Analysis of Acoustic Communication in Parrots
DIPLOMA THESIS

Petr Skřípal *
petr.skripal@seznam.cz
Lodžská 463, Praha 8 - Bohnice, 181 00
Supervisor: Ing. Jan Koutník

Abstract: The topic of the thesis is the analysis of the acoustic communication and vocalization in parrots. In the experimental part of this work, we are dealing with cepstral and statistical analyses of real phonic samples and we will introduce a method for the extraction of appropriate acoustic signal features. The document describes a proposal for implementation of a system that would use the analytical outcomes of our work for the need of phonetics expression classification. We focus on detailed analyses of the self-organizing neural network and its application during the process of the acquired features classification. An overview of implementation of the above-mentioned classification system and of its use with real data constitutes an integral part of this study.
Finally in this paper, we will discuss the possibilities of application of the Parrot Speech Toolbox and briefly present some crucial results as the outcomes of the whole work.

Keywords: Cepstral analysis, statistical analysis, cluster analysis, neural networks, self-organizing map, acoustic signal processing, clustering, parrot vocalization

1 Introduction

Human speech and bird vocalization are complex communicative behaviors with notable similarities in development and underlying mechanisms [3]. There is an important difference between humans and birds in the way vocal complexity is generally produced [4]. However, it has been proposed that, analogous to human speech production, tongue movements observed at parrot vocalizations modulate formant [18] characteristics independently from the vocal source [9], [2].

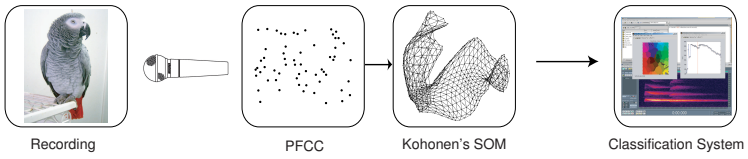


Fig. 1: *Processing chain of parrot vocalization recognition. The acoustic data are preprocessed using a new PFCC filter and Kohonen's self-organizing neural network before it is processed by our classification system.*

* Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University, Technická 2, 166 27, Prague

Irene Pepperberg's research [10] with captive African grey parrots (Psittacus erithacus), including parrot Alex, has shown that these parrots are capable of associating human words with their meanings, at least to some extent. While comparative judgements of animal intelligence are always very difficult to make objectively, greys are generally regarded as being amongst the most intelligent of birds. Figure 1 shows a processing chain of parrot vocalization recognition.

The analysis of human speech is usually based on a spectral analysis of the signal's formants F_x and its fundamental frequency – pitch F_0. Together with an inherent knowledge of the language, its rules and phonetic consequences the analysis serves as an initial base upon which the speech recognition (SR) system can be built.

This is of course valid for human speech, but considering a non-human acoustic signal that contains a given phonetic context, preliminary steps have to be made before any SR is developed. An example of such an acoustic signal is the vocal communication of grey parrots. The phenomenom of avian "speech" has been studied e.g. in [13] and recent experiments of Pepperberg's research support such context, though we are not aware of any recognition system that would model parrot speech or analytically classify sounds produced by parrots.

In our work, we focus on the analysis of parrot vocalization and features of transformed acoustic signal that can be utilized in a further patterning and in a subsequent classification procedure. There are of course many inter-related areas concerning the subject of parrot speech, e.g. linguistical, and more generally, semiological studies of the phenomena; nevertheless, a tool for primary clustering and classification is necessary. Evidence of the existence of *phonemes* or further linguistical derivatives in parrot speech could be built on such a classifier which would facilitate the survey and allow for the required resolution for a distinction among sounds. A low-level approach of this work expect an occurrence of *phones* as described e.g in [20].

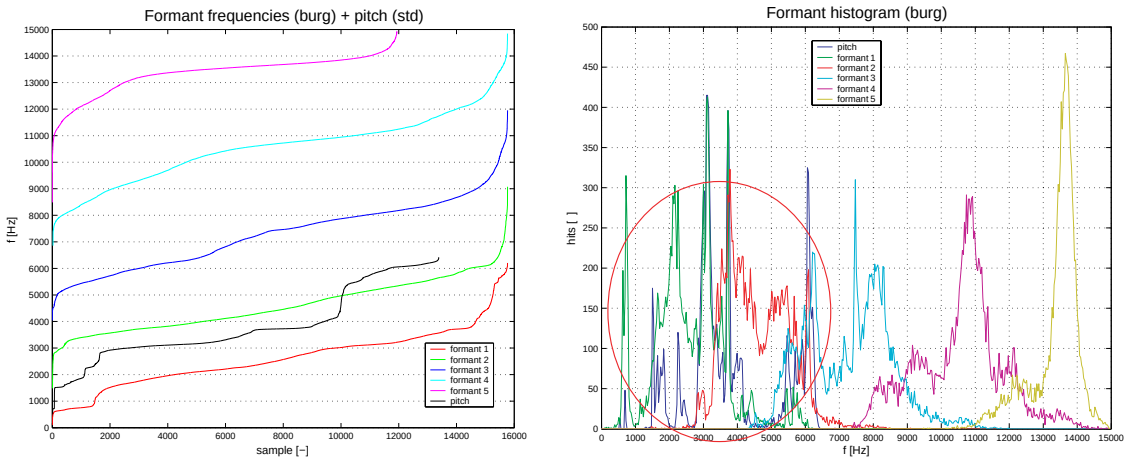
2 Experimental Results

2.1 Cepstral Analysis

This part of our experiments is dedicated to the use of cepstral and filter-bank analysis [11] in the feature extraction from the acoustic data of the recordings. As the outcome, we propose a new scale adapted to parrot vocalization: *Par scale*. The scale is integrated to the frequency warping function f_{par} of the *par scaled filter-bank* (see Figure 3) implementation (two models of the function are proposed – (*smooth* and *strict* version).

A derivation of the function is based on the course of *mel-scaled* function f_{mel} [19]. To find a correct frequency warping, we performed a profound analysis of the formant frequency histogram of vocalization data, see Figure 2. The region that we call the critical bandwidth interval – Γ reflects the highest density of significant frequencies and the filter should provide high resolution within this area. To find such minimal formant resolution, we based its estimation Θ on the deviation of formants, which were computed for several selected recordings. The implementation of the filter consists of $\mathcal{K}$ filter channels which satisfy the formant resolution and the critical bandwidth. The exact values are defined as:

$$\boxed{\Theta = 230\text{Hz}, \Gamma = \langle 500, 6500 \rangle \text{ Hz}, \mathcal{K} = 80}$$



(a) Formant frequencies (b) Formant histogram

Fig. 2: *Formant frequencies and their histogram. A high density of formant and pitch f_x within the range $\langle 500, 6500 \rangle\, Hz$ – critical bandwidth interval is apparent.*

Finally, the application of *par scaled filter-bank* together with the cepstral transformation of the data set leads to the appropriate extraction of the features. Analogically to the well known *Mel Frequency Cepstral Coefficients* we call the derived parrot-adapted features *Par Frequency Cepstral Coefficients* – PFCC.

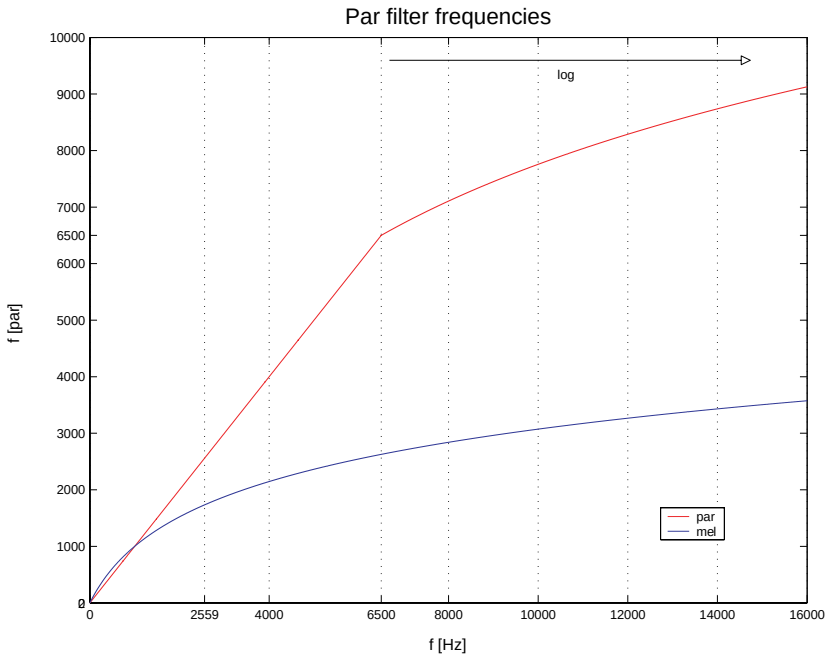


Fig. 3: *Center frequencies of the par scaled filter. The picture shows the frequency warping function f_{par} of par filter (top curve) in comparison with the original* mel-scale *f_{mel} (bottom curve).*

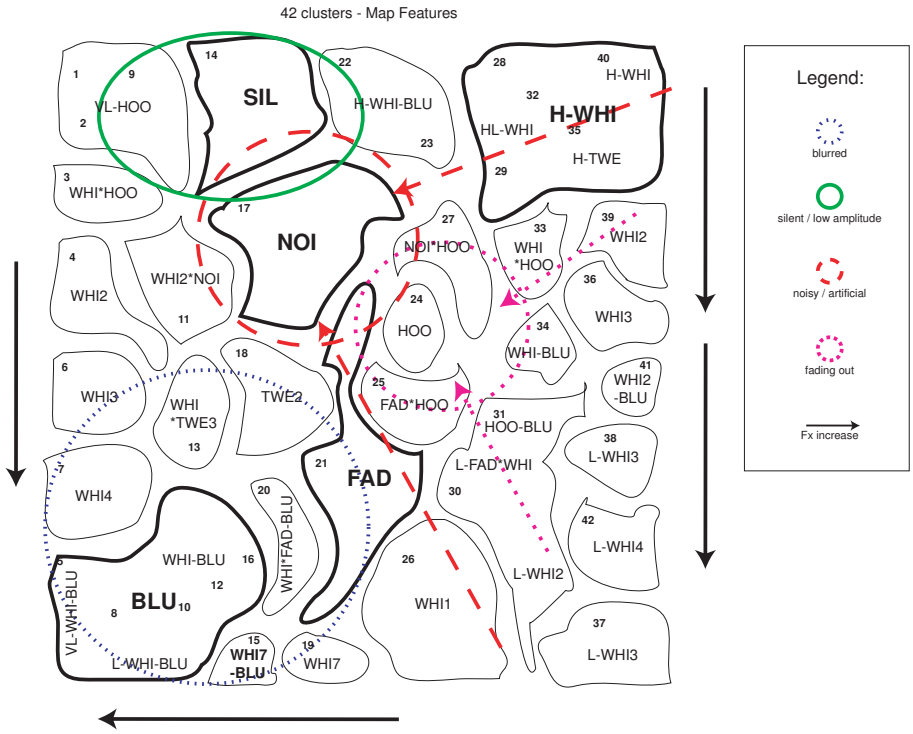


Fig. 4: *Data flow map: map overall features. To point out their inner mutual similarity, some classes were manually condensed to greater regions. Circulation of the formants and the traces of selected acoustic attributes are both depicted by arrows. We have emphasized several main classes: NOI – noise, SIL – silence, BLU – blurred (low-frequency noise), FAD – fade out and H-WHI – high-frequency whistle.*

2.2 Self-Organizing Map

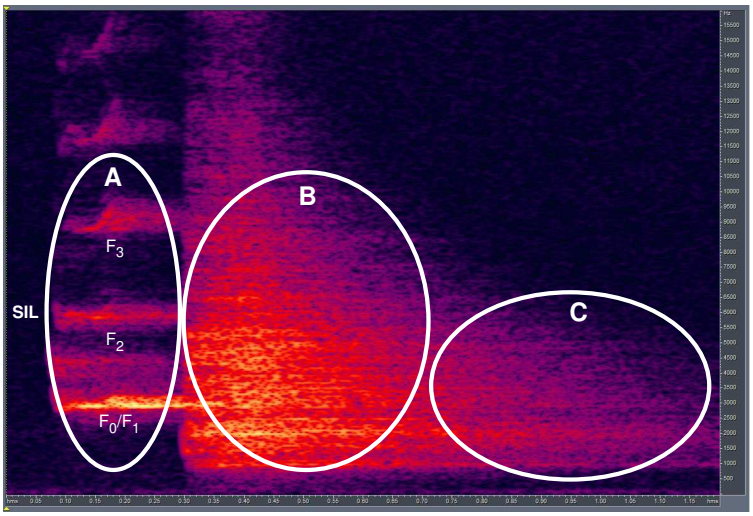
One of the neural network approaches – the self-organizing map (SOM) [5] – serves as a vector quantization and projection method applied on the acquired acoustic features. The SOM constitutes a non-linear mapping of the data onto a two-dimensional grid – the map. We clustered a resulting map and manually labeled it with names which correspond to particular acoustic classes. The number of classes is derived from often commonly used clustering methods: k-means algorithm and distance matrix clustering algorithm [17]. The final number of classes is acquired:

$$\boxed{\text{number of classes} = 42}$$

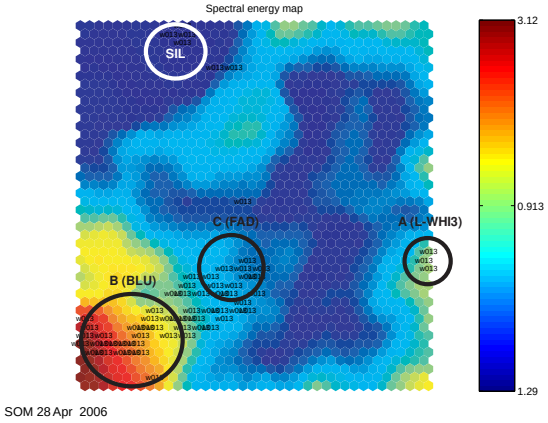
An interpretation of such a classification is depicted in Figure 4 where the main classes are emphasized. There are 42 classes, roughly representing a similar number of *phones* of parrot vocalization. Some of the classes were manually condensed to more extensive regions to point out their inner mutual similarity. The classified sound is decomposed to frames that are, sample by sample, assigned to the particular classes. The map is capable of recognizing silence (SIL) and artificial noisy sounds (NOI) in the recording. Low-frequency noise (BLU), fade-out of the sounds (FAD) or high-frequency whistle (H-WHI) are examples of other classes used in the classification process. The map also provides detailed information for every classified sound: the average number of formants found in every class. The flow of formants evokes the circulation of phones across the map.

The resolution of the map is much higher than the capability of human hearing for which most of the phones coalesce. The application of SOM supports refined distinction within a class as well.

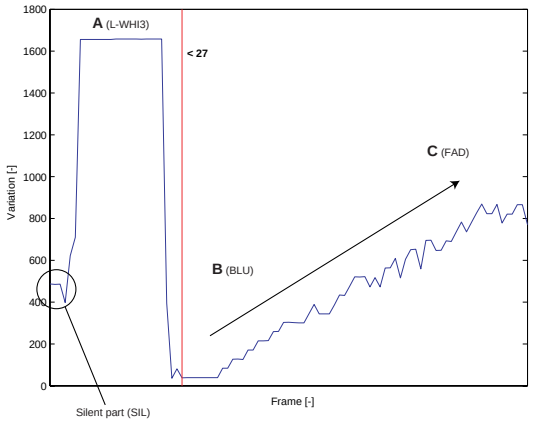
2.3 Parrot Speech Toolbox



(a) sample spectrogram



(b) classified parts of the sample shown on the top of the *spectral energy map*



(c) classifier interface. X axis is time, Y axis is variance of the classification

Fig. 5: *An example of a classification: (b) The sample is divided into four parts that are assigned to different classes: A, B, C and SIL. The different parts can be preliminarily distinguished directly in the spectrogram (a). The map provides furthermore their accurate classification by showing a sample frame assignment (c). One can distinguish whether the classification of particular frame is correct and what kind of sound it provides.*

The final map of phones, together with an interface providing the monitoring of the assignment of different samples to the classes, is called a classifier. It was implemented as the part of a robust software platform MATLAB [8] and named **Parrot Speech Toolbox**. The application provides easy-to-use classification of parrot sounds and facilitates several

visualization modes (and various data mining methods). Figure 6 shows a screenshot of the MATLAB workspace currently running Parrot Speech Toolbox and Adobe Audition [1] in the background. The toolbox supports frame-by-frame tracking of the sample classification and shows a variation of frame hits to the classes as well. Figure 5 displays an example of a classification process. In the example, a sample has been assigned to a total of four classes with a silence (SIL) class at the beginning. A spectral energy map (one which can be considered a loudness map) is used for visualization in this case.

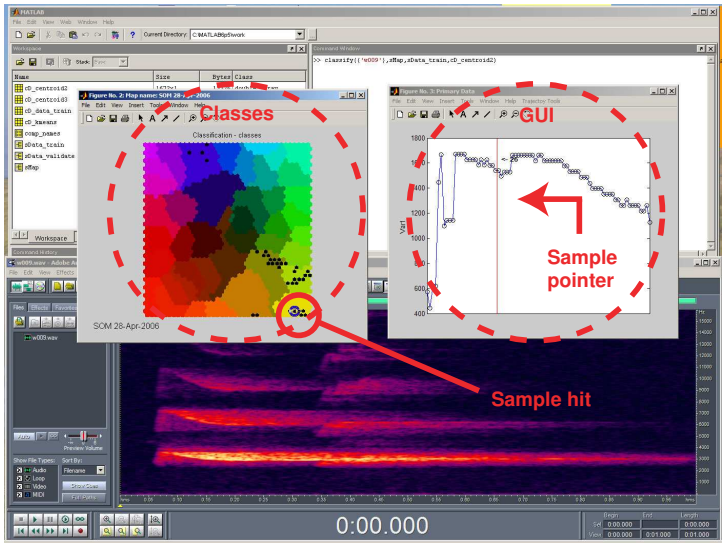


Fig. 6: *A screenshot of the workspace with the applied classification of a sound sample. A user can use a sample pointer to move across the recording frame by frame and track its class assignment. Variation within the assignment is visualized in the GUI window.*

3 Conclusion and Future Work

In this work, the use of Cepstral Transformation and Self-Organizing Map (SOM) is successfully performed on the real data – parrot vocalization. The results were encouraging. We have focused mainly on the methods of feature extraction and on the adaptation of these methods to a parrot-specific characteristic of an acoustic signal. We furthermore focused on the utilization of the above-mentioned methods for the classification system.

The **Parrot Speech Toolbox** has been implemented and it is immediately exploitable for the purposes of ethological research at the FHS – Laboratory of Interspecies Communication at Charles University in Prague. The implementation of the classifier – **classification system** that supports further exploration of the phenomena has been done. The **map of classified parrot vocalization** has been created and preliminary class labels for the samples have been proposed. A profound **analysis of SOM training algorithm** while using different learning parameters has been performed during the work. A proposal and an application of a new **combined quality measure** for the quality evaluation of the SOM has been done. A **detailed statistical survey** of the acoustical data using Principal Component Analysis [14] and common statistical methods has been performed. The design and the implementation of two variants of **par-scaled filters** used for the

feature extraction – *smooth* and *strict* filter – have been incorporated to the filter-bank. The design of a **new frequency warping scale** – *par-scale* f_{par} that is well-adapted for the parrot sound production has been drawn.

Finally, we conclude that the goals we had pointed out were accomplished in their most effective way. Hopefully, this work will inspire other researchers, fellow data engineers as well as ethologist to benefit and exploit the potential of the SOM as a versatile data mining / patterning tool. It should also encourage them to continue the research on avian vocalization and its correlations to human speech – the most important means of communication.

Future work

There are several fields of possible improvements and of future investigations. One of the main problems we are aware of is the **restricted set of recorded parrot-sounds** that were selected for the training phase of the final map – the number of recorded parrots was limited and a possible imbalance in the variety of sounds could occur. Cepstral analysis is one of the options we used for the extraction of feature vectors. There are several **other feature extraction approaches** that can be applied before the par-scaling is utilized (e.g. par-scaled FFT, LDA or wavelets [6]). Two-dimensional output space of SOM is not a requirement in any case. There are several practical examples where the usage of a **hyper-cubical SOM** lattice improved the overall quality of the quantization of TIMIT speech feature vectors, see e.g. [15]). In order to evaluate the potential number of parrot native phones (classes) more exactly, several **other clustering methods** might be performed. There are concepts of fuzzy clustering, soft competitive learning or the Learning Vector Quantizer (LVQ) [5]. Basic SOM is not designed for the **recognition of vector sequences**, though there are several variants of the algorithm that are developed for this purpose, as e.g. DTW SOM (Dynamic Time Warping) [16]. SOM itself can be alternatively utilized in the construction of HMM (Hidden Markov Models) [7], [12] which is currently the most widespread speech recognition method. **The real-time application** of the classifier needs its new implementation on a software independent platform, since the target group of users is often not acquainted with MATLAB environment.

4 Acknowledgements

The contributions and cooperation of FHS – Laboratory of Interspecies Communication at Charles University in Prague is gratefully acknowledged.

Bibliography

1. Adobe. Adobe audition. http://www.adobe.com/products/audition/.
2. Denice K. Warren, D. K. P., and Pepperberg, I. M. Mechanism of american english wovel production in a grey parrot (psittacus erithacus). *Acoustical Society of America*, 113 (1996), 41–58.
3. Doupe, A., and Kuhl, P. Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neuroscience 22*, 26 (1999), 567–631.
4. Gabriel J.L. Beckers, B. S. N., and Suthers, R. A. Vocal-tract filtering by lingual articulation in a parrot. *Current Biology 14* (2004), 1592–1597.
5. Kohonen, T. *Self-Organizing Maps*, 3rd ed. Springer-Verlag, 2001.
6. Lee, Y., and Hwang, K.-W. Selecting good speech features for recognition. *ETRI 18*, 1 (1996).
7. Leek, T. R. Information extraction using hidden Markov models. Master's thesis, UC San Diego, 1997.
8. Mathworks. Mathworks matlab 6.5. http://www.mathworks.com/products/.
9. Patterson, D., and Pepperberg, I. A comparative study of human and parrot phonation: Acoustic and articulary correlates of vowels. *Acoustical Society of America*, 96 (1994), 634–648.
10. Pepperberg, I. The alex foundation. http://www.alexfoundation.org.
11. Plannerer, B. An introduction to speech recognition. Tech. rep., University of Munich, Germany, 2003.
12. Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition. In *Proc. of the IEEE, vol. 77, No. 2* (1989).
13. Scanlan, J. *Analysis of avian "speech": patterns and production.* PhD thesis, London University College, 1988.
14. Smith, L. I. A tutorial on principal component analysis, 2002 . http://www.chem.agilent.com/cag/bsp/SiG/Downloads/pdf/pca.pdf.
15. Somervuo, P. Speech dimensionality analysis on hypercubical self-organizing maps. Tech. rep., Helsinki University of Technology, Neural Networks Research Centre, 2001.
16. Varsta, M. *Self-Organizing Maps in Sequence proccesing.* PhD thesis, Helsinki University of Technology, 2002.
17. Vesanto, J., and Alhoniemi, E. Clustering of the self-organizing map. *IEEE-NN 11*, 3 (May 2000), 586. http://citeseer.ist.psu.edu/vesanto00clustering.html.
18. Wikipedia. Formant. http://en.wikipedia.org/wiki/Formant.
19. Wikipedia. Mel scale. http://en.wikipedia.org/wiki/Mel_scale.
20. Wikipedia. Phoneme. http://en.wikipedia.org/wiki/Phoneme.