ARCHIVES OF ACOUSTICS
31, 4 (Supplement), 205–210 (2006)

THE APPLICATION OF KOHONEN AND MULTILAYER PERCEPTRON NETWORKS IN THE SPEECH NONFLUENCY ANALYSIS

I. SZCZUROWSKA

Agricultural University of Lublin
Faculty of Agricultural Engineering, Department of Physics
Akademicka 13, 20-950 Lublin, Poland
e-mail: izabela.szczurowska@ar.lublin.pl

W. KUNISZYK–JÓŹKOWIAK, E. SMOŁKA

Maria Curie-Skłodowska University
Institute of Informatics, Laboratory of Biocybernetics
Pl. Maria Curie-Skłodowska 1, 20-031 Lublin, Poland

(*received June 15, 2006; accepted September 30, 2006*)

Paper reports the neural network tests on ability of recognition and categorising the non-fluent and fluent utterance records. 40 of 4-second fragments containing the blockade before words starting with stop consonants (p, b, t, d, k and g) and including from 1 to 11 stop consonant repetitions and 40 recordings of the speech of the fluent speakers containing the same fragments were applied. Two various networks were examined. The first, Self Organizing Map (Kohonen network), with 21 inputs and 25 neurons in output layer, was used to reduce the dimension describing the input signals. As a result of the analysis we achieved vectors consisting of the neurons winning in a particular time point. Those vectors were taken as an input for the next network that was Multilayer Perceptron. Its various types: with one and two hidden layers, different kinds and time of learning were examined.

Key words: neural networks, speech disfluency, Kohonen network, Multilayer Perceptron network, stuttering.

1. Introduction

Artificial neural networks are used as a tool in speech analysis both of the fluent and non-fluent speakers [1, 2, 4, 6, 7, 9]. It is due to their characteristic structure, which is patterned on model showing only basic essence of brain working, the learning process similar to that of human [6–8] and the possibility of realizing a part of perception of both: a non-fluent speaker and a hearer [6–9]. Their most important abilities are that they could solve nonlinear problems and reduce dimensions describing investigating issues, which helps us to propose a proper solution. The knowledge of all the principles accompanying non-fluent signals can help to create a recognition of speech and an au-

tomatic system of diagnosing the speech disfluency types. It could also enable to devise the right kind of therapy and follow its progress.

The authors of the following article applied Kohonen and Multilayer Perceptron Networks to recognize and classify fluent and disfluency containing speech utterances.

2. Methodology

The research material were recordings taken from eight stuttering people by using the Sound Blaster with the sampling frequency 22050 Hz and sampling precision of 16 bits. The utterances were recorded before a therapy as well as during its various stages and included two situations: reading story fragments and describing illustrations. All of the patients had a blockade before words starting with stop consonants (p, b, t, d, k and g) and contained from 1 to 11 stop consonant repetitions. That type of disfluency was chosen because of difficulty with eliminating it during therapy and, as previous researches showed [6–7], is well recognised by artificial neural networks.

Table 1. Stuttering people characteristic.

Person initials	Sex	Age (years)	Intensity of the stuttering (in number of mistakes on hundred words)	
			Reading	Narrating
P	F	12	1	1
MJ	F	21	31	17
MSZ	M	11	42	36
RK	M	13	73	67
RCZ	M	23	9.5	33
AM	M	16	20	27
PH	M	18	9	42
Pap	M	23	11	37.5

From the recordings, forty of 4-second fragments containing disfluency were selected and the speech of fluent speakers containing the same fragments was recorded. Following this, all utterances were analysed by FFT 512 with the use of a 21 digital 1/3-octave filters of centre frequencies between 100 and 10000 Hz and an A-weighting filter. FFT time resolution was 23 ms, which transformed every 4-second sample into 21 vectors consisting of 171 time points. The first network (Fig. 1), with 21 inputs and 25 neurons in output layer [6] was used to reduce the dimension describing the input signals. The network was thought with following parameters: training time – 100 epochs, learning rate – 0.1 and neighbourhood – 1.

As a result of the signal analysis (Fig. 2a) vectors consisting of the number of the neuron winning in a particular time point (Fig. 2b) for non-fluent (on the left) and fluent sample were obtained.

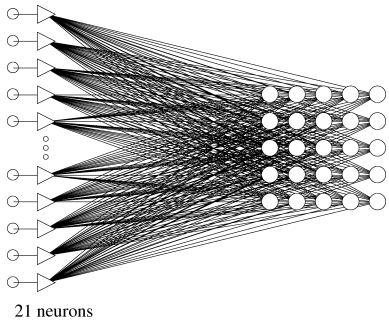

Fig. 1. Kohonen network with 21 inputs and 25 neurons in output layer.

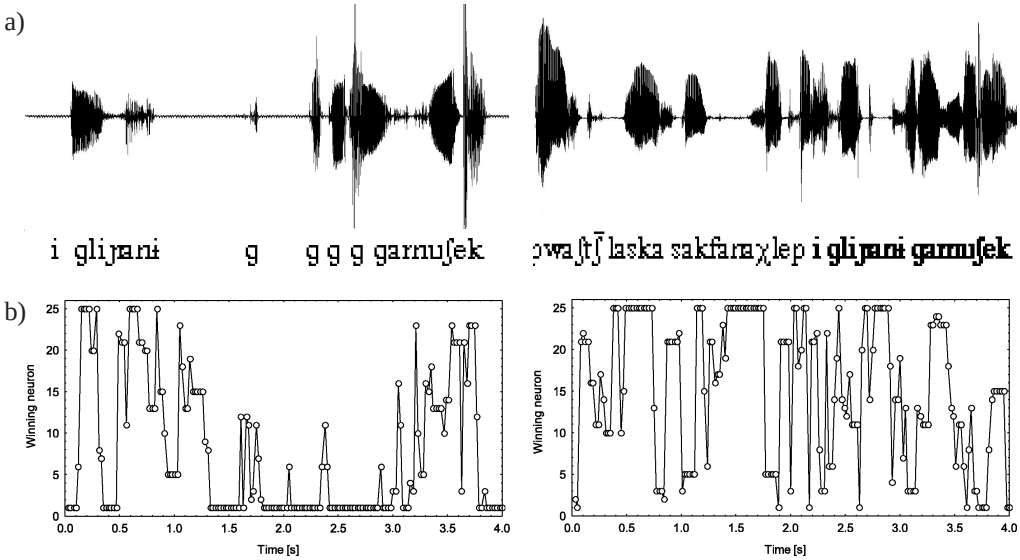

Fig. 2. Oscillograms (a) and number of winning neuron (b) for non-fluent (left) and fluent sample (right).

The numbers of the winning neurons were put into one data table that was taken as an input for the Multilayer Perceptron Network. Data table was divided into three groups: teaching (40 vectors), verifying (20 vectors) and testing (20 vectors). Various types of MLP (Fig. 3) were tested and various kinds of learning algorithm were used to check their ability to classify samples into two groups. All tested networks had 171 neurons in input layer due to receiving 171 time points for each sample, and 1 neuron in the output because answer “non-fluent” or “fluent” (which means that output neuron was activated or not) was expected. Networks have one or two hidden layers with different number of neurons.

All networks were subjected to the process of teaching by using back propagation algorithm throughout one hundred epochs. Pre- and post processing values were established as 0.7 and 0.4. As an activation function logistic function was used, error was calculated using cross-entropy error function. This error is the sum of the products of

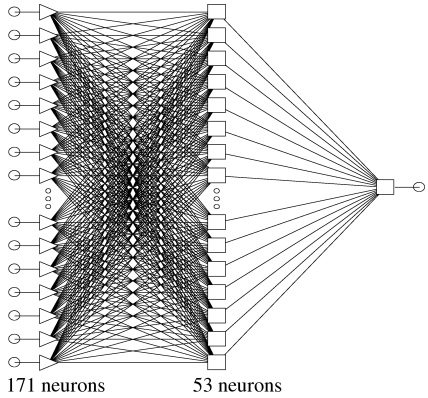

Fig. 3. MLP with 53 neurons in hidden layer.

the target value and the logarithm of the error value on each output unit. The cross-entropy error function is specially designed for classification problems, where it is used in combination with the logistic (single output).

Two best networks were checked with thirty samples that did not take part in the learning. Fifteen of them were non-fluent, second half were fluent equivalents.

3. Results

Table 2 shows effects of learning process on chosen MLP networks. Not all utterances used in that learning were categorised properly. All fluent samples were grouped rightly, only five non-fluent were recognized as fluent. Too short time period between the last repetition and the fluent part of an utterance (Fig. 4), number of repetitions (Fig. 5) and total break time (Fig. 6) may have been the reason for the wrong allocation to the group.

Table 2. MLP architectures and classification effects.

Network	Number of neurons				Learning parameters		Best classification [%]		
	Input	Hidden		Output	Learning rate	Momentum	Training	Verifying	Testing
		I	II						
1	171	18	18	1	0.5–0.1	0.2	100	85	85
2	171	18	–	1	0.5	0.3	100	85	85
3	171	53	–	1	0.4	0.3	100	90	85
4	171	121	121	1	0.3	0.2	100	85	85
5	171	35	35	1	0.7–0.2	0.6	100	85	85
6	171	86	–	1	0.6	0.3	100	90	85

MLP marked as 3 and 6 (Table 2) achieved the best equivalent result so they were checked on their ability to classify not-known samples. Network build of 53 neurons

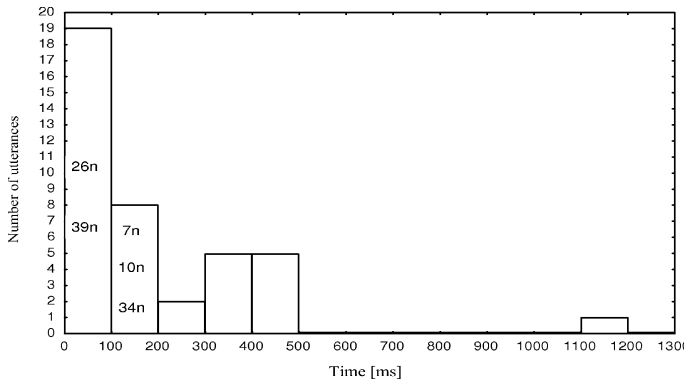

Fig. 4. Length of the pauses after the last repetition with wrongly grouped utterances (n-non-fluent).

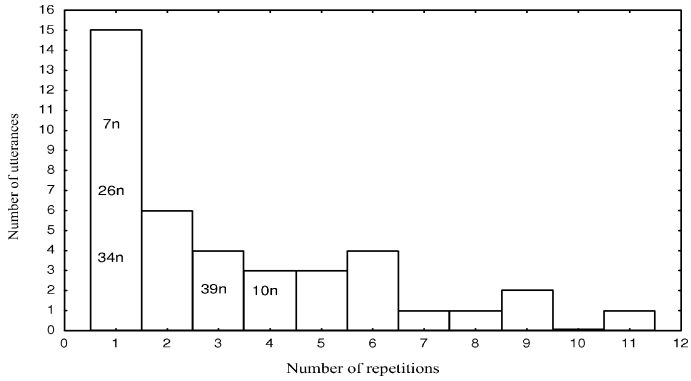

Fig. 5. Relation between the number of utterances and the number of repetitions with wrongly grouped utterances (n-non-fluent).

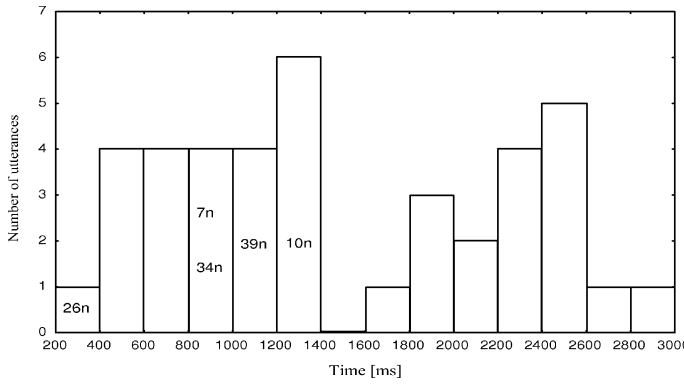

Fig. 6. Total breaks time with wrongly grouped utterances (n-non-fluent).

in hidden layer (number 3) was a little better than second one (number 6). Those MLP correctly classified 23 (when network with 86 neurons in hidden layer classify 22) of utterances which account for 76.67% of correct answers.

4. Conclusion

Our research has shown that artificial neural networks can be a useful tool in speech analysis, especially non-fluent one. Application of the first network allow us to reduce the dimensions describing the input signals and make possible to say that Kohonen network can be used in speech describing. As we can notice Kohonen network gives syllabic structure of utterance very well, exposing fragments containing disfluency. Multilayer Perceptron Network was checked out for correct classification samples into two groups. The best network, built of 171 input neurons, 53 neurons in hidden layer and one output, managed to classify correctly 76.67% of samples.

Neural networks are the area of great possibilities in a field of speech researches. It can help people suffering from various speech disorders by facilitating diagnostic procedures and helping to choose the right therapy. Although more detailed analysis is required, it is a promising start.

Acknowledgments

The research was supported by Grant Deputy Rector for Science of Maria Curie-Skłodowska University.

References

[1] IZWORSKI A., WSZOŁEK W., *Artificial intelligence methods in diagnostics and processing of the pathological acoustics signals* [in Polish], [in:] Speech and Language Technology, vol. 3, 299–319, Poznań 1999.

[2] LEINONEN L., KANGAS J., TORKKOLA K., JUVAS A., *Dysphonia detected by pattern recognition of spectral composition*, J. Speech Hear. Res., **35**, 287–295 (1992).

[3] LEINONEN L., HILTUNEN T., LAAKSO M.,L., POPIUS H., *Categorization of voice disorders with six perceptual dimensions*, Folia Phoniatr. Logop., **49**, 9–20 (1997).

[4] LEINONEN L., HILTUNEN T., TORKKOLA K., KANGAS J., *Self-organized acoustic feature map in detection of fricative-vowel coarticulation*, J. Acoust. Soc. Am., **93**, 6, 3468–3474 (1993).

[5] MUJUNEN R., LEINONEN L., KANGAS J., TORKKOLA K., *Acoustic pattern recognition of /s/ misarticulation by the self-organsing map*, Folia Phoniatr., **45**, 135–144 (1993).

[6] SMOŁKA E., KUNISZYK–JÓŹKOWIAK W., SUSZYŃSKI W., *Reflection of fluent and non-fluent words in Kohonen Network* [in Polish], XLIX Open Seminar on Acoustics OSA'2002, 371–376, Warszawa – Stare Jabłonki 2002.

[7] SMOŁKA E., KUNISZYK–JÓŹKOWIAK W., SUSZYŃSKI W., DZIEŃKOWSKI M., SZCZUROWSKA I., *Speech nonfluecy recognition in two stages of Kohonen Networks*, Structures-Waves-Human Health, 139–142, Zakopane 2004.

[8] TADEUSIEWICZ R., *Speech recognition with application of neural networks* [in Polish], Seminar of Polish Phonetical Society, 137–150, Warszawa 1994.

[9] WSZOŁEK W., TADEUSIEWICZ R., *The evaluation of effectiveness of various neural network types in pathological speech analysis* [in Polish], XLVII Open Seminar on Acoustics OSA'2000, vol. II, 721–728, Rzeszów – Jawor 2000.