

# GRAPHICAL USER INTERFACE FOR SELF-ORGANIZING MAPS

*Petr Zetocha*

Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Circuit Theory,  
Laboratory of Artificial Neural Networks Application,

## Abstract

**This paper describes the MATLAB software package for one part of the automatic speech analysis. We are using the Self-Organizing Map (SOM) [1] application for a speech signal processing in our Laboratory of Artificial Neural Networks Application (LANNA) [4]. Experiments in [2] documented the motivation for creating new software package. This package called the SOM Laboratory (SOMLab) is more suitable for our application. It takes advantage of label information by input data and neural networks.**

**The SOMLab is a complex system which was created as a user friendly application of the artificial neural networks in speech analysis. The project consists of the tools necessary for utilizing the SOMs. Individual tools can be divided into three categories: Pre-processing tools: tools for data preparation, data creation and analysis. Processing tools: tools for working with the SOMs, such as creating and training. Post-processing tools: tools for visualization, comparison and analysis of speech. Some of main advantages are; the unsystematic errors during the hand data preparation process are eliminated by the automatic approach; an exact selection of parameters and they saved or loaded is a useful for verification results; a unification of analysis leads to a more facile team research; and in the end knowledge of the MATLAB script language or the SOM Toolbox functions is not necessary for working with the SOMLab.**

## 1 Introduction

It is necessary to oblige students to take part in projects or experiments. More suitable conditions have to be created. One of these conditions is also the SOM Laboratory, the SOMLab for short. It is a tool which enable students to make their own or collective projects. The SOMLab aims at a support of students who want to work with artificial neural networks. We are using the Self-Organizing Map (SOM) [1] application for a speech signal processing in our Laboratory of Artificial Neural Networks Application (LANNA) [5]. Experiments in [2] and [3] documented a motivation for creating a new program packed. The SOMLab takes advantage of class information by input data and neural networks.

Next part chapter describes the Self-Organizing Map. Chapter three gives an overview of the SOMLab and the last chapter describes one of many others experiments that was used the SOMLab.

## 2 Self-Organizing Map

The Self-Organizing Map, in short the SOM (Fig. 1) is a subtype of artificial neural networks. It is trained using unsupervised learning to produce low dimensional representation of the training samples while preserving the topological properties of the input space. This makes SOM reasonable for visualizing low-dimensional views of high-dimensional data, akin to multidimensional scaling. The model was first described by the Finnish professor Teuvo Kohonen and is thus sometimes referred to as a Kohonen map.

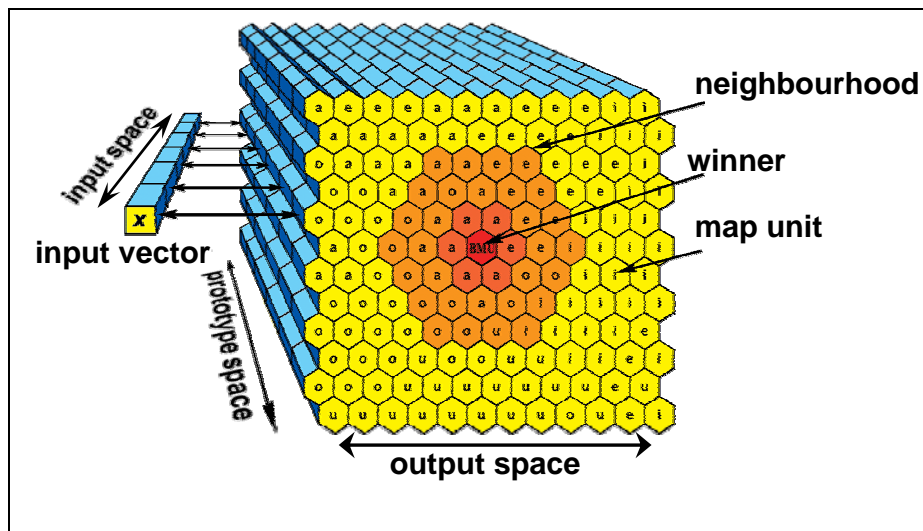


Fig. 1: Self-Organizing Map

The SOM is a single layer feedforward network where the output syntaxes are arranged in low dimensional (usually 2D or 3D) grid – an output space. Each input is connected to all output map units. Attached to every map unit there is a prototype vector with the same dimensionality as the input vectors. The number of input dimensions is usually a lot higher than the output grid dimension. SOMs are mainly used for dimensionality reduction rather than expansion.

The goal of the learning in the SOM is to associate different parts of the SOM lattice to respond similarly to certain input patterns. Like most artificial neural networks, the SOM has two modes of operation:

1. During the training process a map is built, the neural network organises itself, using a competitive process. The network must be given a large number of input vectors, as much as possible representing the kind of vectors that are expected during the mapping
2. During the mapping process a new input vector may quickly be given a location on the map, it is automatically classified or categorised. There will be one single winning map unit: the map unit whose prototype vector lies closest to the input vector. This can be simply determined by calculating the Euclidean distance between input vector and prototype vector.

### 3 SOM Laboratory

The SOMLab works under the computational system MATLAB 7, Release 14. The software, called SOM Toolbox, was used for M-files in the SOMLab in addition to basic Matlab toolboxes. The SOM Toolbox was developed in the Laboratory of Information and Computer Science (CIS) in the Helsinki University of Technology and it is built using the Matlab script language. The SOM Toolbox is available free of charge under the General Public License (GNU) from [4]. The SOM Toolbox is a software library for Matlab 5 (version 5.2 at least) implementing the SOM algorithm. The SOMLab is a graphical user interfaces extension.

Highlights of the SOMLab include the following:

- Modular programming style: the code utilizes Matlab structures and the functions are constructed in a modular manner.
- Component names, masks, labels and normalizations: to facilitate data mining process, the input vector components may be given with class information. Also, the components may be masked, or weighted, according to their relative importance.
- Batch or sequential training: in data analysis applications, the speed of training may be considerably improved by using the batch version. There are also other training variants, like Supervised SOM, LVQ1 and LVQ3.
- Map dimension: maps may be N-dimensional although visualization is not supported when

$N > 2$ .

- Advanced graphics: building on the Matlab's strong graphics capabilities, attractive figures can be easily produced for visualisation of network grid or mapping input vectors to the map.

The SOMLab consists of the tools necessary for utilizing the SOMs. Individual tools can be divided into three categories:

- Pre-processing tools: tools for data preparation, data creation and analysis.
- Processing tools: tools for working with the SOMs, such as creating and training.
- Post-processing tools: tools for visualization, comparison and analyses of speech.

The following text describes particular tools and their interconnection in the SOMLab project (Fig. 2). The main module is “som\_create” for initialization and training SOM. File \*.mat includes information about trained or initialized SOM. Select the menu called Utilities from the menubar is possible to active all others modules. A module “som\_labeling” creates phonemes boundaries according to input wav-files. These information about phonemes are saved to label-files (\*.lbl). Label-files and wav-files are used in a “som\_inputdata” module, which prepares input data (\*.dat) for the SOM. The selected phonemes from wav-files are extracted and pre-processed. Modules “som\_visualize” and “som\_analyse” analyse tested data in trained SOMs. Outputs are pictures of tested vectors in maps. Some statistic information is displayed in the “som\_analyse” module. All modules mutually communicate by extern files, like \*.dat (input data, info), \*.mat (SOM) or \*.lbl (label-file).

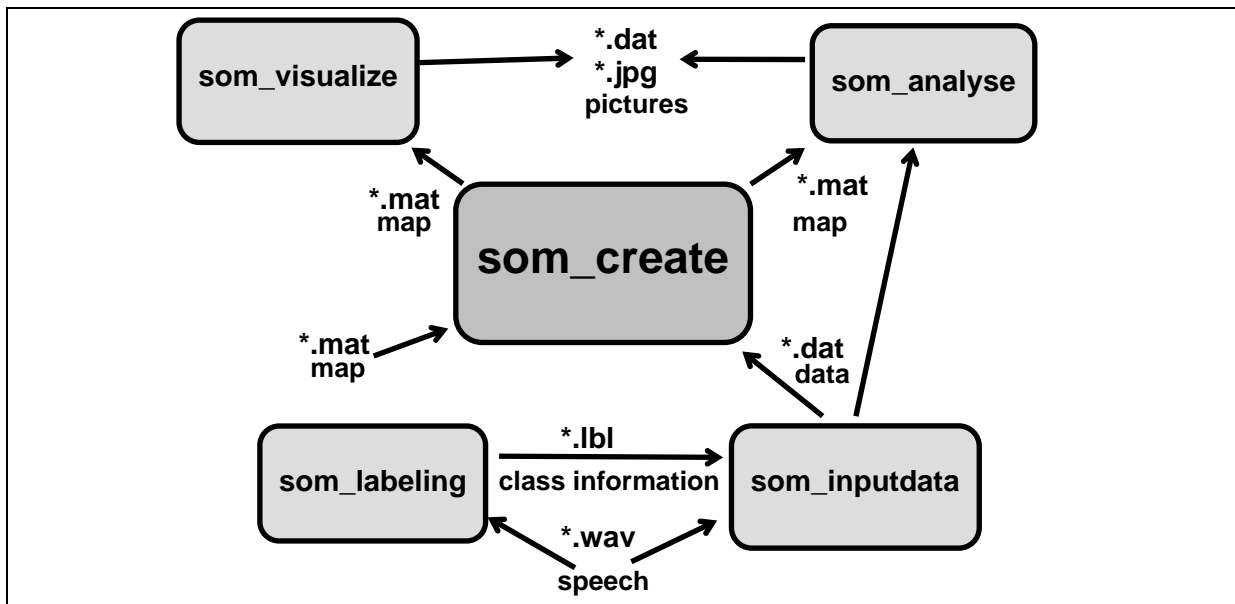


Fig. 2: The SOM Laboratory – block diagram

#### 4 The SOM Laboratory in experiment

References in [2] and [3] document the SOMLab using. There was worked on the presumption that special neurological defect, called developmental dysphasia, can influence a shift of formant frequencies in spectral characteristics compared with the formant frequencies of healthy children. For that reason, this hypothesis says that the improvement in a successful classification means the improvement in health of the patient (i.e. the reduction of developmental dysphasia).

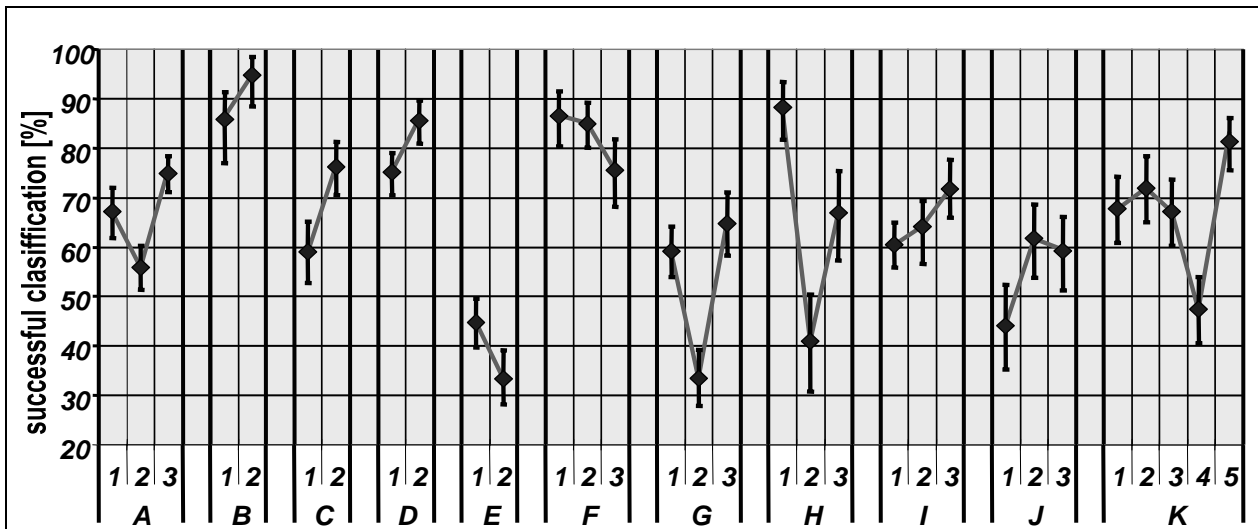


Fig. 3: Result of experiment in [3]. Numerals on the axis-x are measures. Letters are patients.

Experiments were based on a vowel classification by a Supervised Self-Organizing Maps (Supervised SOM). The Supervised SOM is one of the types of artificial neural networks with better cluster separation based on the Kohonen map. The maps trained by input data of healthy children were compared with the input data of eleven patients observed before and after their first and second part of a medical therapy. The differences of the successful classification between recorded speeches were observed by confidence intervals of a binomial distribution. The results are shown in Fig. 3. So far our hypothesis was supported by neurologists in one case. In this case the hypothesis was accepted by our results with the probability of error smaller than 5 %.

Trained and tested data sets were prepared in the “som\_inputdata” module. Maps used in experiments were initialized and trained by the “som\_create” modul.

## 5 Conclusion

In the LANNA work two groups nowadays. One group works on speech children analyses like and the second one works on an improvement of prosody. The first one includes about seven students who participate in development of the SOM Laboratory (SOMLab).

Some of main advantages of the SOMLab are; the unsystematic errors during the hand data preparation process are eliminated by the automatic approach; an exact selection of parameters and they saved or loaded is a useful for verification results; a unification of analysis leads to a more facile team research; and in the end knowledge of the MATLAB script language or the SOM Toolbox functions is not necessary for working with the SOMLab.

## Acknowledgements

Real data sets have been kindly provided by the LANNA. The work has also been supported by the Higher Education Development Fund (FRVS, Grant # 2453) „Design and realization of children speech database“.

## References

- [1] T. Kohonen, *Self-Organizing Maps*. 3-rd edition, Berlin Heidelberg, Springer Series in Information Sciences, Springer-Verlag, 2001.
- [2] J. Tučková and P. Zetocha, *Speech Analysis of Children with Developmental Dysphasia by Supervised SOM*. In International Journal on Neural and Mass-Parallel Computing and Information Systems - Neural Network World, p. 533-545, 2006

- [3] J. Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, *SOM Toolbox for Matlab 5*, SOM Toolbox Team, Helsinki University of Technology, Finland, Homepage of SOM Toolbox: [www.cis.hut.fi/projects/somtoolbox](http://www.cis.hut.fi/projects/somtoolbox)
- [4] J. Tučková, M. Bártů, P. Zetocha and M. Žůrek, *Laboratory of Artificial Neural Networks Application*, Prague, Czech Republic, [amber.feld.cvut.cz/lanna](http://amber.feld.cvut.cz/lanna)
- 

Petr Zetocha,  
Czech Technical University in Prague, Faculty of Electrical Engineering  
Department of Circuit Theory  
Address: Technická 2, 166 27 Praha 6  
Phone: (+420) 22435 2288  
E-mail: [zetocp1@fel.cvut.cz](mailto:zetocp1@fel.cvut.cz)