accuracy improvement. We also intend to integrate a BAF into some species recognition system to check the assumption a BAF should even improve a species identification accuracy.

9 Improving automatic bird identification by data merging

9.1 Introduction

We were considering first a data merging idea by the time we processed a small amount of animal recordings. The basic idea of this method is to make a model more robust and suitable by joining data even when there is no chance to obtain more songs.

Section 3.2.3 exhibits essential variants of the bird songs experiments involving work with raw recordings or with songs extracted from the recordings. This merging method deals with a kind of a midway approach. We used neither simple songs nor joined the recordings, but we tried to merge just a few songs into the one. The main motivation for the experiment was to handle only the cases when a limited amount of recorded data is available.

9.2 Method principle

In the nature, ornithologist battle with overlap, background noise, and a long distance between recorder and singer. [SED15] deals with species recognition over as-is recordings by comparing two methods: spectrogram point counting and automated processing. Swiston and Mennill [SWI09] compared manual and automated methods for identification of specific sounds (i.e. particular type of syllables) in continuous recordings. For manual method, the authors manually checked spectrograms. For automated, they used Data Template Detector software (Harold Figueroa, Ithaca, NY) for tracking “lernal” song types. The Data Template Detector was based on signal thresholds level setting. The result indicates the manual scanning outdo an automated method. Nevertheless, there is in question if authors selected the right automated method for the application.

One has to choose an efficient strategy to work with long-lived recordings (provided by an ornithologist) dealing with either bird individual identification or species recognition systems. Let us suppose we cut off the songs from the recordings. Then there are two options: at first, we join all songs, which belong to the one bird individual, into the one recording and then we extract features from this *big song data*. The second option is to extract features, *one by one*, from the single songs.

One such data merging method combines both approaches. Instead of joining all recordings into the one bulk, we merge just a defined number of data, and a so-called *merging level* defines a number of joined songs. After the data merging we extract features from this data (i.e. *merged recordings*, *merged songs*). The principal idea of the method is demonstrated on the Figure 9.1.

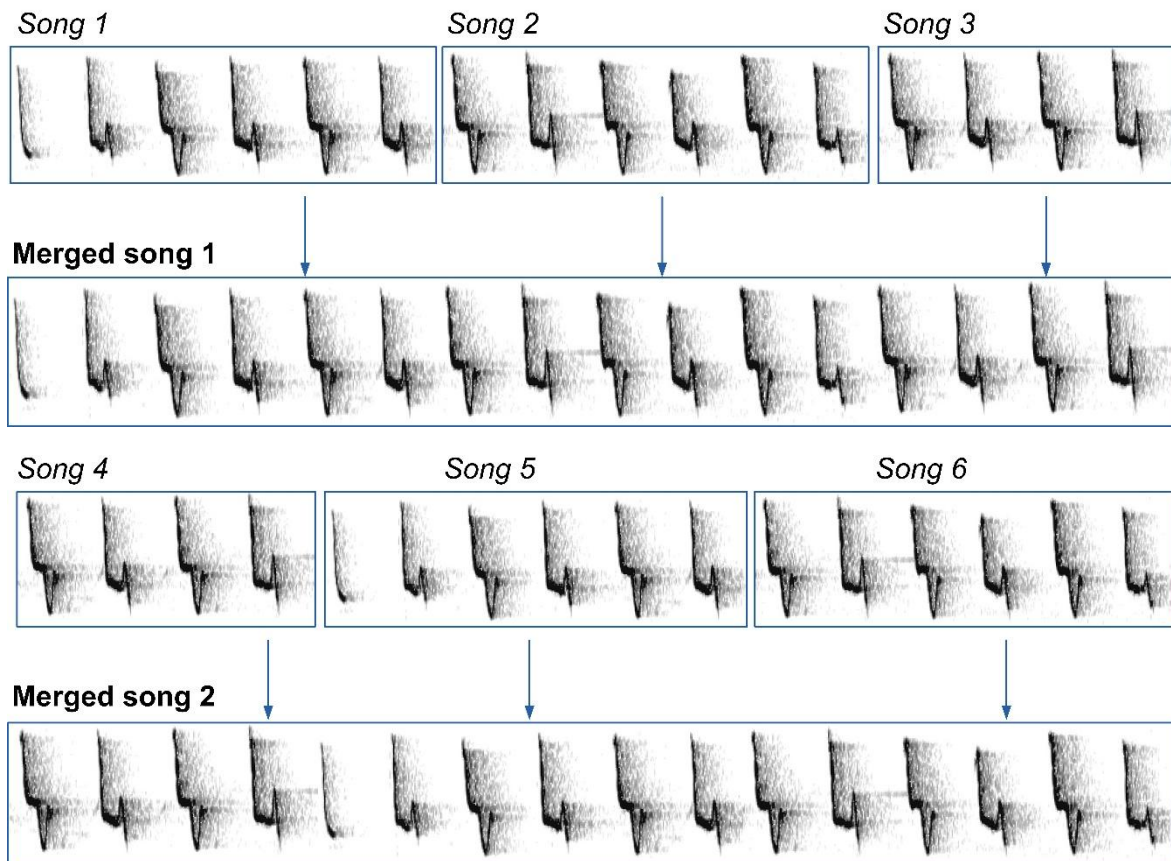


Figure 9.1: The principal idea of the data merging method. A defined number of recordings composes the train data. The figure shows an example for merging level = 3.

We designed a new tool in the Matlab for data merging to test the method. We merged just the songs of the same recording origin because there could be a significant difference between recordings (see section 3.2). It was impossible to collect necessary amount of songs for some data sets, mainly for highest merging, see so-called IDA (insufficient data amount) cells in the Table 23. We also prepared an automation experiment tool in Matlab, see chapter Development Framework.

9.3 Results

We identified bird individuals by the GMM-UBM method (see section 3.4.2). For results, see Table 23. Twelve sets of chiffchaff recordings were evaluated; each consists of approximately 2,000 songs. We processed in total 1.3 million trials. The number of bundled data increased from two to ten.

Set File	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10	Set 11	Set 12
1	19.4%	21.9%	21.3%	22.3%	22.3%	22.1%	32.1%	16.1%	22.9%	21.8%	15.5%	21.1%
2	18.0%	21.7%	18.9%	21.5%	21.3%	22.1%	31.8%	16.1%	24.2%	20.4%	15.8%	20.0%
3	17.6%	24.0%	19.4%	20.1%	23.1%	20.2%	32.9%	16.6%	26.2%	26.2%	14.0%	17.0%
4	16.6%	22.1%	16.3%	18.5%	17.0%	15.0%	30.0%	13.4%	23.6%	22.4%	14.6%	19.0%
5	18.3%	20.7%	18.0%	20.5%	18.7%	15.0%	29.5%	12.5%	20.7%	24.0%	17.2%	26.9%
6	16.8%	19.0%	14.6%	16.4%	18.0%	14.0%	28.3%	11.3%	27.3%	30.7%	14.8%	19.0%
7	22.8%	19.0%	16.3%	21.8%	18.0%	14.0%	28.6%	8.8%	16.2%	16.3%	23.9%	19.0%

8	14.0%	19.0%	15.5%	17.8%	18.0%	14.0%	27.3%	IDA	16.2%	16.3%	13.1%	IDA
9	10.2%	IDA	10.7%	7.9%	19.0%	14.0%	28.6%	IDA	27.3%	27.4%	IDA	IDA
10	12.2%	IDA	7.9%	IDA	12.8%	12.7%	IDA	IDA	IDA	IDA	IDA	IDA

Table 22: Data merging: Experiment results, total EER. First line gives the standard EER without any data merging, labelled as EER_0 . See the EER suffix for particular merging level. For instance, the row EER_5 gives the EER for merging level 5. Notice the experiments were not performed for some merging levels because of insufficient data amount (IDA).

Merging level	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10	Set 11	Set 12
2	1.4%	0.2%	2.4%	0.8%	1.0%	0.0%	0.3%	0.0%	-1.3%	1.4%	-0.3%	1.1%
3	1.8%	-2.1%	1.9%	2.2%	-0.8%	1.9%	-0.8%	-0.5%	-3.3%	-4.4%	1.5%	4.1%
4	2.8%	-0.2%	5.0%	3.8%	5.3%	7.1%	2.1%	2.7%	-0.7%	-0.6%	0.9%	2.1%
5	1.1%	1.2%	3.3%	1.8%	3.6%	7.1%	2.6%	3.6%	2.2%	-2.2%	-1.7%	-5.8%
6	2.6%	2.9%	6.7%	5.9%	4.3%	8.1%	3.8%	4.8%	-4.4%	-8.9%	0.7%	2.1%
7	-3.4%	2.9%	5.0%	0.5%	4.3%	8.1%	3.5%	7.3%	6.7%	5.5%	-8.4%	2.1%
8	5.4%	2.9%	5.8%	4.5%	4.3%	8.1%	4.8%	IDA	6.7%	5.5%	2.4%	IDA
9	9.2%	IDA	10.6%	14.4%	3.3%	8.1%	3.5%	IDA	-4.4%	-5.6%	IDA	IDA
10	7.2%	IDA	13.4%	IDA	9.5%	9.4%	IDA	IDA	IDA	IDA	IDA	IDA
Average	2.8%	1.0%	5.4%	3.8%	3.5%	5.8%	2.2%	2.5%	0.2%	-1.0%	-0.6%	0.8%

Table 23: Data Merging Identification improvement. The table is based on the EER results from previous table. Each line contains differences between particular merging level EER and EER with no merging. For example a third line (Merging level 4) contains differences between EER_4 and EER_0 .

Following graphs display an accuracy improvement in detail; see Figure 9.2, Figure 9.3, Figure 9.4, and Figure 9.5. Notice the data origins from Table 23.

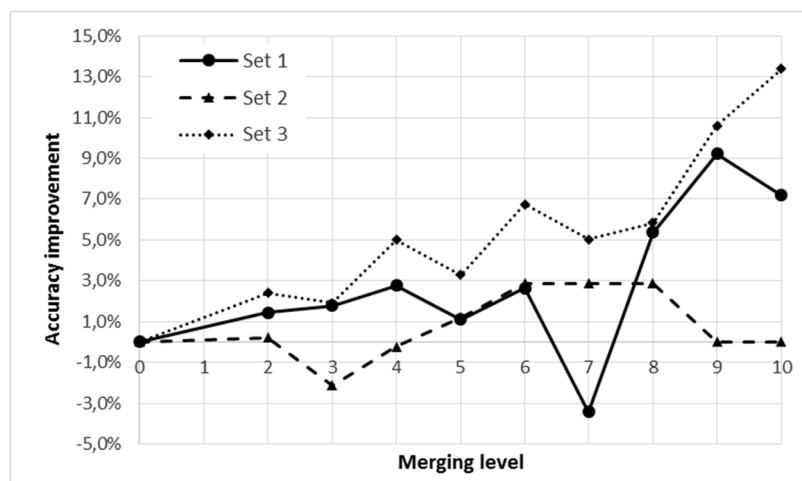


Figure 9.2: Accuracy improvement: Data sets 1, 2, and 3.

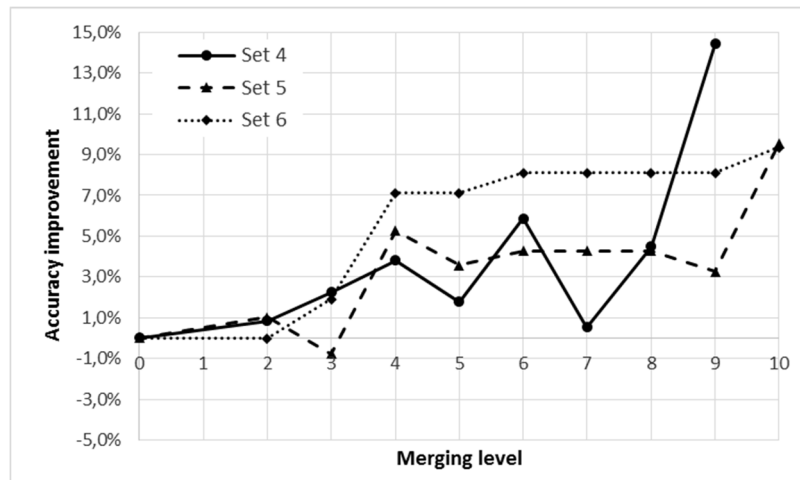


Figure 9.3: Accuracy improvement: Data sets 4, 5, and 6.

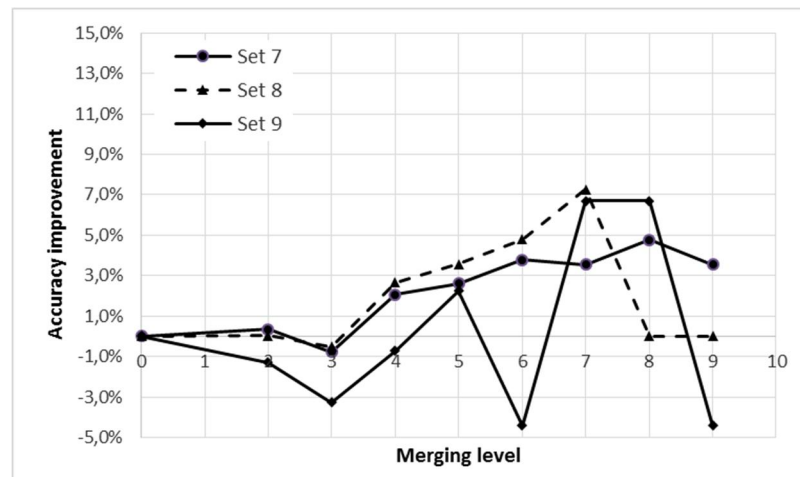


Figure 9.4: Accuracy improvement: Data sets 7, 8, and 9.

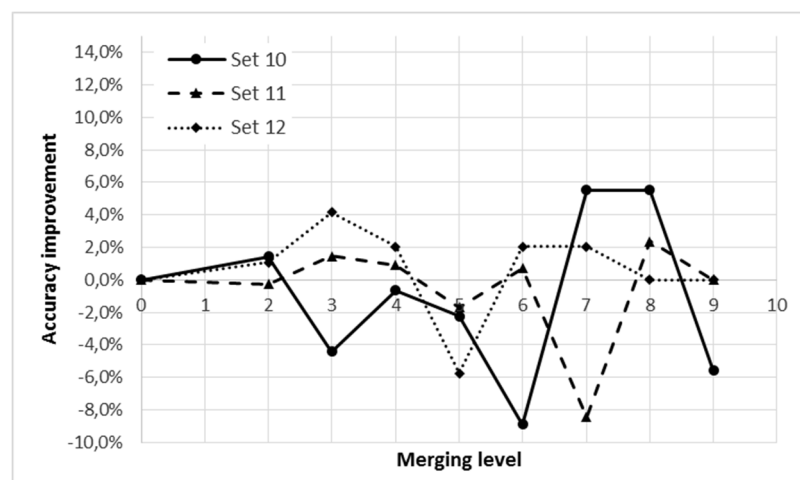


Figure 9.5: Accuracy improvement: Data sets 10, 11, and 12.

We provided 94 experiments i.e. combination of all sets x nine merging levels. The recognition improvement accuracy was achieved for 77.7% of these experiments (73 cases). The absolute achieved improvement was 3.0 % on average, see Table 24. The highest value is 14.4% (set 4, merging level 9), the lowest -8.9% (set 10, merging level 6). The best results achieved the merging level 9, where an average of 9.9% improvement was achieved. However, just a few sets were involved because of insufficient amount of data.

Merging level	2	3	4	5	6	7	8	9	10	Average
Average improvement	0.6%	0.1%	2.5%	1.4%	2.4%	2.8%	5.0%	4.9%	9.9%	3.0%
Number of experiments	12	12	12	12	12	12	10	8	6	
Number of insufficient data amount (IDA)	0	0	0	0	0	0	2	4	8	

Table 24: Data merging identification improvement, summary.

9.4 Contribution

Researchers dealing with an automation processing of animal vocalization still suffers from an insufficient amount of data. In such cases any method, which could improve an identification accuracy, is very welcomed. Although the basic idea of data merging is simple, the results indicate the contributory of the method. We introduced and discussed the research in [PTA15c].

9.5 Summary

The results reveal the method could improve individual/species identification accuracy in the case we operate with insufficient recording number. We provided the Wilcoxon signed rank test (left-tailed) for levels 0, 2, and 4 with following results: significance between *level 0* and *level 2*: $p = 0.033$ and between *level 0* and *level 4*: $p = 0.048$.

Although with the positive Wilcoxon test, we are keenly aware that the method cannot be apprehended as a universal tool for identification accuracy improvement. The partial result analysis (see previous graphs) leads to the assumption that the method could improve the accuracy for only some data. For some data sets the improvement grows at a rate proportional to merging level (e.g. Figure 9.2 and Figure 9.3); for some, the accuracy changes with high deviation (e.g. Figure 9.4, set 7), even it is worse (e.g. Figure 9.5, data set 10 and 11).

Before the application of the method, we suggest users tune the identification system by increasing the merging level gradually. The decision to use or not to use the method will need to be made after the results have been achieved.

10 Bird Song Database

10.1 Introduction

There is many amateur and professional ornithologists over the world, there is many people recording the bird song. Nevertheless, they do not share their recordings compared to linguists. The speech-to-text systems use language corpuses and speech databases (i.e. human-speech data provided by National Institute of Standards and Technology) for a long time.

There are commercial databases of bird and animal sounds on the market. Ornithology laboratories (e.g. *Cornell Lab of Ornithology* or *Borrer Laboratory*) collect bird songs and release them as Audio CDs. Professional ornithologists have also released Audiobooks (e.g. *Elliott, L., Read, M.: Common Birds and Their Songs*), or CDs as a book appendix (e.g. *Borrer, D., J.: Common Bird Songs*). These CDs are sold in many stores (e.g. *amazon.com*, *discogs.com*), often categorized under *Field Recordings*. *British Library* offers Wildlife and environmental sounds which is the largest collection of its kind in Europe and the most comprehensive in the world. *Cornell Lab of Ornithology*, Macaulay Library archive contains over 250,000 digital audio and video recordings of birds, mammals, amphibians, reptiles, fishes, and insects from around the world. Users may request access to download media from the traditional archive for research, educational, and personal uses. We explored the Cornell database for purposes of our thesis but it contains just 24x recordings of Common Chiffchaff (*Phylloscopus collybita*), on the other way an user find about 190 chiffchaff photos. *Avibase* is an extensive database information system about all birds of the world, containing over 17 million records about 10,000 species and 22,000 subspecies of birds, including distribution information, taxonomy, synonyms in several languages, also in Czech. *Borrer Laboratory* of Bioacoustics archive contains over 40,000 animal sound recordings. They provide recordings for research, education, management, and other uses. The recordings requests from these libraries are under licensing fees except research purposes.

There are many amateur ornithologists the world over, and many record the bird songs. Nevertheless, their bird song database is not open source. Few non-commercial databases exist to share recordings, the most known is *xeno-canto.org*. Another database was introduced in [ARR15]. It serves for storing and annotating of sequence of vocal sounds (song, phrase). Over 1000 recordings were collected for more than 30 bird species up to date. Main purpose of the database is to collect long-recordings and phrase tagging providing besides. Although some valuable function, actually this database could not be used for similar purpose as BSC, because it is not based strictly on rules for bird individual identification, for instance an individual identity classification requisite (just as optional filed), lack of some data (distance, localization, weather), song structure decomposition missing (song->phrase->syllables), etc.

The main problem of these databases is the *inconsistent quality of recordings*. Amateur ornithologists make the recordings under different climatic and noise conditions with different recording devices. Nevertheless, there is great potential for an automated bird recognition tool(s) because a bird song listening, recording, and collecting has been a popular hobby for a hundred years.

Scientists usually do not share recordings. If yes, so only with a close community like an institute, or faculty. However, a huge amount of data is necessary for an ARSBI design. This led the author to aim to build up a Bird Song Database (BSC) for scientific purposes. The BSC was designed and developed in the last three years and is under a commercial preparation process now.

The main BSC benefits will be data sharing, experiment repeating, and ARSBI tuning. The database should, similarly like human speech corpus, serve the scientists for data exchange and experiment making with those data. Thanks to BSC, it will be easier compare the achieved results and optimize the used methods.

10.2 Requirements

The general requirement for BSC is an easy manipulation with large data amount. Using the term data we mean recordings of birds and / or other animal species. Manipulation means uploading, viewing and downloading.

BSC is designed for cataloguing of bird recordings using a web interface - browser. It also allows to present data online very simply. The system requirements can be divided into two parts - Frontend, which is the public outcome, and Backend, the system administration part. Both parts physically run at the same server, in both HW and SW, see Figure 10.1.

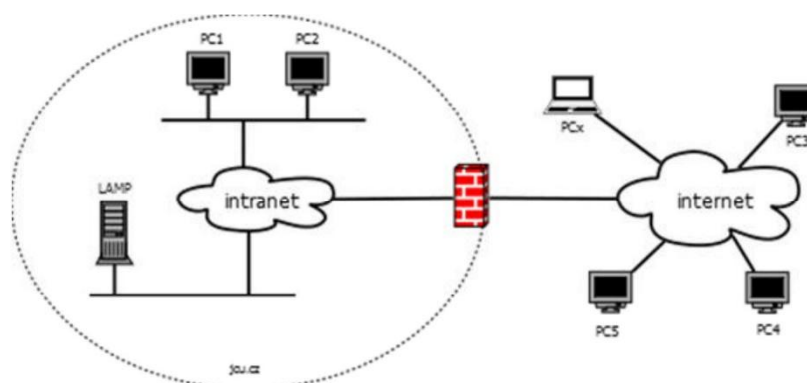


Figure 10.1: BSC infrastructure.

The system operates three levels of user authorization, see Table 25.

	superadmin	admin	User
User administration	yes	no	no
Log administration (access, activity)	yes	no	no
LOV administration	yes	yes	no
Recording upload and download	yes	yes	yes
Self-user profile management	yes	yes	yes

Table 25: User access levels.

When uploading data (sound recordings), the files are sent to the server using HTTP protocol. After a successful saving into the server file system, only the records of the recording attributes such as its size, type and file link are filed into the record chart. The names are generated with a md5 cipher so the file cannot be easily addressed though HTTP.

10.3 Application functionality

The necessary functions resulted from the analysis carried out among the future users, ornithologists and engineers responsible for system administration.

10.3.1 Backend

1. Access only for registered users according to their roles - superadmin, admin, user
2. Enable the superadmins manage user accounts
3. Catalogues individuals (birds), connect them to the table of keys, especially the order, family, species
4. Enable system admins edit data
5. Catalogue individual recordings
6. Enable system users to upload such recordings
7. Enable system users to edit their user accounts (password, contacts, etc.)
8. English is the primary language with the option to switch into Czech.

Frontend

1. Public (unauthorized) approach - web pages
2. Data aggregation at different levels (taxonomy, order, location)
3. Possibility to search for and sort data within these aggregations
4. Possibility to download the chosen recordings (sounds) from the web.
5. English is the primary language with the option to switch into Czech.

10.4 Database model

The data model was created after the initial discussion with the project participants and collecting of the basic requirements for the database functions, see Figure 10.2.

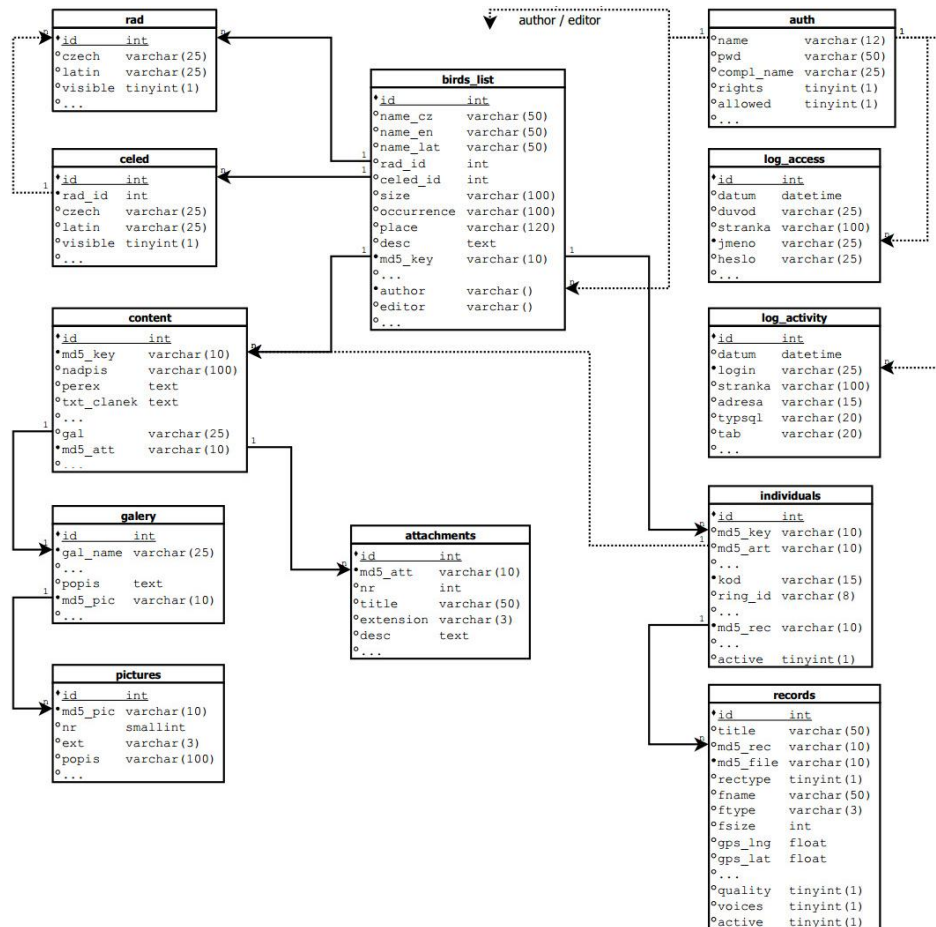


Figure 10.2: Relational model.

The list of the main charts and their utilization is given in Table 26.

Table name	Description
rad	List of values (LOV) Order
celed	LOV Family
birds_list	LOV Bird
auth	Users
content	Articles
gallery	Names of galleries
pictures	LOV Pictures
gallery attachments	LOV Attachments
individual	LOV Individuals
records	LOV Recording

Table 26: List of application tables.

10.5 General description

The aim was to create the whole application using Open Source. The obvious advantage of such solution is the costs, when there is no need to pay the licence fee. The main parts are: Web Server including an OS, and Database Server.

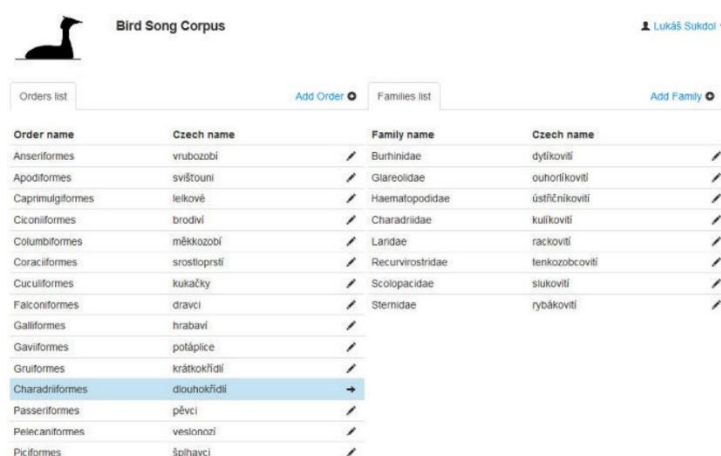
Essential HW requirement for the system is the availability of a fast Internet connection, since the programme response time and comfort of the manipulation: upload and download the large size files, are based on this feature.

Backend entities important for BSC administration and operation are:

- List of taxonomy
- LOV Birds
- Mass import
- LOV Individuals
- LOV Recordings
- LOV Gallery and pictures

10.5.1 List of taxonomy

The main content of the Administrator section is filling, and administration of List of values (LOV). One of the basic is *Taxonomy LOV*. For the needs of the Ornithologists it is sufficient to use just two levels: Orders, and Families.

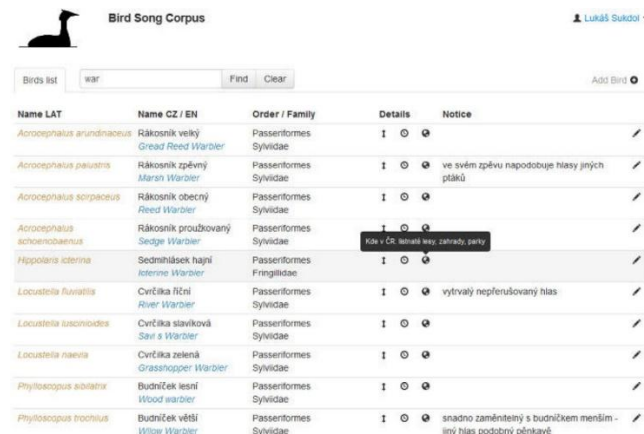


Order name	Czech name	Family name	Czech name
Anseriformes	vrubozobí	Burhinidae	dytíkovití
Apodiformes	svířtouni	Glareolidae	ouhoříkovití
Caprimulgiformes	lelkové	Haematopodidae	ústřčnickovití
Ciconiiformes	brodvní	Charadriidae	kulíkovití
Columbiformes	měkkozobí	Landae	rackovití
Coraciiformes	srostloprstí	Recurvirostridae	tenkozobcovití
Cuculiformes	kukačky	Scolopacidae	slukovití
Falconiformes	dravci	Sternidae	rybákovití
Galliformes	hrabaví		
Gaviiformes	potáplice		
Gruiformes	krátkokřídlí		
Charadriiformes	dlouhokřídlí		
Passeriformes	pěvci		
Pelecaniformes	vestionozí		
Piciformes	špihavci		

Figure 10.3: *LOV Administration.*

10.5.2 LOV Birds

The essential table of keys of the project is LOV Birds. The ambition of the project was not to create a complete database of known birds, but initially just birds that occur in the Czech Republic. Combining the information from public web pages and check by the scientists of the University of South Bohemia we managed to get the basic list of 224 of CZ with bound taxonomy.



Name LAT	Name CZ / EN	Order / Family	Details	Notice
<i>Acrocephalus arundinaceus</i>	Rákosník velký Great Reed Warbler	Passeriformes Sylviidae	1 0 0	
<i>Acrocephalus palustris</i>	Rákosník zpěvný Marsh Warbler	Passeriformes Sylviidae	1 0 0	ve svém zpěvu napodobuje hlasy jiných ptáků
<i>Acrocephalus scirpaceus</i>	Rákosník obecný Reed Warbler	Passeriformes Sylviidae	1 0 0	
<i>Acrocephalus schoenobaenus</i>	Rákosník proužkovaný Sedge Warbler	Passeriformes Sylviidae	1 0 0	
<i>Hippolais icterina</i>	Sedmháček hajní Icterine Warbler	Passeriformes Fringillidae	1 0 0	
<i>Locustella fluviatilis</i>	Črvička říční River Warbler	Passeriformes Sylviidae	1 0 0	vyrvalý nepřerušovaný hlas
<i>Locustella lusitana</i>	Črvička staviková Savi's Warbler	Passeriformes Sylviidae	1 0 0	
<i>Locustella naevia</i>	Črvička zelená Grasshopper Warbler	Passeriformes Sylviidae	1 0 0	
<i>Phylloscopus collybita</i>	Budníček lesní Wood warbler	Passeriformes Sylviidae	1 0 0	
<i>Phylloscopus trochilus</i>	Budníček větší Willow Warbler	Passeriformes Sylviidae	1 0 0	snadno zaměnitelný s budníčkem menším - jiný hlas podobný plénkavě

Figure 10.4: LOV Birds.

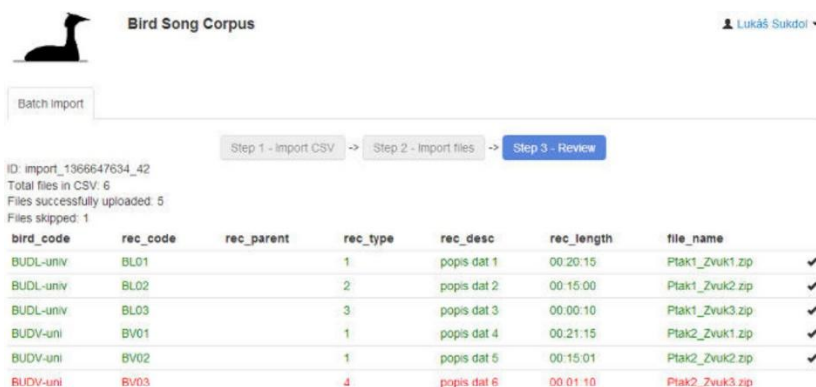
The key item of LOV is the Latin name, displayed in orange, that is together with the Czech name a compulsory item of the table of keys. Placing the mouse over the icon in the list shows some editable details - size of the bird, time and habitat in the Czech Republic.

10.5.3 Mass data import

Because of the need to import a very large amount of data initially, Mass Data import was implemented. That is a multi-step import, which requires preparation from the system administrator.

The actual import is performed in three steps:

1. A CSV file is selected. Its content is automatically analysed and the import information appears in the application.
2. If the file is OK, an instruction to upload the files appears. The administrator highlights the files designated for import and drags them onto the upload instruction. The files are item-by-item compared with the CVS records and are transferred to the server. If the transfer is successful, the data will be shown and the chart with imported records will update. Records with the same name as the uploaded items will be marked green.
3. By clicking the Proceed button the green marked recordings will be saved into the database and renamed with a coded name. In the end the import statistics appears, see Figure 10.5.



bird_code	rec_code	rec_parent	rec_type	rec_desc	rec_length	file_name
BUDL-uniV	BL01	1	popis dat 1	00:20:15	Plak1_Zvuk1.zip	✓
BUDL-uniV	BL02	2	popis dat 2	00:15:00	Plak1_Zvuk2.zip	✓
BUDL-uniV	BL03	3	popis dat 3	00:00:10	Plak1_Zvuk3.zip	✓
BUDV-uni	BV01	1	popis dat 4	00:21:15	Plak2_Zvuk1.zip	✓
BUDV-uni	BV02	1	popis dat 5	00:15:01	Plak2_Zvuk2.zip	✓
BUDV-uni	BV03	4	popis dat 6	00:01:10	Plak2_Zvuk3.zip	

Figure 10.5: Mass import, step 3. Green colour signs the correctly imported records.
Red colour signs an import error occurs.

10.5.4 LOV Individuals

To enable matching of the recordings with the particular bird, there is an aggregation element: chart Individuals, between the bird table of keys and table of recordings. Their relation is visible from the Data model (see Figure 10.2). It is possible to bound more individuals with one record in the bird table of keys, as well it does not have to be bound at all. An ornithologist might just want to enter an individual that had not been 100% identified yet, and bound it later. More recordings can be bound to one individual.

10.5.5 LOV Recordings

The table of recordings is the key chart of the whole system. Every record there determines, using a unique code, the name of the physical recording file in the file server storage. The record can be bound to the table of individuals, but this is not a condition.

For an easier handling of recording, the Location function was implemented. It is possible to match the GPS location with every record. This function is implemented using publically available functions API Google Maps. With one click into the map and click on the Save button the location record in database is updated. It is possible to navigate the map as in Google Maps, change the scale, and switch over between the display modes (Standard, Satellite, etc.). It is possible to display the location of every record with GPS position, see Figure 10.6. This list also takes into consideration the extract filter - it only displays the positions of the filtered part of the records.

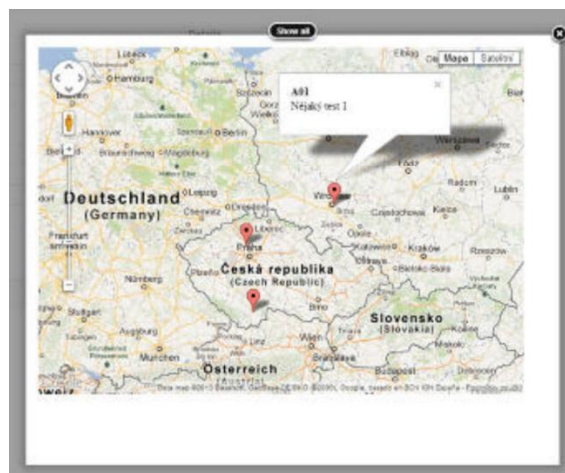


Figure 10.6: Navigate the map.

10.6 Contribution

The BSC will serve for data sharing between ornithologists; it makes comparing the achieved results easier for scientists, and so it helps to optimize the used methods. Contrary to speaker verification, the number of bird records is always limited. While creating a database for the human voice is theoretically unlimited and the researcher needs only “time and money“. The building up of a database of songs of a particular bird is strictly limited. Recording depends on season, weather, bird

mood and condition, accessibility of a nest, and many more factors, the last being the circumstantial. The bird recordings sharing avoids a researcher dependency on just a small amount of own data.

To date, the BSC is out of operation because it has been moved from České Budějovice to NTIS, Pilsen. If we follow the plan, the BSC will be operational at the end of 2016. We plan to introduce the BSC in an article after a pilot run.

10.7 Summary

Automatic recognition of animal sounds is a relatively new field that is supported by a small amount of resources. The reason might be a lower attractiveness of the topic and its limited applicability compared to the human voice recognition. Another reason might be the vast variety of species and sounds in the animal kingdom as opposed to the “only one” human speech. That is why there has not been any database similar to BSC created yet. The only available resources are web pages like e.g. *xenocanto.org*, that serve for saving just a limited amount of birdsongs, and do not allow more effective work with a large data amount.

11 Mashona mole-rat identification

11.1 Introduction

The zoologist of the University of South Bohemia, Faculty of Science have a one of the largest colony of the Mashona mole-rat over the world [DVO13]. Because the mole-rats live in the soil under the surface, its vocalization is very important for communication. Many experiments were performed to reveal the detail of its vocalization, communication, and the capability of the mutual recognition. Purpose of our experiment was to prove a possibility of the mole-rat individual identification based just on its mating calls. Zoologist of University of South Bohemia, Faculty of Science required system accuracy³ η to be at least $\eta \geq 65\%$. We used the Framework described in the chapter Development Framework.

11.2 Mashona mole-rat

The subterranean ecotype is unique in many aspects and has a great impact on the sensory biology of its inhabitants (reviewed in [BUR90], [FRA00]. Ubiquitous darkness prevents visual communication; reduced airflow limits olfactory sense. Acoustic signals can disperse here over medium distances of a few meters [HET86], [LAN07]. Under such conditions, vocalization becomes a crucial means of communication in mammals living underground

The Mashona mole-rat (*Fukomys darling*, in Czech *rypoš*) belongs to the African endemic rodent family of African mole-rats (Bathyergidae, Rodentia) see Figure 11.1. This species was known as *Cryptomys*. It is an herbivorous, social subterranean rodent. These mole-rats live in Zimbabwe, Mozambique and southern Malawi in small families up to nine animals. There is strict hierarchy in their families [GAB96], [BED13].



Figure 11.1: The Mashona mole-rat individual.

³ System accuracy η is defined as simply the ratio of correctly identified mole-rats to total amount of individuals.

11.3 Data

The data were recorded by Veronika Dvořáková, M.Sc., University of South Bohemia, Faculty of Science. She recorded two types of vocalizations of adult individuals based on the vocal repertoire of the Mashona mole-rat. Mating calls were used for experiments with individuality. The Snort sound was used for testing what information mole-rats are able to obtain from the signaller. The recordings were taken with the MD 431 II Sennheiser dynamic microphone (frequency range 40-16.000 Hz) and recorded with the Marantz card audio recorder PMD660 (sample frequency 44.1 kHz, resolution 16 bit). The mole-rats were simultaneously recorded using a Panasonic SDRH60EP-S camera to enable repeated checks of the testing sessions.

For individual identification, vocalizations of five dominant (breeding) females were used. The recordings were divided into 20-second tracks. Each recordings was named by a combination of a single letter (A,B,..E) represents a particular female, and a number represents a no. of recordings, e.g. "A_04", or "E_11".

Families or pairs were kept in terrariums with horticultural peat used as substrate and supplemented with plastic tubes as imitations of tunnels and flowerpots to simulate the nest see Figure 11.2. University of South Bohemia, Faculty of Science has mole-rat breeding that belong among the most representative collections of the underground mammals in the world. The experiments were carried out on these breeding and they serve to verify and add wild nature findings. The room was lighted in 12D/12L (lights on at 0700 h). The temperature was kept at 25 ± 1 °C. Animals were fed *ad libitum* with carrots, potatoes, apples and dry rodent food.



Figure 11.2: The Mashona mole-rat colony in the University of South Bohemia, Faculty of Science.

11.4 Vocalization

This species possess two types of mating calls; see Figure 11.3, both emitted mostly by females during courtship [DVO13]. This call is often produced in a series when one type alternates the other. A cluck is a very short vocalization, with the mean duration of 0.03 s. The range of

frequency is very low and it usually does not exceed 5 kHz. A shriek is a sound similar to a cluck, but it has a main frequency lower than a cluck and does not show a rising frequency towards the end.

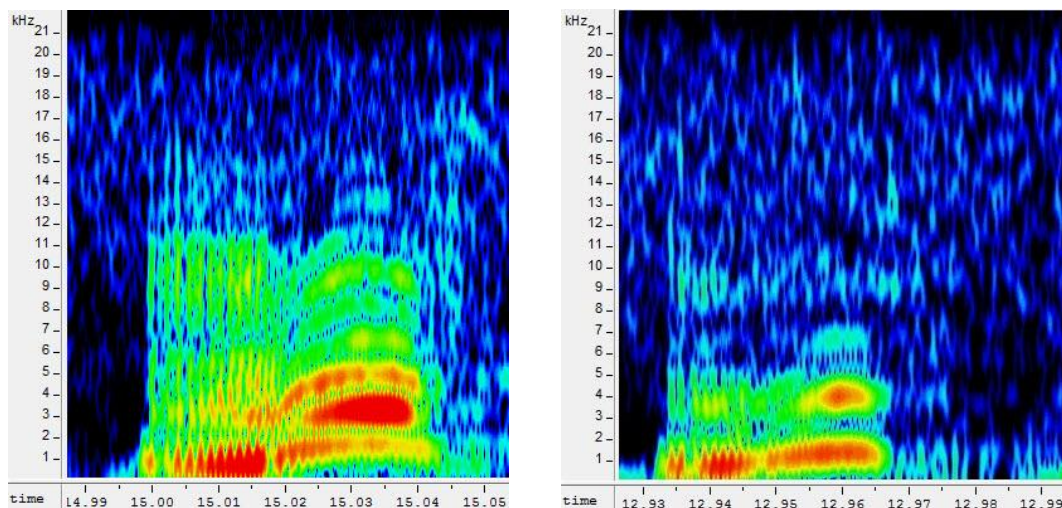


Figure 11.3: Spectrograms of the mating calls: cluck (left), and shriek (right).

11.5 Testing procedure

To ensure the objectivity of testing, two rounds which differed in the animal used for the UBM were performed for every target female.

Four older recordings from female B (recorded in 2010) were used. These four recordings were of poor quality as they were recorded by a different type of audio recorder (Sony Digital Audio Tape-corder TCD-D100) and probably negatively affected the results when used for estimating GMM of female B. On the other hand, these recordings did not have any impact when used for estimating UBM and in the verification phase. Hence, we removed these recordings from experiments when female B was used as the target animal, but used them when other females were used as target animals.

For the system, setup parameters see Table 27.

Parameter	Value
Window type	Hamming
Window length	20 ms
Window overlap	10 ms
Number of filters	23
Number of cepstral coefficients	20
Compute zero coefficient logE	Yes
High pass filter	250 Hz
Low pass filter	not used
Delta coefficients	Yes
VAD detector	No

Preemphase	0.99
Linear/MelFilter scale	Mel

Table 27: Parametrization set up values.

11.6 Results

Table 28 shows the overall success rate of individual identification. The number of obtained recordings (20 s soundtracks) varied from 10 to 40.

Target	# recordings	Experiment round	# test	# errors	Success rate (%)	Overall success rate (%)
A	33	round 1	1394	332	76.20%	82.30%
		round 2	1904	253	86.70%	
B	26	round 1	1067	223	79.10%	79.30%
		round 2	1012	207	79.50%	
C	40	round 1	1860	391	79.00%	79.00%
		round 2	1700	358	78.90%	
D	10	round 1	540	202	62.60%	59.20%
		round 2	1386	584	57.90%	
E	29	round 1	1190	192	83.90%	83.50%
		round 2	1260	212	83.20%	
OVERAL			13313	2954	77.80%	

Table 28: Mole-rat identification results.

The success rate of correct identification of particular individual varied between 59.2% and 83.5%. The lowest number of correct identification was obtained in female D, which unfortunately died at the time of recordings. I obtained only 10 soundtracks from this female, which is significantly less compared to any other female (26, 29, 33, 40). When results of recognition of female D are excluded, the overall success rate increases to 80.9%.

11.7 Contribution

The experiments confirmed the hypothesis that the mole-rats' vocalization also holds individuality identification [BED13]. As described above, this hypothesis is of high importance because of the mole-rats' life environment. Based on our knowledge, these are the first experiments dealing with mole-rat automatic individual identification. Although the experiments were introduced in [DVO13], it was not published yet. We plan together with Ema Hrouzkova⁴, Ph.D. to extend the experiments into winter 2016, and to proceed to submit the results for an article. The aim is to support the hypotheses about the mole-rat vocalization individuality described in [BED13].

11.8 Summary

The GMM-UBM based automatic system used for individual recognition was able to match the recordings to the particular female with an overall success rate of 77.8% (even more 80.9% if the

⁴ Department of Zoology, Faculty of Science, University of South Bohemia Ceske Budejovice, Czech Republic

female that died, i.e. with the lowest number of the sounds recorded is excluded). The overall percentage is thus high enough to show that the mating calls of the Mashona mole-rat can carry information about mole-rat individuality [DVO13].

Results demonstrated that mole-rats chose to follow the sound of the subordinate male. Females preferred subordinate, probably less dangerous individuals. Avoiding dominant males reflects the experiences from one's own family where the dominant male is not the one with which to interact.

12 Conclusion

Our thesis deals with automatic recognition and identification of bird individuals. The main goals as set in the Thesis goals chapter were fulfilled:

First, we designed and tested ARSBI algorithms and methods for a chiffchaff individual identification using as-is recordings i.e. live recordings made by ornithologist in nature without any pre-processing. The accuracy of the designed ARSBI reached 78.5%. Thus, the accuracy $\eta \geq 70\%$ required by the ornithologists of University of South Bohemia, Faculty of Science was met (see section 1.1 Overview).

Second, we propose a new feature extraction optimized for bird song. A new filter bank distribution BAF optimized for bird vocalization was designed. The EER improvements achieved using the test data were BAF to Linear 1.94% and BAF to Mel 4.42%. For the altered distribution BAF 1/3, an improvement of 3.56% to linear and 6.03% to Mel were reached.

Further the partial goals of our thesis were reached:

A new Bird Audiogram Unified Equation was found for mathematical expression of audiograms. We discovered the coefficients for forty-seven species and for four aggregated audiograms (all birds, Non-Passeriformes, Passeriformes, and Strigiformes).

A BSC database was created for storing the annotated recordings of bird songs. The aim of the authors is to collect the highest possible quantity of recordings annotated by independent ornithologists. After being filled with data, BSC enables to carry out experiments with both ARSBI and ARSBS, and with techniques belonging to the State of the Art. The desired step would be training of matrix iVector so that similar or even better results as for training of human speech. The planned BSC live launch in the end of 2016.

Experiments of bird identification using iVector were carried out. The system correctly identified nine birds of thirteen bird (69.2%). Based on our knowledge these are the first experiments dealing with bird individual identification by iVectors. Although the small scope of the experiments, we proved iVectors identification ability for Bird Individual Identification on the Closed Set.

Data Merging method was designed for improving the accuracy of identifying ARSBI and ARSBS. An accuracy improvement varied between 0.1% and 9.9%, and 3.0% in average. Although the method cannot be apprehended as a universal tool for accuracy improvement, it could be useful in cases we battle with insufficient number of data.

The last activity described in the thesis is identification of the mole-rats individuals. The success rate of individual identification varied between 59.2% and 83.5%, and 77.8% in average. Based on our knowledge these are the first experiments dealing with mole-rat automatic individual identification. Even it may seem the activity has no connection with the main topic, we include it into the thesis mainly because we used the identical Matlab Tool as in the whole research. That supports the idea, that the identification of animal individuals can be used more or less universally for any species, also for those with limited vocalization (number, frequency). The only limitation is the realistic chance of collecting sufficient amount of quality recordings.

It was proved that the ARSBI methods and algorithms we have introduced could be utilized for non-contact identification of birds, in our case for chiffchaff. At the same time, it was verified that ARSBI could be used not only for birds, but also for other animal species. We hope that the presented thesis brings new knowledge that may lead to the creation of a universal system of automatic recognition of birds in the wilderness (ARSBI and ARSBS). We believe that this method shall be an essential contribution to the study of the diverse and colourful world of birds, the research of which is now limited by the restricted possibilities of ringing or DNA testing.

12.1 Application of an automatic bird identification

Automatic bird identification offers a wide spectrum of applications, for instance:

Territory survey. A researcher installs automatic record machines near nests, triggered when the level of a bird song exceeds a limit. The recorders do not need an operator. He or she downloads records after a while then uses them for automatic recognition. At present, similar systems are used for night bird recordings.

Migration birds mapping. Ornithologists from different countries could share the data. From these records, a database of individual birds could be established after creating a precise bird model set. Then every user of this registry may be verified if the recorded bird is included in the registry.

Inaccessible breeding grounds observation. After installation of automatic recorders, data can be collected automatically. After a while, an ornithologist marks the recorders down. The recorders then do not need an operator: either they can run continually or they are triggered after a set volume levels is exceeded (Intensity Trigger, Limiter level). There are starting some projects having tens of recorders record continually for several days or weeks, see for example CIBRA project (University of Pavia). However, the major setback is the processing of the collected material that is dealt with manually at present, with only partial automatization. Full automatization would mean an essential breakthrough for the data processing.

Environmental protection. Thanks to a more sophisticated way of mapping the way of bird life, more accurate information about the life and habits of birds shall be gained. More accurate estimations about the possible influence of human activity on the individual species shall be provided. Specific cases may be, for example, the locality of NP Šumava, highway constructions, planning buildings near nesting sites, etc.

Protection of air traffic. Air traffic is very often put in danger because of birds. More detailed observation together with automatically evaluated records shall determine migration habits of individuals more accurately.

Others. As a matter of interest, we would like to mention also the area, where according to the author's opinion, automatic recognition can also be useful: for the leisurely ornithologist activities, in particular, the competitions of canary breeders. The canary songs are evaluated by the referees. The canaries are then evaluated with points according to the given criteria, as for example in gymnastics, and there is a final ranking in the end. The referees can never be fully objective; the automatic recognition can thus serve as additional means for independent evaluation. That can contribute to a better objectivity of such competitions.

12.2 Future work

Many interesting areas appeared that deserve a more detailed research. The topics we want to study in the future are:

- An automatic identification of bird individuals throughout months even years.
- To define parameters for maximum result achievement.
- To determine the required minimum of the recordings for which the system will still be capable of identifying a bird individual as well as the minimum number of recordings to achieve acceptable results (in the sense of recording quality, amount of data, etc.). For every ornithologist it is essential to know how many bird songs or recordings of certain quality have to be recorded to obtain a reasonable recognition accuracy of the ASRBI.
- To specify the influences and obstacles that would affect the ARSBI ability.
- Expansion of the mole-rat individual identification experiments, based just on its mating calls.
- BSC deployment and operation.

12.3 Personal note

Ornithology is a traditional scientific field. No matter how developed the opportunities of automatization and data processing are today, I still find ornithology has not acquired much in utilizing of such technologies. One of the reasons may also be its conservative approach: *why should we do it in a different way, it had been done for years and it works*. I have noticed this opinion at many ornithologists. They promptly refuse the automatic recognition as something not useful at all, because ringing and binoculars do the job. Unfortunately, more advanced methods (cepstral analysis, wavelet analysis, correlation analysis, etc.). At present there are also not many software solutions tailored to animal vocalization processing. For bird song analysis and gaining information from recordings, a spectrogram analysis is widely used in ornithology. Spectrograms are sometimes, with regret, perceived as the universal source of information about vocalization. The effort to mine features from the spectrum thus results, features that are not possible to gain at all or get inaccurately (pitch, vocal unique parameters, humour, age of an animal, etc.). In addition, the insufficient knowledge of frequency analysis appears and related restrictions (e.g. relation between n-point DFT and the signal length, Nyquist frequency, aliasing, quantization error, loss of time information during time-to-frequency transformation, etc.). Similarly ornithology approaches to signal filtering. Spectrum is most often understood as a picture, where the filtration is made by a simple cut out of a part of the picture, in other words the unrequired part of the frequency pattern. Possible restrictions (filter quality, steepness, wavelet, spectre leakage, frequency dependence of phase shift, etc.) are not taken into consideration, resulting from a poor knowledge of technical principles.

One of my colleagues, ornithologist, worked on the spectrographic recordings of bird singing and looked for mutual differences as part of his thesis. He dedicated one whole year of work to this study. I am not afraid to say that the automatization would shorten this activity to hours.

The above-described approach of ornithologists towards technology is understandable. Ornithology is a complex and traditional science, including a lot of knowledge that is necessary to acquire if a person wants to work on a certain advanced level. It is a humanities field requiring a different way of thinking compared to technical fields. Ornithologist understand the IT technology as the tool for getting the required information, and do not analyse its principles. A technician would most probably not be interested in a top-level ornithology, as such level requires memorizing a huge amount of encyclopaedic knowledge.

Thanks to the appearance of the new generation of scientists, we can observe a partial deflection from such philosophy. The young generation takes tablets, smartphones, and social networks as an everyday part of life and sees the potential of those technologies.

Another reason for the minimum penetration of automatization into ornithology I consider the distance or even fear zoologists have towards technologies. It is apparent at any humanitarian person, and every technician is more than familiar with it. The orientation of zoologists and any living person is simply different from that of a geek whose life is filled by gadgets and modern technology toys. Such differences are an endless source of many films and TV series (Beautiful Mind), music albums (Tata Boys Nanoalbum) or books (Surely You're Joking, Mr. Feynman! Adventures of a Curious Character).

In my opinion, the interdisciplinary cooperation can bring, apart from an original view of the problems, another positive side and that is the mutual enrichment. The technician helps with implementation of modern technologies. The scientist shows the technician that life is about not only zeros and ones, and that nature offers endless variety. I fully respect ornithology and appreciate the opportunity of the cooperation. Because of the nature of my job I consider myself an engineer, I am glad my colleagues - ornithologists - help me mediate an original view of the world.

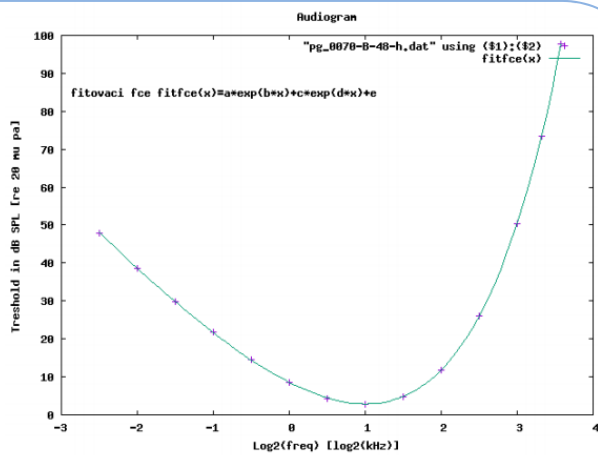
13 Appendix

The stated outputs were obtained from *GNUPlot* software, see section 7.2. The first page contains aggregate audiograms as shown in Table 18 (All Birds, Non-Passeriformes, Passeriformes, and Strigiformes). The remaining parts contain audiograms for the particular analysed species. Their complete list as in Table 17.

Legend

pg_0070-B-48 :
all birds ()

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 48: Audiogram all birds ()

After 1442 iterations the fit converged.
final sum of squares of residuals : 0.11888
rel. change during last iteration : -6.85045e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.11493
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0132089

Final set of parameters	Asymptotic Standard Error
a = 11.8718	+/- 0.6325 (5.328%)
b = 0.734457	+/- 0.009855 (1.342%)
c = 524.5	+/- 151.5 (28.88%)
d = -0.0360242	+/- 0.00966 (26.82%)
e = -527.81	+/- 152.1 (28.81%)

[DOO02b] page number, audiogram type.

Audiogram parameters see Table 16.

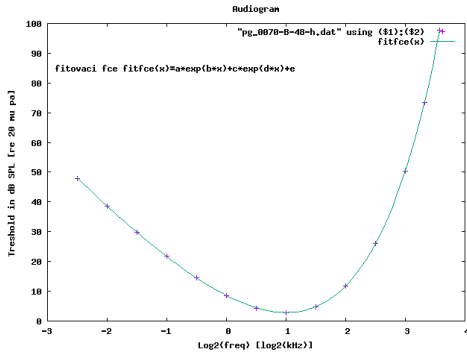
A final audiogram. The purple crosses represent localization points we put on the original audiogram. The coloured line is the graph of f_1 see Figure 7.3. Notice the frequency axis is in \log_2 scale.

GNUPlot iteration results for f_1 see equations (74), (75), and (76).

Final set of parameters see Table 17 and Table 18. Parameter asymptotic standard error.

pg_0070-B-48 :
all birds ()

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 48: Audiogram all birds ()

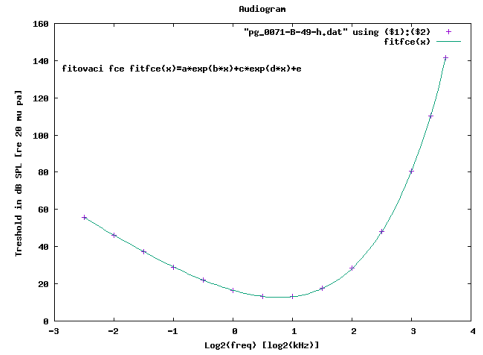
After 1442 iterations the fit converged.
final sum of squares of residuals : 0.11888
rel. change during last iteration : -6.85045e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.11493
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0132089

Final set of parameters	Asymptotic Standard Error
a = 11.8718	+/- 0.6325 (5.328%)
b = 0.734457	+/- 0.009855 (1.342%)
c = 524.5	+/- 151.5 (28.88%)
d = -0.0360242	+/- 0.00966 (26.82%)
e = -527.81	+/- 152.1 (28.81%)

pg_0071-B-49 :
01-12 (Non Passeriformes)

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 49: Audiogram 01-12 (Non Passeriformes)

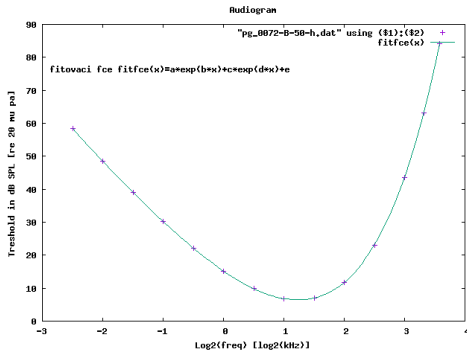
After 4276 iterations the fit converged.
final sum of squares of residuals : 0.0541087
rel. change during last iteration : -3.57858e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0775376
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00601208

Final set of parameters	Asymptotic Standard Error
a = 18.7657	+/- 0.6405 (3.413%)
b = 0.692153	+/- 0.005916 (0.8547%)
c = -2884.21	+/- 2526 (87.59%)
d = 0.0076418	+/- 0.006818 (89.22%)
e = 2881.98	+/- 2527 (87.68%)

pg_0072-B-50 :
13-35 (Passeriformes)

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 50: Audiogram 13-35 (Passeriformes)

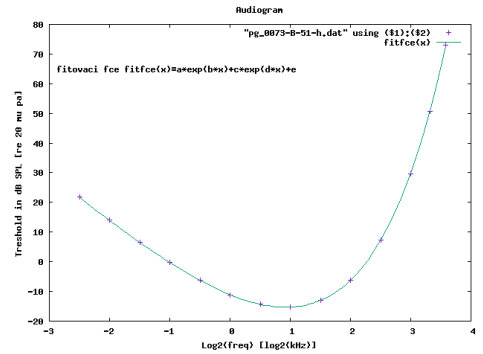
After 3740 iterations the fit converged.
final sum of squares of residuals : 0.0105172
rel. change during last iteration : -3.29397e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0341844
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00116857

Final set of parameters	Asymptotic Standard Error
a = 12.0993	+/- 0.2333 (1.928%)
b = 0.712241	+/- 0.003446 (0.4839%)
c = 1904.88	+/- 507.4 (26.64%)
d = -0.0110408	+/- 0.002869 (25.99%)
e = -1901.79	+/- 507.7 (26.69%)

pg_0073-B-51 :
36-47 (Strigiformes)

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 51: Audiogram 36-47 (Strigiformes)

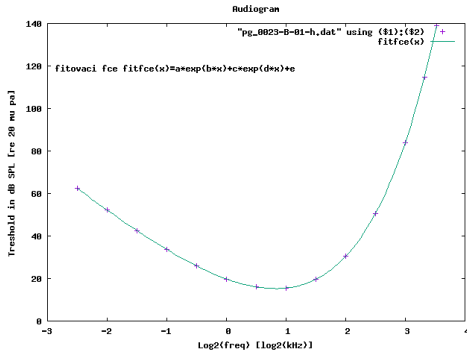
After 3726 iterations the fit converged.
final sum of squares of residuals : 0.0134215
rel. change during last iteration : -3.11217e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0386171
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00149128

Final set of parameters	Asymptotic Standard Error
a = 12.2142	+/- 0.265 (2.17%)
b = 0.710617	+/- 0.003876 (0.5455%)
c = 1379.51	+/- 465.1 (33.71%)
d = -0.012282	+/- 0.004028 (32.8%)
e = -1402.81	+/- 465.3 (33.17%)

pg_0023-B-01 :
Mallard Duck (*Anas platyrhynchos*)

15.57 2.00 0.32 5.22 1.28 4.91



Obrázek 1: Audiogram Mallard Duck (*Anas platyrhynchos*)

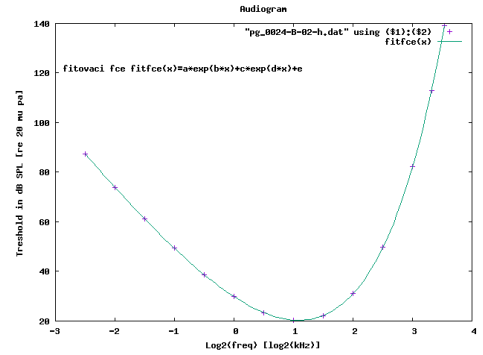
After 4316 iterations the fit converged.
final sum of squares of residuals : 0.155299
rel. change during last iteration : -1.20649e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.13136
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0172554

Final set of parameters	Asymptotic Standard Error
a = 19.6151	+/- 1.17 (5.966%)
b = 0.690624	+/- 0.01041 (1.507%)
c = -3256.3	+/- 4966 (152.5%)
d = 0.00727713	+/- 0.0113 (155.2%)
e = 3256.49	+/- 4967 (152.5%)

pg_0024-B-02 :
Australian Grey Swiftlet (*Collocalia spodiopygia*)

20.31 2.00 0.49 5.71 1.66 5.23



Obrázek 2: Audiogram Australian Grey Swiftlet (*Collocalia spodiopygia*)

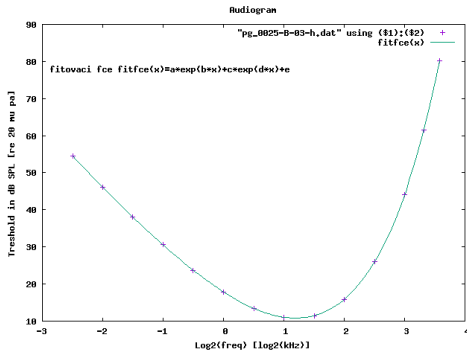
After 3747 iterations the fit converged.
final sum of squares of residuals : 0.0593927
rel. change during last iteration : -1.82682e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0812354
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00659919

Final set of parameters	Asymptotic Standard Error
a = 18.5348	+/- 0.5943 (3.206%)
b = 0.708868	+/- 0.005754 (0.8117%)
c = 2419.5	+/- 1091 (45.07%)
d = -0.0118689	+/- 0.005207 (43.87%)
e = -2408.24	+/- 1091 (45.31%)

pg_0025-B-03 :
Oilbird (*Steatornis caripensis*)

20.31 2.00 0.49 5.71 1.66 5.23



Obrázek 3: Audiogram Oilbird (*Steatornis caripensis*)

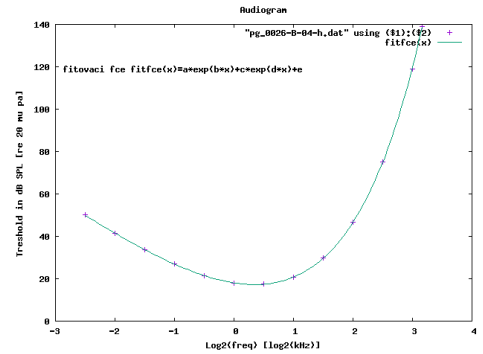
After 3963 iterations the fit converged.
final sum of squares of residuals : 0.00903014
rel. change during last iteration : -1.84354e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0316757
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00100335

Final set of parameters	Asymptotic Standard Error
a = 10.9649	+/- 0.2267 (2.067%)
b = 0.70453	+/- 0.003666 (0.5204%)
c = 1858.69	+/- 613.8 (33.02%)
d = -0.00974333	+/- 0.003147 (32.3%)
e = -1851.8	+/- 614 (33.16%)

pg_0026-B-04 :
Emu (*Dromaius novaehollandiae*)

17.39 1.41 0.20 4.04 0.88 3.85



Obrázek 4: Audiogram Emu (*Dromaius novaehollandiae*)

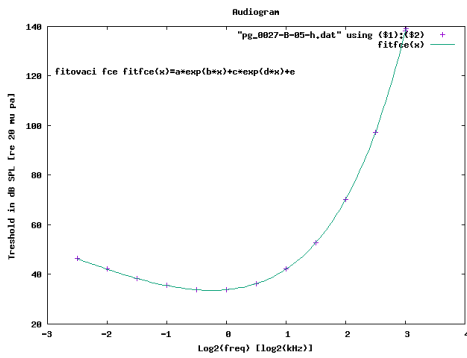
After 3429 iterations the fit converged.
final sum of squares of residuals : 0.709296
rel. change during last iteration : -9.9957e-06

degrees of freedom (FIT_NDF) : 8
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.297762
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0886621

Final set of parameters	Asymptotic Standard Error
a = 22.8571	+/- 3.698 (16.18%)
b = 0.699756	+/- 0.0299 (4.273%)
c = -2892.04	+/- 1.531e+04 (529.4%)
d = 0.00707593	+/- 0.03816 (539.3%)
e = 2887.24	+/- 1.531e+04 (530.4%)

pg_0027-B-05 :
Plains Wanderer (*Pedionomus torquatus*)

33.80 0.71 0.05 3.56 0.44 3.50



Obrázek 5: Audiogram Plains Wanderer (*Pedionomus torquatus*)

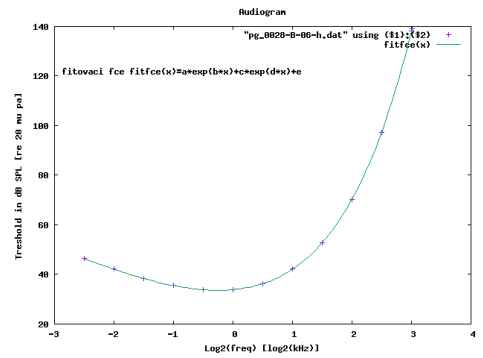
After 4241 iterations the fit converged.
final sum of squares of residuals : 0.0170089
rel. change during last iteration : -3.08153e-07

degrees of freedom (FIT_NDF) : 8
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0461098
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00212611

Final set of parameters	Asymptotic	Standard Error	
a	= 20.2978	+/- 0.6944	(3.421%)
b	= 0.689831	+/- 0.006341	(0.9192%)
c	= -1208.69	+/- 1387	(114.8%)
d	= 0.00979192	+/- 0.01154	(117.9%)
e	= 1222.24	+/- 1388	(113.6%)

pg_0028-B-06 :
Pigeon (*Columbia livia*)

16.90 1.41 5.67 0.13 5.80 5.67



Obrázek 6: Audiogram Pigeon (*Columbia livia*)

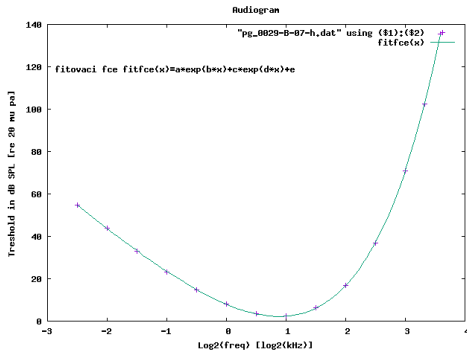
After 4241 iterations the fit converged.
final sum of squares of residuals : 0.0170089
rel. change during last iteration : -3.08153e-07

degrees of freedom (FIT_NDF) : 8
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0461098
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00212611

Final set of parameters	Asymptotic	Standard Error	
a	= 20.2978	+/- 0.6944	(3.421%)
b	= 0.689831	+/- 0.006341	(0.9192%)
c	= -1208.69	+/- 1387	(114.8%)
d	= 0.00979192	+/- 0.01154	(117.9%)
e	= 1222.24	+/- 1388	(113.6%)

pg_0029-B-07 :
American Kestrel (*Falco sparverius*)

2.42 2.00 0.36 5.25 1.37 4.89



Obrázek 7: Audiogram American Kestrel (*Falco sparverius*)

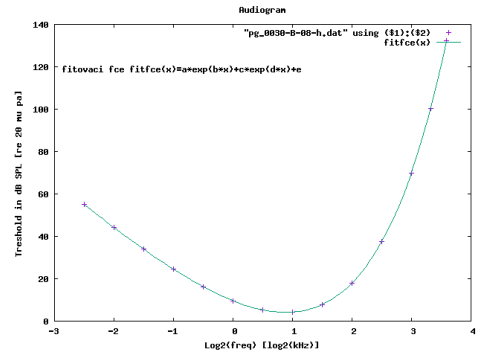
After 1629 iterations the fit converged.
final sum of squares of residuals : 0.264687
rel. change during last iteration : -2.31468e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.171493
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0294097

Final set of parameters	Asymptotic	Standard Error	
a	= 23.9205	+/- 1.911	(7.99%)
b	= 0.662058	+/- 0.01315	(1.987%)
c	= -782.247	+/- 281.2	(35.94%)
d	= 0.035272	+/- 0.01397	(39.61%)
e	= 766.256	+/- 283	(36.94%)

pg_0030-B-08 :
European Sparrowhawk (*Accipiter nisus*)

4.27 2.00 0.35 5.39 1.37 5.04



Obrázek 8: Audiogram European Sparrowhawk (*Accipiter nisus*)

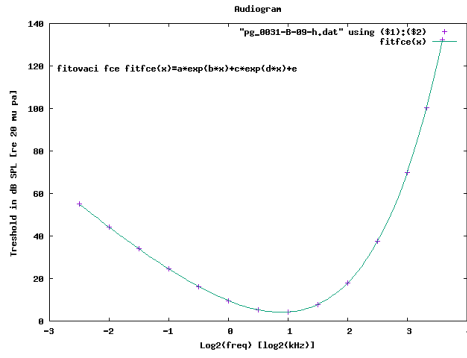
After 4517 iterations the fit converged.
final sum of squares of residuals : 0.0422705
rel. change during last iteration : -7.13187e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0685327
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00469673

Final set of parameters	Asymptotic	Standard Error	
a	= 19.8187	+/- 0.5828	(2.941%)
b	= 0.687179	+/- 0.005069	(0.7377%)
c	= -3101.63	+/- 2049	(66.06%)
d	= 0.00801726	+/- 0.005401	(67.36%)
e	= 3091.44	+/- 2049	(66.29%)

pg_0031-B-09 :
Bobwhite Quail (*Colinus virginianus*)

13.15 2.00 2.13 8.70 1.35 6.57



Obrázek 9: Audiogram Bobwhite Quail (*Colinus virginianus*)

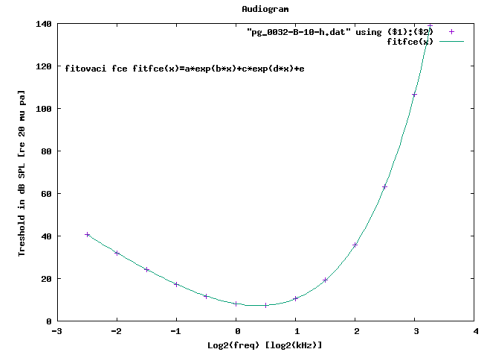
After 4517 iterations the fit converged.
final sum of squares of residuals : 0.0422705
rel. change during last iteration : -7.13187e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0685327
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00469673

Final set of parameters	Asymptotic Standard Error
a = 19.8187	+/- 0.5828 (2.941%)
b = 0.687179	+/- 0.005069 (0.7377%)
c = -3101.63	+/- 2049 (66.06%)
d = 0.00801726	+/- 0.005401 (67.36%)
e = 3091.44	+/- 2049 (66.29%)

pg_0032-B-10 :
Chicken (*Gallus gallus*)

7.37 1.41 0.20 4.10 0.91 3.90



Obrázek 10: Audiogram Chicken (*Gallus gallus*)

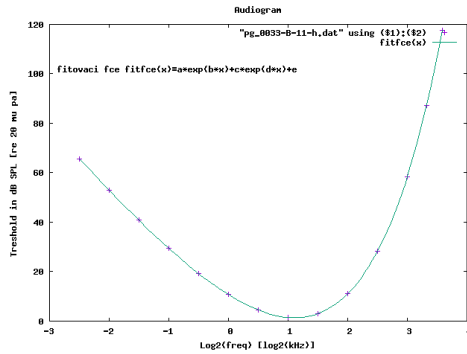
After 3893 iterations the fit converged.
final sum of squares of residuals : 0.0983042
rel. change during last iteration : -1.11988e-06

degrees of freedom (FIT_NDF) : 8
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.110851
variance of residuals (reduced chisquare) = WSSR/ndf : 0.012288

Final set of parameters	Asymptotic Standard Error
a = 23.155	+/- 1.283 (5.539%)
b = 0.692576	+/- 0.01002 (1.446%)
c = -2599.53	+/- 4174 (160.6%)
d = 0.00799712	+/- 0.01311 (163.9%)
e = 2584.57	+/- 4175 (161.5%)

pg_0033-B-11 :
Japanese Quail (*Coturnix coturnix japonica*)

1.40 2.00 0.47 5.90 1.66 5.43



Obrázek 11: Audiogram Japanese Quail (*Coturnix coturnix japonica*)

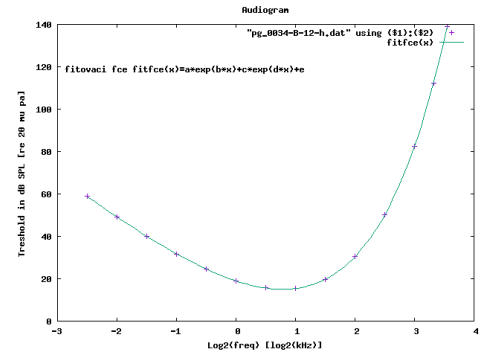
After 3957 iterations the fit converged.
final sum of squares of residuals : 0.0164719
rel. change during last iteration : -1.07258e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.042781
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00183021

Final set of parameters	Asymptotic Standard Error
a = 17.6339	+/- 0.3036 (1.722%)
b = 0.705661	+/- 0.003057 (0.4332%)
c = 2703.4	+/- 759.5 (28.1%)
d = -0.0101645	+/- 0.002791 (27.46%)
e = -2710.31	+/- 759.8 (28.03%)

pg_0034-B-12 :
Turkey (*Meleagris gallopavo*)

15.43 2.00 0.29 5.25 1.22 4.96



Obrázek 12: Audiogram Turkey (*Meleagris gallopavo*)

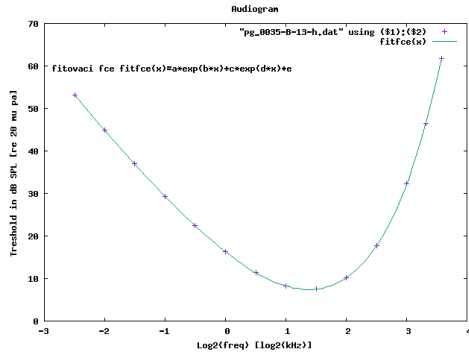
After 4323 iterations the fit converged.
final sum of squares of residuals : 0.0935137
rel. change during last iteration : -3.53741e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.101933
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0103904

Final set of parameters	Asymptotic Standard Error
a = 18.9526	+/- 0.8856 (4.672%)
b = 0.690194	+/- 0.008121 (1.177%)
c = -2959.8	+/- 3515 (118.7%)
d = 0.00754701	+/- 0.009129 (121%)
e = 2959.84	+/- 3516 (118.8%)

pg_0035-B-13 :
American Robin (*Turdus migratorius*)

7.49 2.83 0.34 8.73 1.72 8.39



Obrázek 13: Audiogram American Robin (*Turdus migratorius*)

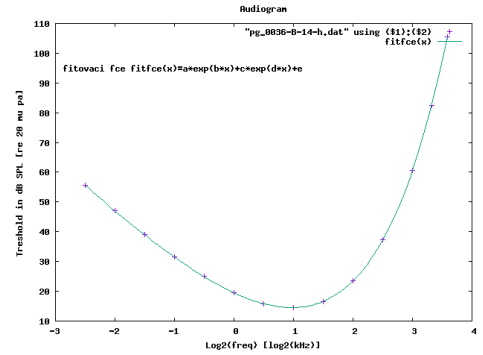
After 3941 iterations the fit converged.
final sum of squares of residuals : 0.00725465
rel. change during last iteration : -7.74861e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0283914
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000806073

Final set of parameters	Asymptotic Standard Error
a = 9.29235	+/- 0.2004 (2.157%)
b = 0.706929	+/- 0.003832 (0.5421%)
c = 1876.33	+/- 586.1 (31.23%)
d = -0.00941182	+/- 0.002878 (30.58%)
e = -1869.34	+/- 586.3 (31.36%)

pg_0036-B-14 :
Blue Jay (*Cyanocitta cristata*)

14.46 2.00 0.28 6.31 1.33 6.03



Obrázek 14: Audiogram Blue Jay (*Cyanocitta cristata*)

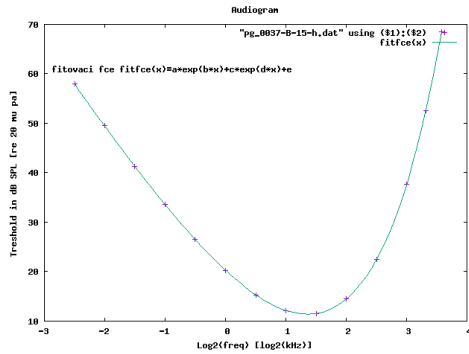
After 4101 iterations the fit converged.
final sum of squares of residuals : 0.00880649
rel. change during last iteration : -6.85812e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.031281
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000978499

Final set of parameters	Asymptotic Standard Error
a = 13.2034	+/- 0.2243 (1.699%)
b = 0.704129	+/- 0.003012 (0.4278%)
c = 1889.04	+/- 595.9 (31.55%)
d = -0.00983117	+/- 0.003033 (30.85%)
e = -1882.75	+/- 596.2 (31.66%)

pg_0037-B-15 :
Brown-headed Cowbird (*Molothrus ater*)

11.50 2.83 0.35 8.50 1.72 8.15



Obrázek 15: Audiogram Brown-headed Cowbird (*Molothrus ater*)

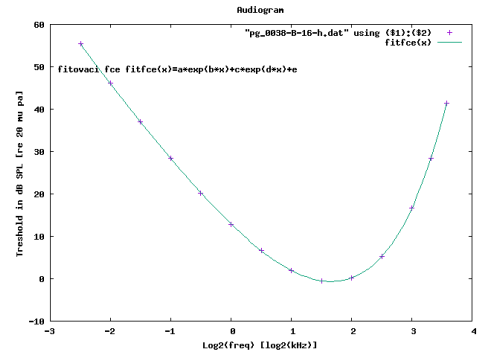
After 3892 iterations the fit converged.
final sum of squares of residuals : 0.00880884
rel. change during last iteration : -9.09264e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0312851
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000978759

Final set of parameters	Asymptotic Standard Error
a = 9.68272	+/- 0.2215 (2.288%)
b = 0.70611	+/- 0.004063 (0.5755%)
c = 1810.32	+/- 573.6 (31.68%)
d = -0.00999705	+/- 0.003096 (30.97%)
e = -1799.75	+/- 573.8 (31.88%)

pg_0038-B-16 :
Bullfinch (*Pyrrhula pyrrhula*)

-0.50 2.83 0.48 10.20 2.21 9.72



Obrázek 16: Audiogram Bullfinch (*Pyrrhula pyrrhula*)

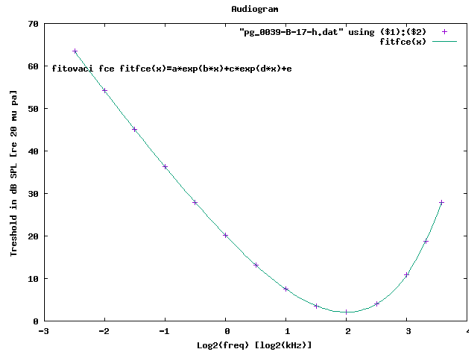
After 6254 iterations the fit converged.
final sum of squares of residuals : 0.00441906
rel. change during last iteration : -9.99889e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0221587
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000491007

Final set of parameters	Asymptotic Standard Error
a = 8.73047	+/- 0.1656 (1.897%)
b = 0.700806	+/- 0.00334 (0.4766%)
c = 5081.55	+/- 2713 (53.38%)
d = -0.00389487	+/- 0.00206 (52.9%)
e = -5077.4	+/- 2713 (53.43%)

pg_0039-B-17 :
Chipping Sparrow (*Spizella passerina*)

2.06 4.00 0.59 12.90 2.75 12.31



Obrázek 17: Audiogram Chipping Sparrow (*Spizella passerina*)

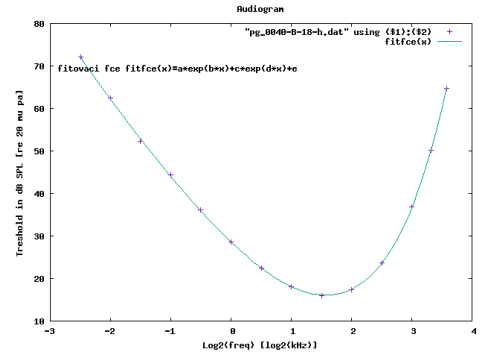
After 3713 iterations the fit converged.
final sum of squares of residuals : 0.0118744
rel. change during last iteration : -2.169e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0363232
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00131937

Final set of parameters	Asymptotic	Standard Error	
a	= 6.40234	+/- 0.249	(3.889%)
b	= 0.711876	+/- 0.006937	(0.9744%)
c	= 2095.93	+/- 775.7	(37.01%)
d	= -0.00919708	+/- 0.003334	(36.25%)
e	= -2082.22	+/- 775.9	(37.26%)

pg_0040-B-18 :
Common Canary (*Serinus canarius*)

15.98 2.83 0.47 9.37 2.08 8.90



Obrázek 18: Audiogram Common Canary (*Serinus canarius*)

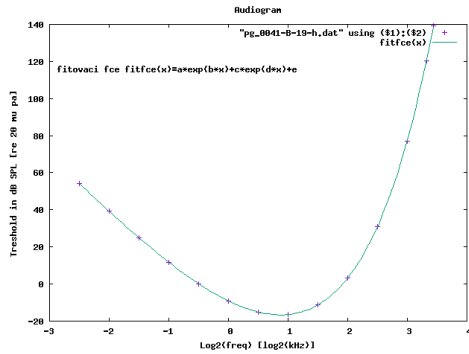
After 1175 iterations the fit converged.
final sum of squares of residuals : 0.728832
rel. change during last iteration : -9.54049e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.284572
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0809813

Final set of parameters	Asymptotic	Standard Error	
a	= 6.99756	+/- 1.342	(19.17%)
b	= 0.756999	+/- 0.03619	(4.781%)
c	= 438.35	+/- 252	(57.48%)
d	= -0.0427403	+/- 0.02259	(52.86%)
e	= -416.765	+/- 253.3	(60.77%)

pg_0041-B-19 :
Common Crow (*Corvus brachyrhynchos*)

-16.41 2.00 0.47 4.57 1.46 4.10



Obrázek 19: Audiogram Common Crow (*Corvus brachyrhynchos*)

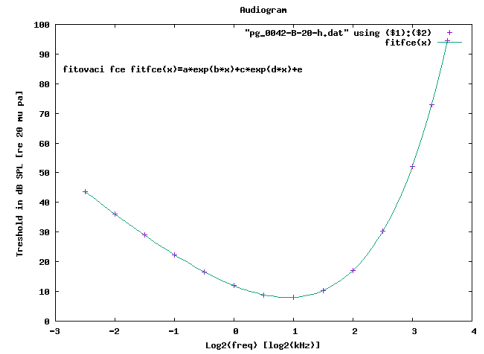
After 4196 iterations the fit converged.
final sum of squares of residuals : 0.77613
rel. change during last iteration : -1.2792e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.293661
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0862367

Final set of parameters	Asymptotic	Standard Error	
a	= 27.3755	+/- 2.828	(10.33%)
b	= 0.692923	+/- 0.01826	(2.635%)
c	= -4876.45	+/- 1.238e+04	(253.9%)
d	= 0.00710128	+/- 0.01835	(258.4%)
e	= 4840	+/- 1.238e+04	(255.8%)

pg_0042-B-20 :
European Starling (*Sturnus vulgaris*)

8.00 2.00 0.23 6.43 1.20 6.20



Obrázek 20: Audiogram European Starling (*Sturnus vulgaris*)

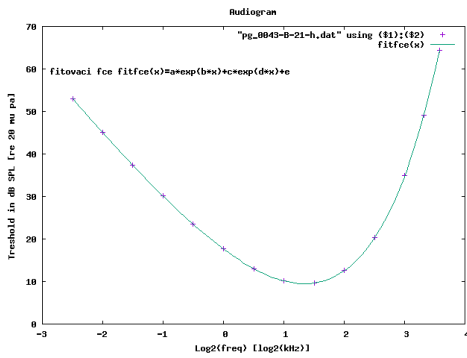
After 4154 iterations the fit converged.
final sum of squares of residuals : 0.00906184
rel. change during last iteration : -4.718e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0317312
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00100687

Final set of parameters	Asymptotic	Standard Error	
a	= 12.2944	+/- 0.2284	(1.858%)
b	= 0.703478	+/- 0.003292	(0.4679%)
c	= 1658.01	+/- 588.1	(35.47%)
d	= -0.00997581	+/- 0.003459	(34.67%)
e	= -1658.35	+/- 588.3	(35.47%)

pg_0043-B-21 :
Field Sparrow (*Spizella pusilla*)

9.61 2.83 0.32 8.65 1.65 8.33



Obrázek 21: Audiogram Field Sparrow (*Spizella pusilla*)

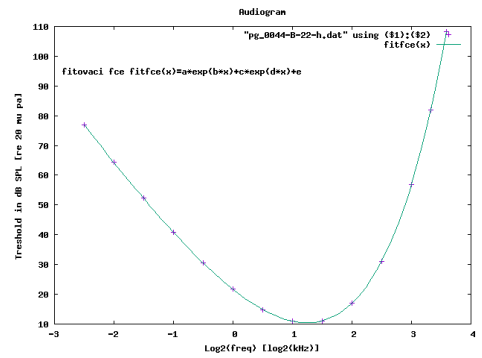
After 3920 iterations the fit converged.
final sum of squares of residuals : 0.0090912
rel. change during last iteration : -9.46687e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0317826
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00101013

Final set of parameters	Asymptotic Standard Error
a = 9.21394	+/- 0.2243 (2.435%)
b = 0.706816	+/- 0.004326 (0.612%)
c = 1758.36	+/- 619.5 (35.23%)
d = -0.009687	+/- 0.003339 (34.47%)
e = -1749.95	+/- 619.7 (35.41%)

pg_0044-B-22 :
Fire finch (*Lagonosticta senegala*)

10.89 2.00 0.50 6.49 1.79 5.99



Obrázek 22: Audiogram Fire finch (*Lagonosticta senegala*)

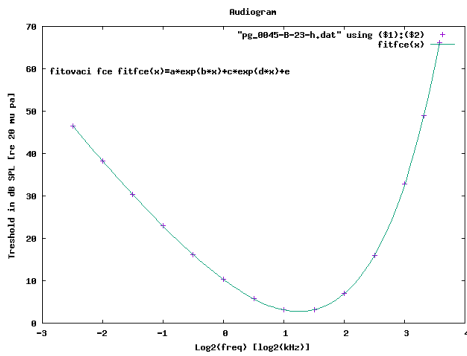
After 3892 iterations the fit converged.
final sum of squares of residuals : 0.0180105
rel. change during last iteration : -9.68491e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0447344
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00200117

Final set of parameters	Asymptotic Standard Error
a = 15.7395	+/- 0.3181 (2.021%)
b = 0.705706	+/- 0.003587 (0.5083%)
c = 2847.53	+/- 913.9 (32.09%)
d = -0.00947517	+/- 0.002976 (31.41%)
e = -2841.59	+/- 914.2 (32.17%)

pg_0045-B-23 :
Great tit (*Parus major*)

3.07 2.00 0.32 8.17 1.60 5.02



Obrázek 23: Audiogram Great tit (*Parus major*)

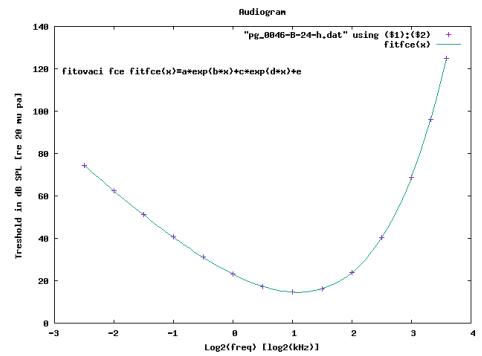
After 3908 iterations the fit converged.
final sum of squares of residuals : 0.00988154
rel. change during last iteration : -2.38496e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0331353
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00109795

Final set of parameters	Asymptotic Standard Error
a = 10.2496	+/- 0.2355 (2.298%)
b = 0.705533	+/- 0.004079 (0.5781%)
c = 1783.18	+/- 615.8 (34.54%)
d = -0.00993607	+/- 0.003355 (33.77%)
e = -1783.19	+/- 616.1 (34.55%)

pg_0046-B-24 :
House finch (*Carpodacus mexicanus*)

14.55 2.00 0.44 6.00 1.61 5.56



Obrázek 24: Audiogram House finch (*Carpodacus mexicanus*)

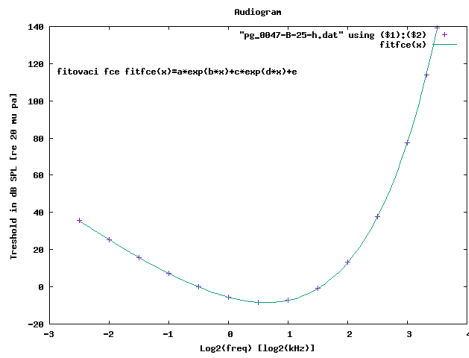
After 4048 iterations the fit converged.
final sum of squares of residuals : 0.0193257
rel. change during last iteration : -2.18601e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0463389
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0021473

Final set of parameters	Asymptotic Standard Error
a = 16.7576	+/- 0.3323 (1.983%)
b = 0.704423	+/- 0.003515 (0.499%)
c = 2746.77	+/- 970.1 (35.32%)
d = -0.00937526	+/- 0.003241 (34.57%)
e = -2740.42	+/- 970.4 (35.41%)

pg_0047-B-25 :
House Sparrow (*Passer domesticus*)

-8.31 1.41 0.29 4.55 1.13 4.27



Obrázek 25: Audiogram House Sparrow (*Passer domesticus*)

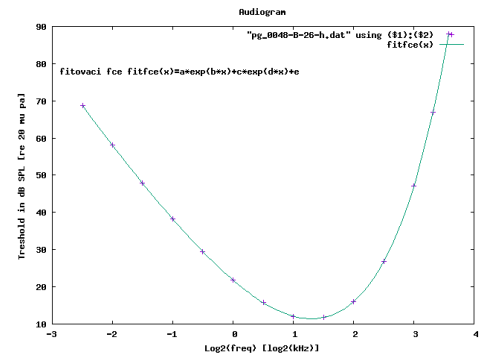
After 4147 iterations the fit converged.
final sum of squares of residuals : 0.309682
rel. change during last iteration : -1.57487e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.185497
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0344091

Final set of parameters	Asymptotic Standard Error
a = 22.2117	+/- 1.663 (7.488%)
b = 0.693419	+/- 0.01314 (1.895%)
c = -3497.61	+/- 8035 (229.7%)
d = 0.00683571	+/- 0.01597 (233.6%)
e = 3469.87	+/- 8037 (231.6%)

pg_0048-B-26 :
Pied Flycatcher (*Ficedula hypoleuca*)

11.70 2.83 0.44 7.34 1.79 6.90



Obrázek 26: Audiogram Pied Flycatcher (*Ficedula hypoleuca*)

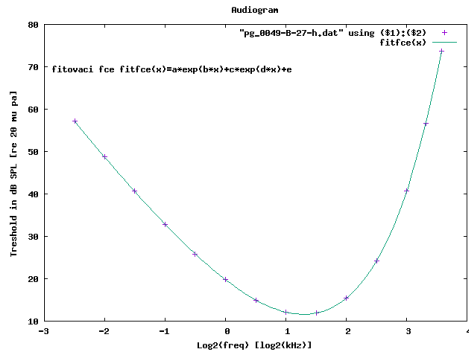
After 3929 iterations the fit converged.
final sum of squares of residuals : 0.0129479
rel. change during last iteration : -3.14935e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0379296
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00143866

Final set of parameters	Asymptotic Standard Error
a = 12.6422	+/- 0.2687 (2.125%)
b = 0.70622	+/- 0.003774 (0.5344%)
c = 2349.14	+/- 742.4 (31.6%)
d = -0.00967388	+/- 0.002991 (30.92%)
e = -2340.01	+/- 742.7 (31.74%)

pg_0049-B-27 :
Red-winged Blackbird (*Agelaius phoeniceus*)

11.85 2.83 0.33 8.20 1.64 7.87



Obrázek 27: Audiogram Red-winged Blackbird (*Agelaius phoeniceus*)

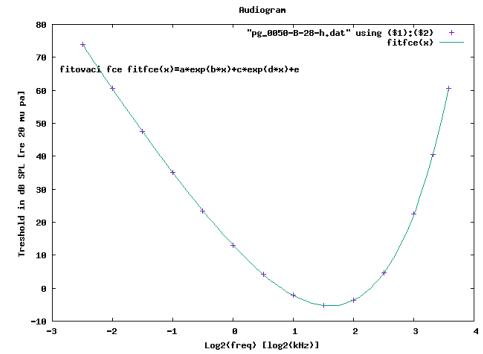
After 3974 iterations the fit converged.
final sum of squares of residuals : 0.0105685
rel. change during last iteration : -7.70427e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0342677
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00117427

Final set of parameters	Asymptotic Standard Error
a = 10.2342	+/- 0.2441 (2.385%)
b = 0.705722	+/- 0.004231 (0.5995%)
c = 2016.04	+/- 771.7 (38.28%)
d = -0.00902441	+/- 0.003384 (37.5%)
e = -2006.53	+/- 772 (38.47%)

pg_0050-B-28 :
Slate-colored Junco (*Junco hyemalis*)

-5.29 2.83 0.68 8.25 2.36 7.57



Obrázek 28: Audiogram Slate-colored Junco (*Junco hyemalis*)

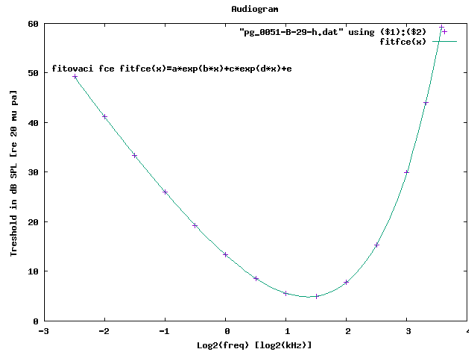
After 3903 iterations the fit converged.
final sum of squares of residuals : 0.0202193
rel. change during last iteration : -4.1674e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0473982
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00224659

Final set of parameters	Asymptotic Standard Error
a = 12.5593	+/- 0.3295 (2.623%)
b = 0.709278	+/- 0.004671 (0.6586%)
c = 2905.14	+/- 911.1 (31.36%)
d = -0.0097247	+/- 0.002984 (30.68%)
e = -2904.7	+/- 911.5 (31.38%)

pg_0051-B-29 :
Song Sparrow (*Melospiza melodia*)

4.98 2.83 0.33 8.76 1.69 8.43



Obrázek 29: Audiogram Song Sparrow (*Melospiza melodia*)

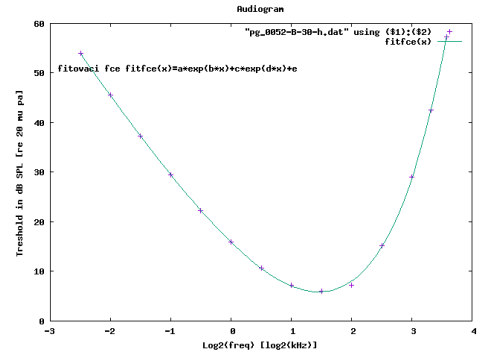
After 5484 iterations the fit converged.
final sum of squares of residuals : 0.00189632
rel. change during last iteration : -9.99183e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0145156
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000210702

Final set of parameters	Asymptotic Standard Error
a = 9.52358	+/- 0.1082 (1.136%)
b = 0.700698	+/- 0.002001 (0.2856%)
c = 3561.02	+/- 1125 (31.61%)
d = -0.00489455	+/- 0.00153 (31.25%)
e = -3557.21	+/- 1126 (31.64%)

pg_0052-B-30 :
Swamp Sparrow (*Melospiza georgiana*)

6.05 2.83 0.37 9.00 1.82 8.63



Obrázek 30: Audiogram Swamp Sparrow (*Melospiza georgiana*)

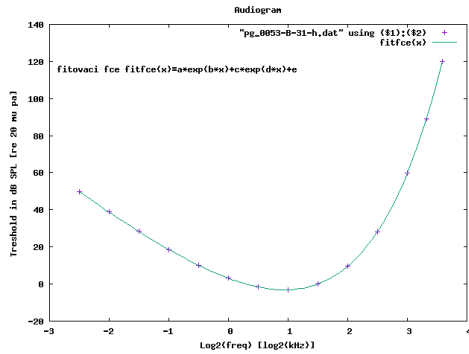
After 2578 iterations the fit converged.
final sum of squares of residuals : 0.772093
rel. change during last iteration : -9.99356e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.292896
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0857882

Final set of parameters	Asymptotic Standard Error
a = 8.6637	+/- 1.936 (22.35%)
b = 0.717871	+/- 0.04007 (5.582%)
c = 1978.67	+/- 6333 (320.1%)
d = -0.0090733	+/- 0.02846 (313.7%)
e = -1971.45	+/- 6335 (321.3%)

pg_0053-B-31 :
Western Meadowlark (*Sturnella neglecta*)

-3.06 2.00 0.37 5.55 1.42 5.18



Obrázek 31: Audiogram Western Meadowlark (*Sturnella neglecta*)

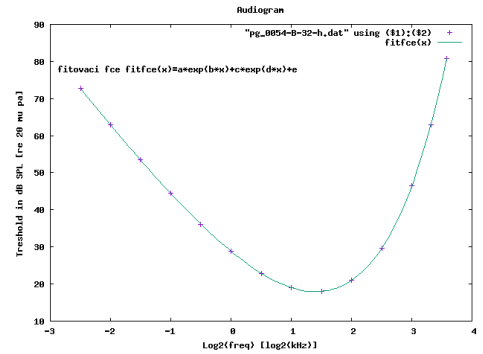
After 4670 iterations the fit converged.
final sum of squares of residuals : 0.030098
rel. change during last iteration : -9.0441e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0578293
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00334422

Final set of parameters	Asymptotic Standard Error
a = 19.555	+/- 0.4958 (2.535%)
b = 0.68609	+/- 0.004365 (0.6363%)
c = -3034.19	+/- 1597 (52.62%)
d = 0.00835449	+/- 0.004487 (53.71%)
e = 3017.73	+/- 1597 (52.92%)

pg_0054-B-32 :
Zebra Finch (*Taeniopygia guttata*)

17.98 2.83 0.44 8.24 1.89 7.81



Obrázek 32: Audiogram Zebra Finch (*Taeniopygia guttata*)

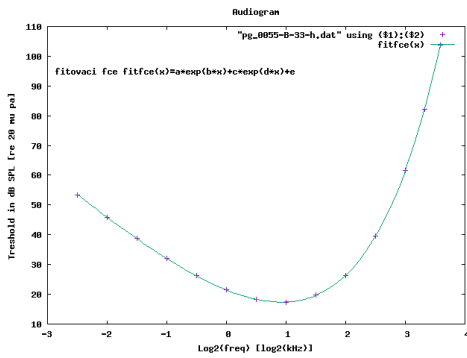
After 5639 iterations the fit converged.
final sum of squares of residuals : 0.00417722
rel. change during last iteration : -9.988e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0215438
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000464136

Final set of parameters	Asymptotic Standard Error
a = 11.268	+/- 0.1611 (1.429%)
b = 0.700425	+/- 0.002516 (0.3593%)
c = 4666.73	+/- 1937 (41.5%)
d = -0.00454736	+/- 0.001867 (41.06%)
e = -4649.21	+/- 1937 (41.66%)

pg_0055-B-33 :
Bourke's Parrot (*Neophema bourkii*)

17.36 2.00 0.23 6.50 1.22 6.27



Obrázek 33: Audiogram Bourke's Parrot (*Neophema bourkii*)

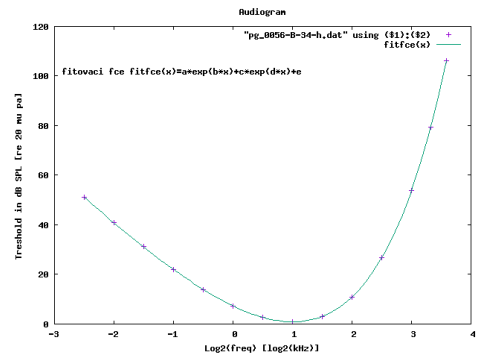
After 4185 iterations the fit converged.
final sum of squares of residuals : 0.0103867
rel. change during last iteration : -6.81314e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0339717
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00115408

Final set of parameters	Asymptotic Standard Error
a = 12.3652	+/- 0.2455 (1.985%)
b = 0.703078	+/- 0.003516 (0.5%)
c = 1755.74	+/- 691.5 (39.39%)
d = -0.00952326	+/- 0.00367 (38.54%)
e = -1746.66	+/- 691.8 (39.61%)

pg_0056-B-34 :
Budgerigar (*Melopsittacus undulatus*)

0.80 2.00 0.36 5.97 1.45 5.62



Obrázek 34: Audiogram Budgerigar (*Melopsittacus undulatus*)

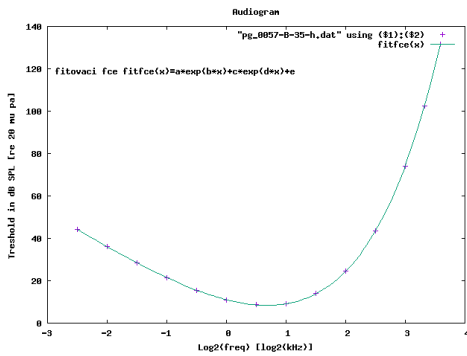
After 4111 iterations the fit converged.
final sum of squares of residuals : 0.0184701
rel. change during last iteration : -2.73268e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0453016
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00205224

Final set of parameters	Asymptotic Standard Error
a = 15.5471	+/- 0.3269 (2.102%)
b = 0.703388	+/- 0.003724 (0.5294%)
c = 2382.28	+/- 945.2 (39.68%)
d = -0.00940287	+/- 0.003652 (38.84%)
e = -2390.49	+/- 945.5 (39.55%)

pg_0057-B-35 :
Cockatiel (*Nymphicus hollandicus*)

8.60 1.41 0.22 5.22 1.08 5.00



Obrázek 35: Audiogram Cockatiel (*Nymphicus hollandicus*)

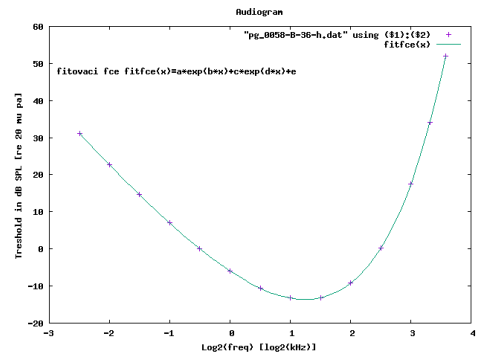
After 1581 iterations the fit converged.
final sum of squares of residuals : 0.16501
rel. change during last iteration : -5.96023e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.135405
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0183344

Final set of parameters	Asymptotic Standard Error
a = 20.0919	+/- 1.456 (7.246%)
b = 0.66705	+/- 0.012 (1.798%)
c = -593.939	+/- 226 (38.05%)
d = 0.0347311	+/- 0.01452 (41.82%)
e = 584.858	+/- 227.4 (38.89%)

pg_0058-B-36 :
African Wood Owl (*Strix woodfordii*)

-13.40 2.40 0.33 7.90 1.60 7.58



Obrázek 36: Audiogram African Wood Owl (*Strix woodfordii*)

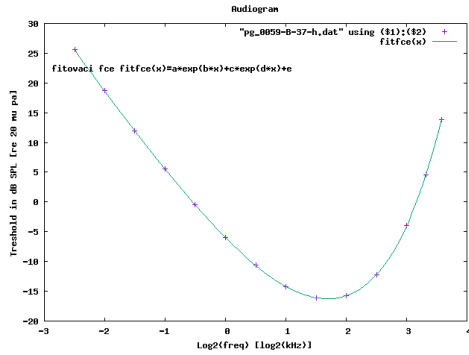
After 3945 iterations the fit converged.
final sum of squares of residuals : 0.00944017
rel. change during last iteration : -1.31468e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0323868
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00104891

Final set of parameters	Asymptotic Standard Error
a = 10.5762	+/- 0.2307 (2.182%)
b = 0.705346	+/- 0.003871 (0.5488%)
c = 1897.42	+/- 648.1 (34.16%)
d = -0.00957758	+/- 0.003201 (33.42%)
e = -1913.98	+/- 648.4 (33.88%)

pg_0059-B-37 :
Barn Owl (Tyto alba)

-16.20 2.83 0.32 12.00 1.95 11.68



Obrázek 37: Audiogram Barn Owl (Tyto alba)

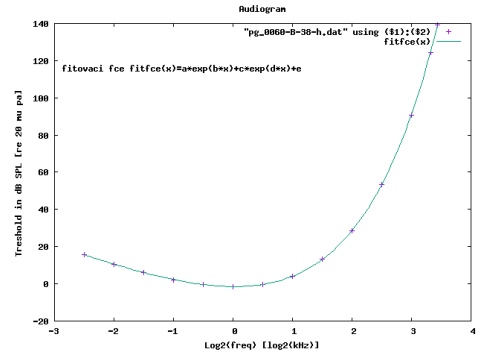
After 3959 iterations the fit converged.
final sum of squares of residuals : 0.00740078
rel. change during last iteration : -6.97308e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0286759
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000822309

Final set of parameters	Asymptotic	Standard Error	
a	= 6.09353	+/- 0.2011	(3.3%)
b	= 0.708372	+/- 0.005868	(0.8284%)
c	= 1654.48	+/- 679.1	(41.04%)
d	= -0.00877175	+/- 0.00353	(40.24%)
e	= -1666.57	+/- 679.3	(40.76%)

pg_0060-B-38 :
Brown Fish Owl (Ketupa zeylonensis)

-1.60 1.00 0.08 4.00 0.57 3.92



Obrázek 38: Audiogram Brown Fish Owl (Ketupa zeylonensis)

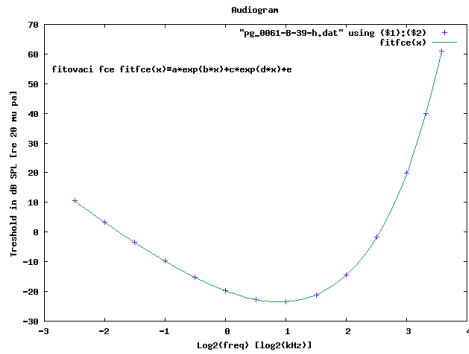
After 1447 iterations the fit converged.
final sum of squares of residuals : 0.555113
rel. change during last iteration : -5.48941e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.248353
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0616793

Final set of parameters	Asymptotic	Standard Error	
a	= 20.0632	+/- 2.752	(13.71%)
b	= 0.684674	+/- 0.0236	(3.447%)
c	= -451.523	+/- 562.1	(124.5%)
d	= 0.0308808	+/- 0.04177	(135.3%)
e	= 429.972	+/- 564.8	(131.3%)

pg_0061-B-39 :
Eagle Owl (Bubo bubo)

-23.48 2.00 0.21 6.52 1.18 6.31



Obrázek 39: Audiogram Eagle Owl (Bubo bubo)

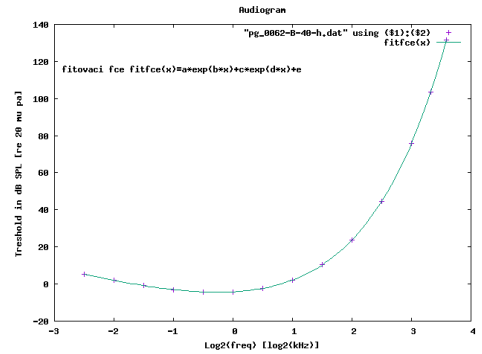
After 4224 iterations the fit converged.
final sum of squares of residuals : 0.00829497
rel. change during last iteration : -1.29973e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0303589
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000921664

Final set of parameters	Asymptotic	Standard Error	
a	= 11.9798	+/- 0.2195	(1.832%)
b	= 0.703116	+/- 0.003244	(0.4613%)
c	= 1701.7	+/- 638.7	(37.53%)
d	= -0.00936734	+/- 0.003442	(36.74%)
e	= -1733.48	+/- 638.9	(36.86%)

pg_0062-B-40 :
Great Horned Owl (Bubo virginianus)

4.31 0.71 0.03 4.15 0.35 4.12



Obrázek 40: Audiogram Great Horned Owl (Bubo virginianus)

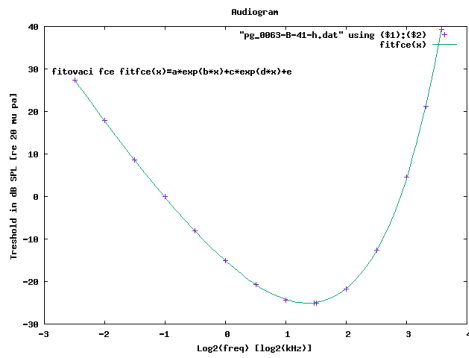
After 4038 iterations the fit converged.
final sum of squares of residuals : 0.0138939
rel. change during last iteration : -5.60679e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0392908
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00154377

Final set of parameters	Asymptotic	Standard Error	
a	= 15.4178	+/- 0.3249	(2.107%)
b	= 0.691312	+/- 0.003648	(0.5276%)
c	= -1183.32	+/- 1309	(110.6%)
d	= 0.00755437	+/- 0.008512	(112.7%)
e	= 1163.68	+/- 1309	(112.5%)

pg_0063-B-41 :
Long Eared Owl (*Asio otus*)

-25.05 2.83 0.41 8.06 1.81 7.65



Obrázek 41: Audiogram Long Eared Owl (*Asio otus*)

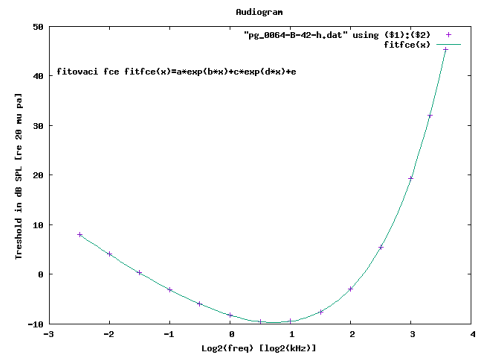
After 1308 iterations the fit converged.
final sum of squares of residuals : 0.160174
rel. change during last iteration : -8.47552e-06

degrees of freedom (FIT_NDF) : 10
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.12656
variance of residuals (reduced chisquare) = WSSR/ndf : 0.0160174

Final set of parameters	Asymptotic	Standard Error	
a	= 9.17166	+/- 0.6615	(7.212%)
b	= 0.737917	+/- 0.0134	(1.816%)
c	= 566.597	+/- 176.9	(31.22%)
d	= -0.0341592	+/- 0.009947	(29.12%)
e	= -590.947	+/- 177.5	(30.04%)

pg_0064-B-42 :
Mottled Owl (*Strix virgata*)

-9.54 1.41 0.06 8.20 0.72 8.14



Obrázek 42: Audiogram Mottled Owl (*Strix virgata*)

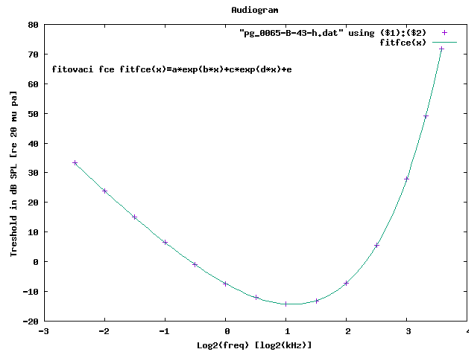
After 3419 iterations the fit converged.
final sum of squares of residuals : 0.00355814
rel. change during last iteration : -2.78023e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0198834
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000395349

Final set of parameters	Asymptotic	Standard Error	
a	= 7.32194	+/- 0.1402	(1.914%)
b	= 0.705148	+/- 0.003404	(0.4827%)
c	= 676.01	+/- 215.8	(31.93%)
d	= -0.0130068	+/- 0.004033	(31%)
e	= -691.558	+/- 216	(31.23%)

pg_0065-B-43 :
Scops Owl (*Otus scops*)

-14.29 2.00 0.34 6.65 1.50 6.31



Obrázek 43: Audiogram Scops Owl (*Otus scops*)

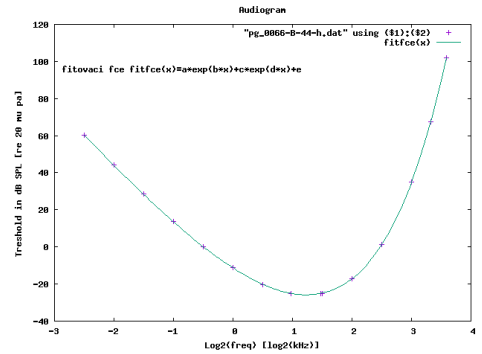
After 4042 iterations the fit converged.
final sum of squares of residuals : 0.0122322
rel. change during last iteration : -4.37224e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0368664
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00135913

Final set of parameters	Asymptotic	Standard Error	
a	= 13.1191	+/- 0.2635	(2.009%)
b	= 0.704679	+/- 0.003563	(0.5056%)
c	= 2081.64	+/- 705.4	(33.89%)
d	= -0.00980306	+/- 0.003249	(33.14%)
e	= -2102.12	+/- 705.7	(33.57%)

pg_0066-B-44 :
Snowy Owl (*Nyctea scandiaca*)

-25.25 2.00 0.63 5.88 1.91 5.26



Obrázek 44: Audiogram Snowy Owl (*Nyctea scandiaca*)

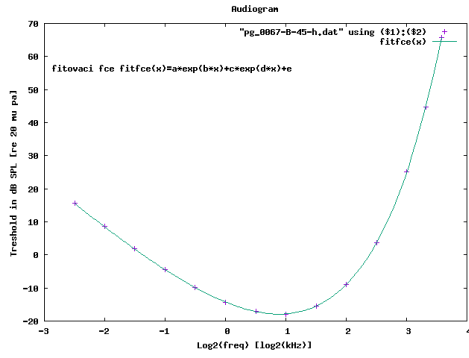
After 3968 iterations the fit converged.
final sum of squares of residuals : 0.0333056
rel. change during last iteration : -7.61092e-06

degrees of freedom (FIT_NDF) : 10
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.057711
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00333056

Final set of parameters	Asymptotic	Standard Error	
a	= 20.4316	+/- 0.395	(1.933%)
b	= 0.706276	+/- 0.003445	(0.4878%)
c	= 3442.97	+/- 987	(28.67%)
d	= -0.0101675	+/- 0.002849	(28.02%)
e	= -3474.7	+/- 987.4	(28.42%)

pg_0067-B-45 :
Spotted Wood Owl (Strix seloputo)

-17.89 2.00 0.21 6.55 1.17 6.34



Obrázek 45: Audiogram Spotted Wood Owl (Strix seloputo)

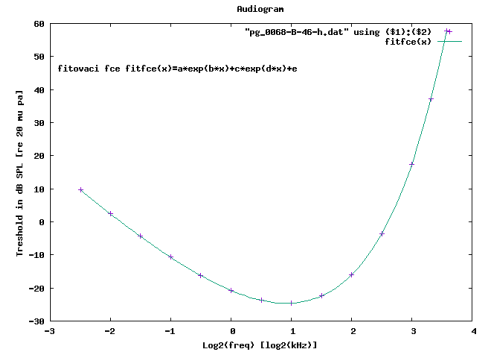
After 4090 iterations the fit converged.
final sum of squares of residuals : 0.00810356
rel. change during last iteration : -2.33467e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0300066
variance of residuals (reduced chisquare) = WSSR/ndf : 0.000900395

Final set of parameters	Asymptotic	Standard Error
a	= 11.7761	+/- 0.2148 (1.824%)
b	= 0.70404	+/- 0.003235 (0.4594%)
c	= 1484.15	+/- 495.1 (33.36%)
d	= -0.0105653	+/- 0.003441 (32.57%)
e	= -1510.22	+/- 495.3 (32.79%)

pg_0068-B-46 :
Tawny Owl (Strix aluco)

-24.62 2.00 0.22 6.62 1.19 6.41



Obrázek 46: Audiogram Tawny Owl (Strix aluco)

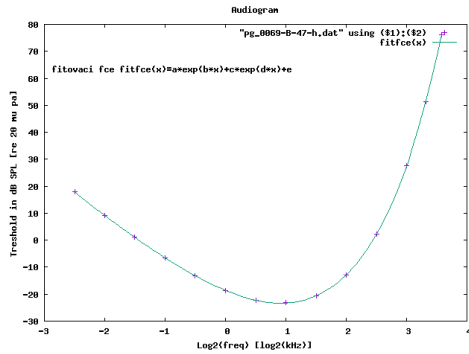
After 4212 iterations the fit converged.
final sum of squares of residuals : 0.0110325
rel. change during last iteration : -1.26409e-07

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0350118
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00122583

Final set of parameters	Asymptotic	Standard Error
a	= 11.7794	+/- 0.2536 (2.153%)
b	= 0.703057	+/- 0.003811 (0.5421%)
c	= 1797.89	+/- 826.2 (45.95%)
d	= -0.00884564	+/- 0.003983 (45.03%)
e	= -1830.43	+/- 826.4 (45.15%)

pg_0069-B-47 :
White-faced Scops Owl (Otus leucotis)

-23.26 2.00 0.28 6.04 1.29 5.76



Obrázek 47: Audiogram White-faced Scops Owl (Otus leucotis)

After 4527 iterations the fit converged.
final sum of squares of residuals : 0.0244658
rel. change during last iteration : -5.13637e-06

degrees of freedom (FIT_NDF) : 9
rms of residuals (FIT_STDFIT) = sqrt(WSSR/ndf) : 0.0521385
variance of residuals (reduced chisquare) = WSSR/ndf : 0.00271843

Final set of parameters	Asymptotic	Standard Error
a	= 15.5641	+/- 0.4429 (2.846%)
b	= 0.687426	+/- 0.004906 (0.7137%)
c	= -2445.67	+/- 1509 (61.7%)
d	= 0.00814554	+/- 0.005127 (62.94%)
e	= 2411.45	+/- 1510 (62.6%)

Shrnutí

Předložená práce se věnuje automatickému rozpoznávání a verifikaci ptáků. Zabývá se návrhem a evaluací nových metod pro automatickou identifikaci jedinců ptáků s využitím živých nahrávek bez nutnosti jejich předzpracování. Automatizované systémy založené na těchto metodách (Automatic Recognition System of Bird Individual, ARSBI) umožní identifikaci jedinců ptáků bez nutnosti jejich kroužkování nebo kontroly DNA. Práce se dále věnuje návrhu nové banky filtrů, optimalizované pro ptačí zpěv (Bird Adapted Filter, BAF).

Současně bylo nezbytné řešit úkoly související s hlavními cíli práce. Mezi ně patří nalezení nového způsobu vyjádření audiogramů ptáků v automatizovaných systémech, návrh a vytvoření databáze ptačích zpěvů (Bird Song Corpus, BSC), ověření možností nových metod při rozpoznávání ptáků a konečně návrh a využití navrženého ARSBI při identifikaci jiných zvířecích druhů, konkrétně ryposů. Navržené metody, popsané v jednotlivých kapitolách, jsou zároveň experimentálně ověřeny.

Členění práce: První část popisuje současný stav obou hlavních oblastí výzkumu, tedy ornitologie a rozpoznávání mluvčího. Druhá část uvádí technické prostředky, které byly při výzkumu využity. Autor vytvořil programový celek v prostředí Matlab a využíval části kódu napsané v jazyce C++. Následují kapitoly, které se věnují plnění hlavních a dílčích cílů práce. Poslední část shrnuje dosažené výsledky a uvádí možnosti dalšího rozvoje. Přílohy obsahují 51 audiogramů (pro 47x druhů a 4x agregované), jejichž definice byly v rámci této práce nalezeny.

Summary

Our thesis deals with automatic recognition and identification of bird individuals. The first goal of the thesis is the design and evaluation of new methods and algorithms for automatic bird individual identification using live recordings, without their pre-processing. An automated system using the suggested methods is going to be called Automatic Recognition System of Bird Individual (ARSBI), and it enables a bird identification without the necessity of catching them for banding or DNA check. The thesis also deals with a new filter bank optimized for bird song (Bird Adapted Filter, BAF).

At the same time, it was necessary to solve the below tasks that are closely connected to the main goals of our thesis. Namely a new mathematical expression of a bird audiograms for ARSBI, bird song database design and development (Bird Song Corpus, BSC), new speaker recognition methods evaluation for bird vocalization, and design and utilization of a new ARSBI for other species not just birds, particularly mole-rats. Experiment evaluations of proposed methods are also described.

Thesis structure: The first chapter deals with State of the Art of both main research fields, ornithology and speaker recognition. The second chapter describes development framework we used for the experiments. Author developed a new framework in Matlab, some C++ code parts are also in use. Then separated chapters describe goals completion. We end with a discourse on the results and future work. The attachment contains 51 audiograms (47x species and 4x aggregate) for which we discovered the setup parameters.

References

- [ARM99] Armstrong, D. P., Castro, I., Alley, J. C., Feenstra, B. & Perrott, J. K. 1999. Mortality and behaviour of hihi, an endangered New Zealand honeyeater, in the establishment phase following translocation. *Biological Conservation*, 89, 329–339.
- [ARR15] Julio G. Arriaga a,1, Martin L. Cody b,2, Edgar E. Vallejo a,3, Charles E. Taylor: Bird-DB: A database for annotated bird song sequences. 2015. *Ecological Informatics*, Volume 27, May 2015, Pages 21–25. doi:10.1016/j.ecoinf.2015.01.007
- [BAL90] Balcombe JP (1990) Vocal recognition of pups by mother Mexican free-tailed bats, *Tadarida brasiliensis mexicana*. *Anim Behav* 39:960–966.
- [BED13] Bednářová R, Hrouzková-Knotková E, Burda H, et al. (2013) Vocalizations of the giant mole-rat (*Fukomys mechowii*), a subterranean rodent with the richest vocal repertoire. *Bioacoustics* 22:87–107.
- [BEE85] Beecher M, Stoddard P, Loesche P (1985) Recognition of parents voices by young cliff swallows. *Auk* 102:600–605.
- [BIB00] Bibby, C. J., Burgess, N. D., Hill, D. A. & Mustoe, S. 2000. *Bird census techniques*. 2nd edn. London: Academic Press.
- [BIM04] F. Bimbot, J.F. Bonastre, C. Fredouille, G. Gravier, M.I. Chagnolleau, S. Meignier, T. Merlin, O.J. Garcia, P. Delacretaz, and D.A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 4:430–451, 2004.
- [BLU11] Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M. and Kirschel, A. N. G. (2011), Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48: 758–767. doi: 10.1111/j.1365-2664.2011.01993.x
- [BOI95] Boinski S, Campbell A (1995) Use of trill vocalizations to coordinate troop movement among white-faced capuchins - a 2nd field-test. *Behaviour* 132:875–901.
- [BOU13] Bouchet H, Blois-Heulin C, Lemasson A (2013) Social complexity parallels vocal complexity: a comparison of three non-human primate species. *Front Comp Psychol* 4:390.
- [BRI12] Forrest Briggs^{1,a}, Balaji Lakshminarayanan¹, Lawrence Neal¹, Xiaoli Z. Fern¹, Raviv Raich¹, Sarah J. K. Hadley², Adam S. Hadley² and Matthew G. Betts²: Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. 2012. *J. Acoust. Soc. Am.* 131, 4640 (2012); <http://dx.doi.org/10.1121/1.4707424>
- [BUD14] Budka M, Wojas L, Osiejuk T. 2014. Is it possible to acoustically identify individuals within a population? *J Ornithol.* 156:481–488. (2015).

- [BUR90] Burda, H.: Constraints of pregnancy and evolution of sociality in mole-rats. With special reference to reproductive and social patterns in *Cryptomys hottentotus* (Bathyergidae, Rodentia). 1990, *Zool Syst Evol-Forsch* 28:26–39.
- [BYE96] Byers, B. E.: Messages encoded in the songs of chestnut-sided warblers. 1996, *Animal Behaviour*, 52, 691–705.
- [CAM06] Campbell, W., Sturim, D., Reynolds, D., and Solomonoff, A., 2006. SVM based speaker verification using a GMM supervector kernel and nap variability compensation. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*, 1:I–I.
- [CAT08] Catchpole, C., K., Slater, P., J., B.: *Bird Song. Biological Themes and Variations*, 2nd edition, Cambridge Press University, 2008
- [CHE10] Cheng, J., Yuehua, S., Liqiang, J.: A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. *Pattern Recognition*, 43, 3846–3852.
- [CLE05] Clemins, P., Johnson, T., M., Leong, K., M., Savage, A., 2005, Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations, *J. Acoust. Soc. Am.*, Vol. 117, 956-963.
- [CHU09] W. Chu and A. Alwan: A correlation-maximization denoising filter used as an enhancement frontem for noise robust bird call classification, *Proc. of Interspeech*, 2009, pp. 2831-2834
- [CHU12] Chu, W., Alwan, A.: FBEM: A filter bank EM algorithm for the joint optimalization of features acoustic model parameters in bird call classification
- [DEH09] Dehak, N.: *Discriminative and generative approaches for long- and short-term speaker characteristics modeling: Application to speaker verification*, Quebec University, Montreal, 2009
- [DEH10] Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., and Ouellet, P., 2010. Front-end factor analysis for speaker verification. *IEEE Transactions on audio, speech, and language processing*.
- [DEH11] Dehak, N., Kenny, P., Dehak, R., Dumouchel, P and Ouellet, P. Front-End Factor Analysis for Speaker Verification *IEEE Transactions on Audio, Speech and Language Processing*, 19(4), pp. 788-798, May 2011.
- [DEM77] Dempster, A., Laird, N., and Rubin, D., 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- [DEN98] Dent, M., Dooling, R., J., Leek, M., Summers, V.: Masking of tones by harmonic complexes in the budgerigar (*Melopsittacus undulatus*). *J. Acoust. Soc. Am.* Volume 104, Issue 3, pp. 1811-1811, Sept. 1998
- [DOO95] Dooling, R., J., Best, C., Brown, S.: Discrimination of synthetic full-formant and sinewave/ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *J. Acoust.Soc.Am.* 97 (3), March 1995.

- [DOO02a] Dooling, R., J., Leek, M., R., Gleich, O., Dent, M. L.: Auditory temporal resolution in birds: discrimination of harmonic complexes. *J. Acoust. Soc. Am.*, roč. 112, č. 2, s. 748–759, srp. 2002.
- [DOO02b] Dooling, R., J.: Avian Hearing and the Avoidance of Wind Turbines. National Renewable Energy Laboratory, 2002. NREL/TP-500-30844.
- [DVO13] Dvorakova, V.: Individual features in vocalization of the Mashona mole-rat (*Fukomys darlingi*). Master thesis, University of South Bohemia in České Budějovice, Faculty of Science, 2013.
- [EHN15] M. Ehnes & J.R. Foote (2015) Comparison of autonomous and manual recording methods for discrimination of individually distinctive Ovenbird songs, *Bioacoustics*, 24:2, 111-121, DOI: 10.1080/09524622.2014.994228
- [FOX08] Fox, E. J. S., Roberts, J. D. & Bennamoun, M. 2008. Call-independent individual identification in birds. *Bioacoustics-the International Journal of Animal Sound and Its Recording*, 18, 51–67.
- [FRA00] Francescoli G (2000) Sensory capabilities and communication in subterranean rodents. University of Chicago Press, Chicago, pp 111 –144
- [GAB96] Gabathuler U, Bennett NC, Jarvis JUM (1996) The social structure and dominance hierarchy of the Mashona mole-rat, *Cryptomys darlingi* (Rodentia: Bathyergidae) from Zimbabwe. *JZool Lond* 240:224–231.
- [GLE11] Ondrej Glembek, Lukas Burget, Pavel Matejka, Martin Karafiat, Patrick Kenny: Simplification and optimization of i-Vector extraction, ICASSP, Prague, Brno University of Technology, 2011
- [GRA10] M. Graciarena, M. Delplanche, E. Shriberg, A. Stolcke, and L. Ferrer: Acoustic front-end optimization for bird species recognition, ICASSP, 2010, pp. 293-296.
- [GRA11] M. Graciarena, M. Delplance, E. Shriberg and A. Stolcke, "Bird species recognition combining acoustic and sequence modeling," in Proc. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 341--344.
- [GRE98] G. Greenberg a M. M. Haraway, *Comparative Psychology: A Handbook*. Routledge, 1998.
- [HEF98] Heffner, H. E. & Heffner, R. S. (1998). Hearing. In G. Greenberg and M. M. Haraway (Eds.), *Comparative Psychology, A Handbook*. (pp. 290-303). Garland: New York.
- [HER90] Hermansky, H.: Perceptual linear predictive analysis of speech, *J. Acoust. Soc. Am.*, 1990
- [HET86] Heth G, Frankenberg E, Nevo E (1986) Adaptive optimal sound for vocal communication in tunnels of a subterranean mammal (*Spalax ehrenbergi*). *Experientia* 42:1287–1289.
- [HOE10] Hoeschele M, Moscicki MK, Otter KA, van Oort H, Fort KT, Farrell TM, Lee H, Robson SWJ, Sturdy CB (2010) Dominance signalled in an acoustic ornament. *Anim Behav* 79:657–664.

[JAN11] Peter Jančovič, Münevver Köküer: Automatic Detection and Recognition of Tonal Bird Sounds in Noisy Environments. 2011. EURASIP Journal on Advances in Signal Processing 982936. DOI: 10.1155/2011/982936

[KEN07] Kenny, P., Boulianne, G., Ouellet, P, Dumouchel, P.: Speaker and Session Variability in GMM-Based Speaker Verification, IEEE Transaction on audio, speech, and language processing, Vol. 15, no. 4, may 2007, 15:1448–1460.

[KOG98] J.A. Kogan and D.Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," JASA, vol. 103, no. 4, pp. 2185-2196, 1998.

[KON70] M. Konishi, „Comparative neurophysiological studies of hearing and vocalizations in songbirds", Z. Vergl. Physiol., roč. 66, č. 3, s. 257–272, zář. 1970.

[KRA09] Alan H. Krakauer, Maura Tyrrell, Kenna Lehmann, Neil Losin, Franz Goller, Gail L. Patricelli: Vocal and anatomical evidence for two-voiced sound production in the greater sage-grouse *Centrocercus urophasianus*. 2009. Journal of Experimental Biology 209 212: 3719-3727; doi: 10.1242/jeb.033076

[KRO04] Kroodsma, D. E. 2004. The diversity and plasticity of bird song. In: Nature's music: the science of birdsong, (Ed. by P. R. Marler & H. Slabbekoorn), pp. 108–131. San Diego, CA: Elsevier Academic Press.

[KUN10] Kuntoro, A., Johnson, M., Osiejuk, T.,: Acoustic censusing using automatic vocalization classification and identity recognition, J.Acoustic. Soc. Am, 2010

[KWA06] C. Kwan, K.C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, C. Rochet: An Automated Acoustic System to Monitor and Classify Birds. 2006. EURASIP Journal on Advances in Signal Processing. December 2006, 2006:096706. DOI: 10.1155/ASP/2006/96706

[LAI07] Laiolo, P., Vogeli, M., Serrano, D. & Tella, J. 2007. Testing acoustic versus physical marking: two complementary methods for individual-based monitoring of elusive species. Journal of Avian Biology, 38, 672–681.

[LAN07] Lange S, Burda H, Wegner RE, Dammann P, Begall S, Kawalika M (2007). Living in a “stethoscope”: burrow-acoustics promote auditory specializations in subterranean rodents. Naturwissenschaften 94:134–138. 23

[LAU07] A. M. Lauer, R. J. Dooling, M. R. Leek, a K. Poling, „Detection and discrimination of simple and complex sounds by hearing-impaired Belgian Waterslager canaries", J. Acoust. Soc. Am., roč. 122, č. 6, s. 3615–3627, pro. 2007.

[LIN12a] Linhart, P., Fuchs, R., Poláková, S. & Slabbekoorn, H. 2012a. Once bitten twice shy: long-term behavioural changes caused by trapping experience in willow warblers *Phylloscopus trochilus*. Journal of Avian Biology, 43, 186–192.

[LIN12b] Linhart, P., Slabbekoorn, H., Fuchs, R. 2012b. The communicative significance of song frequency and song length in territorial chiffchaffs. Behavioral Ecology, 23, 1338–1347.

[LIN13] Linhart, P., Jaška, P., Petrusková, T., Petrušek, A. & Fuchs, R. 2013. Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. *Behavioural Processes*, 100, 139-145.

[MAC74] MacArthur, R. H. & MacArthur, A. T. 1974. On the use of mist nets for population studies of birds. *Proceedings of the National Academy of Sciences*, 71, 3230–3233.

[MAN02] Manser MB, Seyfarth RM, Cheney DL (2002) Suricate alarm calls signal predator class and urgency. *Trends Cogn Sci* 6:55–57.

[MAR63] Marquardt, D. 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics* 11 (2): 431–441. doi:10.1137/0111030.

[MAR04] Marler, P., Slabbekoorn, H.: *Nature's music. The Science of Birdsong*. Hardcover, 504 pages, including 2 CD's. Academic Press/Elsevier, San Diego 2004

[MOL08] Molnár, c., Kaplan, F., et al.: *Classification of dog barks: a machine learning approach*, Springer, 2008

[MUL99] Müller, L.: *Rozpoznávání plynulé mluvené češtiny v úloze se středním slovníkem a redukovanou složitostí jazykového korpusu*. Disertační práce ZČU FAV, Plzeň, 1999

[NOT86] Nottebohm, F., Nottebohm, M. E. & Crane, L. 1986. Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and Neural Biology*, 46, 445–471.

[OKA85] K. Okanoya a R. J. Dooling, „Colony differences in auditory thresholds in the canary (*Serinus canarius*)", *J. Acoust. Soc. Am.*, roč. 78, č. 4, s. 1170–1176, říj. 1985.

[PAY96] Payne, R. B. 1996. Song traditions in indigo buntings: origin, improvisation, dispersal, and extinction in cultural evolution. In: *Ecology and evolution of acoustic communication in birds*, (Ed. by D. E. Kroodsma & E. H. Miller), pp. 198–220. Ithaca, New York: Cornell University Press.

[PET16] Tereza Petrusková*, Iveta Pišvejcová, Anna Kinštová, Tomáš Brinke, Adam Petrušek: 2015. *Methods in Ecology and Evolution*, Volume 7, Issue 3, pages 274–284, March 2016. British Ecological Society. DOI: 10.1111/2041-210X.12496

[POT14] Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K.: Automatic bird sound detection in long real-field recordings: Applications and tools. 2014. *Applied Acoustics* 80:1–9, June 2014. DOI: 10.1016/j.apacoust.2014.01.001

[PSU06] Psutka, J., Müller, L., Matoušek, J., Radová, V.: *Mluvíme s počítačem česky*, Academia, 2006

[PTA15a] Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Müller, L.: Automatic recognition of bird individuals on an open set using as-is recordings, *Bioacoustics*, 2015, DOI: 10.1080/09524622.2015.1089524.

[PTA15b] Ptacek, L., Zajic, Z., Vanek, J., Linhart, P., Müller, L.: Using Identity Vectors for the Bird Individual Identification on the Closed Set. *Conference Listening in the wild*. 2015, London, United Kingdom.

- [PTA15c] Ptacek, L., Vanek, J., Linhart, P., Müller, L.: Improving the Automatic Bird Individual Identification Method by Parametrization Data Merging. International Bioacoustics Council (IBAC). 2015, Murnau, Germany.
- [PTA15d] Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Müller, L.: Application of Speaker Recognition Methods for Chiffchaff Individual Identifications. International Bioacoustics Council (IBAC). 2015, Murnau, Germany.
- [PTA16a] –Ptacek, L., Eisner, J., Vanek, J., Pruchova, A., Muller, L.: Bird Audiogram Unified Equation. 2016, in prep.
- [PTA16b] –Ptacek, J., Vanek, L., Eisner, J., Pruchova, A., Muller, L.: Improving Automatic Bird Species and Individual Identification by Dedicated Bird Adapted Filter. Bioacoustics. 2016, in prep.
- [REN04] Rendall D, Owren MJ, Weerts E, Hienz RD (2004) Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *J Acoust Soc Am* 115:411 –421.
- [REY00] Reynolds, D., A., Quatieri, T., F., Dunn, R., B.: Speaker verification using adapted Gaussian mixture models. 2000, *Digital Signal Processing*, 10(1-3):19 – 41.
- [REY95] Reynolds, D. A. and Rose, R. C.: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models.. *IEEE Transaction of Speech and Audio Processing*, Vol. 3, No. 1, January 1995, pages 72-83.
- [SED15] Ondřej Sedláček, Jana Vokurková, Michal Ferenc, Eric Nana Djomo, Tomáš Albrecht, David Hořák: A comparison of point counts with a new acoustic sampling method: a case study of a bird community from the montane forests of Mount Cameroon. 2015. *Ostrich: Journal of African Ornithology*, Volume 86, Issue 3, 2015. DOI: 10.2989/00306525.2015.1049669
- [SEL05] Selin, A.: Bird sound classification using wavelets, Tampere University of Technology, 2005.
- [SEN10] Senoussaoui, M., Kenny, P., Dehak, N., Dummouchel, P.: An i-vector Extractor Suitable for Speaker Recognition with both Microphone and Telephone Speech, 2010
- [SIL94] Silva KB da, Kramer DL, Weary DM (1994) Context-specific alarm calls of the eastern chipmunk, *Tamias striatus*. *Can J Zool* 72:1087–1092.
- [STU05] Sturim, D. E. and Reynolds, D. A., 2005. Speaker Adaptive Cohort Selection for T-norm in Text-Independent Speaker Verification. *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'05*, 1:741–744
- [SWI09] Swiston, K. A. and Mennill, D. J. (2009), Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology*, 80: 42–50. doi: 10.1111/j.1557-9263.2009.00204.x
- [TAY10] Taylor AM, Reby D (2010) The contribution of source–filter theory to mammal vocal communication research. *J Zool* 280:221 –236.

- [TRA05] Trawicki, M.B., Johnson, M.T., Osiejuk, T.S. (2005) Automatic song-type classification and speaker identification of norwegian ortolan bunting (*Emberiza hortulana*) vocalizations. 2005 IEEE Workshop on Machine Learning for Signal Processing, art. no. 1532913, pp. 277-282.
- [TRI08] Trifa, V. M., Kirschel, N. G., Taylor, Ch. E. 2008. Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. Acoustical Society of America, 2008, pages 2424-2431.
- [ULL16] Ulloa, J., S., Gasc,, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J.: Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest, 2016. Ecological Informatics, Volume 31, January 2016, Pages 91–99. doi:10.1016/j.ecoinf.2015.11.012
- [VEN15] Ventura TM, Oliveira AG, Ganchev TG, Figueiredo JM, Jahn O, Marques MI, Schuchmann KL. 2015. Audio parameterization with robust frame selection for improved bird identification. *Expert Syst. Appl.* 42:8463–8471. ISSN: 0957-4174
- [WEI12] Wei Chu, A. A., 2012. Fbem: A filter bank EM algorithm for the joint optimization of features acoustic model parameters in bird call classification. *IEEE*, pages 1993–1996.
- [WIN82] Wingfield, J. C., Smith, J. P. & Farner, D. S. 1982. Endocrine responses of White-crowned sparrows to environmental stress. *Condor*, 84, 399–409.
- [XIA11] Xia, C., Huang, R., Wei, Ch., Nie, P., Zhang, Y.: Individual identification on the basis of the songs of the Asian Stubtail. 2011. *Chinese Birds* 2011, 2(3) 132-139 DOI: 10.5122/cbirds.2011.0024
- [YOS09] Yosida S, Okanoya K (2009) Naked mole-rat is sensitive to social hierarchy encoded in antiphonal vocalization. *Ethology* 115:823–831
- [YIN08] Yin, S.-C., Rose, R., and Kenny, P. 2008. Adaptive score normalization for progressive model adaptation text independent speaker verification. *ICASSP*, 1:857–860.
- [ZSE15] Zsebök, S., Czabán, D., Farkas, J., Siemers, B., M., von Mertenc, S: Acoustic species identification of shrews: Twittering calls for monitoring. 2015. *Ecological Informatics*. Elsevier. doi:10.1016/j.ecoinf.2015.02.002

List of author's publications

Ptacek, L., Eisner, J., Vanek, J., Pruchova, A., Muller, L.: Bird Audiogram Unified Equation. 2016, in prep.

Ptacek, L., Vanek, J., Eisner, J., Pruchova, A., Muller, L.: Improving Automatic Bird Species and Individual Identification by Dedicated Bird Adapted Filter. 2016, in prep.

Ptacek, L., Jelinek, P.: A comparative study of analysis of coronal oscillations based upon Fourier transformation methods. RadioSun Workshop and Summer School, 2016. České Budějovice.

Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Müller, L.: Automatic recognition of bird individuals on an open set using as-is recordings, Bioacoustics, 2015, DOI: 10.1080/09524622.2015.1089524.

Ptacek, L., Zajic, Z., Vanek, J., Linhart, P., Müller, L.: Using Identity Vectors for the Bird Individual Identification on the Closed Set. Listening in the wild conference, 2015, London, United Kingdom.

Ptacek, L., Vanek, J., Linhart, P., Müller, L.: Improving the Automatic Bird Individual Identification Method by Parametrization Data Merging. International Bioacoustics Council (IBAC), 2015, Murnau, Germany.

Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Müller, L.: Application of Speaker Recognition Methods for Chiffchaff Individual Identifications. International Bioacoustics Council (IBAC), 2015, Murnau, Germany.

Other activities

Member of Local Organization Committee, RadioSun Workshop and Summer School, České Budějovice, 2016.

Müller, L., Ptáček, L., Sukdol, L.: Bird Song Corpus, software, Západočeská univerzita v Plzni, 2014.

Ptacek, L.: Automatic Bird Identification and Verification, Phd thesis report, Západočeská univerzita v Plzni, 2012.

Ptacek, L.: Automatické rozpoznávání zvuků ptáků, Status report, Západočeská univerzita v Plzni, 2011.

Ptacek, L.: Návrh rozhlasové hlasatelny a režie, diplomová práce, České vysoké učení technické v Praze, 1998.

Ptacek, L.: Active noise control in ducts, bakalářská práce, České vysoké učení technické v Praze, 1996.