Voice Pathology Distinction Using Autoassociative Neural Networks

Daria Hemmerling
AGH University of Science and Technology
Department of Measurement and Electronics
Al. Mickiewicza 30
30-059 Krakow, Poland
Email: hemmer@agh.edu.pl

***Abstract*—Acoustic analysis is a non-invasive technique that supports voice disease screening, especially the detection and diagnosis of distinction between chosen voice pathologies and healthy control group. This work put en effort on creation of efficient and accurate system for automatic detection and differentiation of normal and three different voice pathologies. This system ensures non-invasive and fully automated analysis of voice characteristics and decision system based on neural networks. The feature vector describing the vocal tract is set up from 35 parameters. Recordings of patients suffering from hyperfunctional dysphonia, recurrent laryngeal nerve paralysis, laryngitis and healthy control group are considered in our experiments. From the experimental results it is observed that effectiveness of auto-associative neural networks seems to be promising in the application of pathological voice distinction.**

I. Introduction

Voice as a physical and mental health barometer requires comprehensive evaluation. It is a multidimensional phenomenon which consists of a number of elements that influence the overall voice quality and voice characteristics. Every voice is unique. Due to the modern requirements imposed on the clinics, a predetermined huge number of patients and the short time in which a doctor should make an examination, initiates the formation of new devices, algorithms to accelerate the diagnosis. The main method used by the medical community to evaluate a speech production system and diagnose pathologies is direct ones, which requires direct inspection of vocal folds and cause discomfort to the patient [1].

Recently, there is the development of advanced digital processing methods for voice analysis. Classical Fourier acoustic analysis of voice is increasingly being supplemented by the non-linear methods [2], [3]. Non-linear analysis using cepstral coefficients allows to determine precise parameters describing signal excitation (ie. the sounds produced directly by the glottis) on the basis of recorded speech samples [4]. In the literature, several methods for detecting speech pathologies have been presented. Among various pathological voice detection algorithms wavelets [5], neural maps and networks [7], [8] and fractals [9] have been the most often employed. What is more, parameters such as pitch analysis, jitter, shimmer, harmonic to noise ratio, mel- frequency cepstral coefficients (MFCC) has been suggested for improving voice pathology detection systems [10], [11]. The classification systems of pathological voices consists of multi-layer perception [12], Gaussian mixture model [13], Support Vector Machine, hidden Markov model [1], probabilistic neural network [14] and linear discriminant analysis [24], [16]. Most of the methods in the literature focus on binary classification to detect whether the voice is classified as normal or pathological, they do not detect the type of voice pathology. The differentation between 5 voice pathologies is presented in the paper [17]. The classification accuracy is in the range 61% and 69%. The paper [1] presents the automatic system to detect the type of 10 voice pathologies with accuracy being at the level of 65% - 81%. The total number of patients taken into the analysis was 108. Authors of the paper [5] present the classification baed on wavelet packet based features of 3 voice pathologies - A-P squeezing, gastric reflux, hyperfunction and normal voice. The accuracy based on energy and entropy is 96-97%.

The objective of this work is to detect the pathological voice and point the type of pathology. The system consists of the 35-dimensional feature vector made up from long- and short-term parameters. The vector is the input to the classification system based on Auto-associative Neural Networks (ANN). The advantage of ANN is the ability to capture the distribution of feature vectors. By applying ANN, the algorithm learn the boundary regions between patterns belonging to the selected classes by mapping the input patterns into a different dimensional space [6]. The block diagram of proposed system is shown in figure 1.

II. Database

The data used in this work was downloaded from Saarbrucken Voice Database (SVD). Svd is published online and is the property of the Institute of Phonetics of the University of the Saarland [18]. This SVD contains voice recordings from more than 2000 healthy and pathological German speakers. The patients and healthy candidates whose recordings are included in this database phonated vowels: of the /a/, /i/, /u/ vowels spoken at four different pitches: high, normal, low and high-low-normal. The sampling frequency of all the recordings was 50 kHz with 16-bit resolution. Duration of eacg the recording was in the range between 1 and 4 seconds.

ISBN 978-0-9928626-7-1

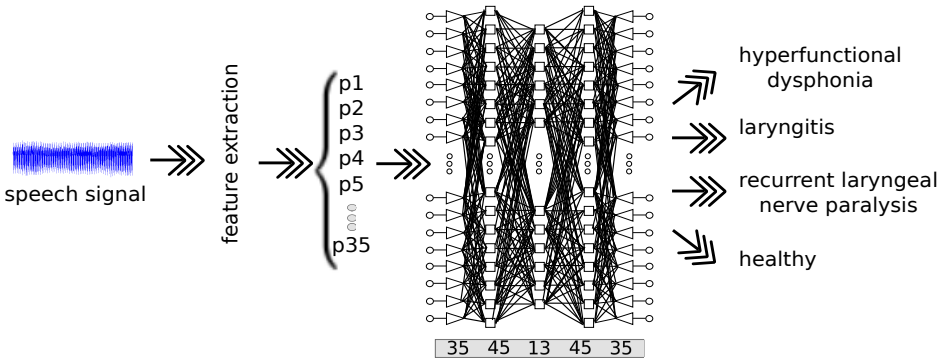

Fig. 1. Block diagram of proposed system.

TABLE I
NUMBER OF PATIENTS TAKEN INTO THE EXAMINATION.

name of the pathology	female	male
hyperfunctional dysphonia	166	45
laryngitis	57	82
recurrent laryngeal nerve paralysis	138	74
healthy	361	201

In this work we used the recordings of vowel /a/ phonated at normal pitch. In the analysis we included recordings of three voice pathologies: hyperfunctional dysphonia, recurrent laryngeal nerve paralysis and laryngitis. The exact numbers of patients whose recordings were analysed in this work are presented in table 1. The number of recordings of healthy candidates was the same as for pathology cases.

III. FEATURE EXTRACTION

Feature extraction is a major part to measure the voice's quality and has been an important area of research for many years. Selection of the features plays an important role and is the main influence on the voice pathology system accuracy. In this paper 35 quantitative voice parameters were computed. The analysis presented in this paper includes the acoustic parameters such as frequency perturbation, amplitude parameters and those characterizing the vibrations of the vocal folds (cepstral parameters). These features are used to analyse aperiodic voice signal with a constant dominant feature, which determines the same fundamental frequency in the course of the acoustic wave [19]. Those parameters contain: fundamental frequency for the vocalisation, jitter coefficient that gives the evaluation of the variability of the pitch period within the analysed voice sample, shimmer coefficient that gives the evaluation of the variability of the peak to peak amplitude, noise to harmonic ratio measures between the frequencies 0-500 Hz, 0-1.5 kHz, 0-2.5 kHz, 0-3.5 kHz to measure the noise present in the vocalization, energy, zeroth-, first-, second-, third-order moment, power factor, 1-, 2-, 3- formant amplitude, 1-, 2-, 3- formant frequency, 12 mel-frequency cepstrum coefficients (MFCCs), approximate entropy [20], kurtosis, turbulent noise index and normalized first harmonic energy [22].

IV. ANN MODEL FOR CAPTURING VOICE PATHOLOGY

Auto-associative Neural Networks are feed forward neural networks using identity mapping of the input space. ANN are used to capture the distribution of the input data [22], [23], [25]. The advantage of ANN is the ability to represent an input space using a non-linear subspace. It is a neural network, wherein the input layer has the same size (same number of neurons), as output layer. In addition, learning of this network aims to faithfully reproduce the output of accepted input. The rationale for the use of such a network structure is the fact that between the input layer and the output layer of the network there is usually at least one hidden layer of neurons, having significantly less number of neurons than the input and output layers. In this intermediate layer compressed representation of the data is produced. What is more, this part of the network that extends from the intermediate layer to the output layer becomes a tool for decompression. The number of hidden layers and neurons in each hidden layer depends on the problem. In the three layer auto-associative neural network activation function on hidden layer transfer the input data in a linear subspace. However, to take full advantage of the non-linear network's we need more than one layer of neurons for each of the two ongoing transformation (both for compression and for decompression). Therefore to create a network for non-linear data compression the topology of the neural network should be composed out of 5 layers [2], [23], [25]. Hidden middle layer is a layer that reduce the dimension of the input signal, and the layer located between the hidden middle layer and the input layer carries out the required non-linear compression of input data. Accordingly, layer network structure, located between the hidden layer and output layer performs the inverse transformation (decompression) of the signal with previously reduced dimensionality. The block diagram of designed system is shown in figure 1. The ANN model is created to capture the distribution of exact voice pathology.

In this study the structure of the ANN model consists of 5 layers: 35L 45N 13N 45N 35L, where L is a linear unit, N is a non-linear unit. During ANN training, the weights of the neural network are adjusted to minimize the mean square error obtained in each feature vector by back propagation algorithm [24]. When the adjustment of weights is finished

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE II
CONFUSION MATRIX [%] FOR FEMALE CLASSIFICATION, HF - HYPERFUNCTIONAL DYSPHONIA, LAR - LARYNGITIS, PAR- RECURRENT LARYNGEAL NERVE PARALYSIS

		output results			
		HF	lar	par	healthy
training data	HF	76,25	17,51	6,25	0,00
	lar	17,86	73,21	8,93	0,00
	par	18,38	2,94	78,68	0,00
	healthy	0,00	0,00	0,00	100,00

for all the features, the network is trained for the epoch. The number of epochs is chosen when the minimum value of mean square error is reached. When the error between the input and output vectors is minimized, the cluster of points in the input space determines the shape of hyper surface obtained by the projection onto the lower dimensional space [22], [26]. The training process ensures the best adaptation and modification of weights in the response to the training input patterns, which are present at the input layer. The weights are calculated using a recursive algorithm, which starts at the output nodes and works back to the hidden layer [27].

V. EXPERIMENTS AND RESULTS

The data used in this experiment includes recordings of vowel /a/ of 3 voice pathologies, each of them were described using acoustic features. Each of the voice pathology consists of 4 ANN models representing each pathology and healthy group.

The ANN training algorithm work as follows:

1) Divide and select the training data from the training set to the network.
2) Run the ANN and compute the output of the network.
3) Compute the error e.
4) Adjust the weights to ensure that the error e is minimized.
   To evaluate the performance of our system the 4 ANN models must be calculated. To do this, the feature vectors from the test groups become the input to presented ANN models. To compare the output of the each model with the initial input we calculated normalized squared error e is calculated. The error e is defined as:

$$e = \frac{||y - o||_2^2}{||y||_2^2} \tag{1}$$

   where y is the input feature vector, o is the output vector. Based on error e a confidence score c is computed, where c is defined as c=exp(-e). The highest confidence score indicates the voice pathology.
5) Repeat steps 1 to 4 for all the training samples.
6) Repeat steps 1 to 5 until the network recognizes the training set or for certain epochs.

Before the application of the ANN, the data were normalised. Figure 2 presents the distribution of the 3 selected parameters from ANN of vowel /a/ for women.

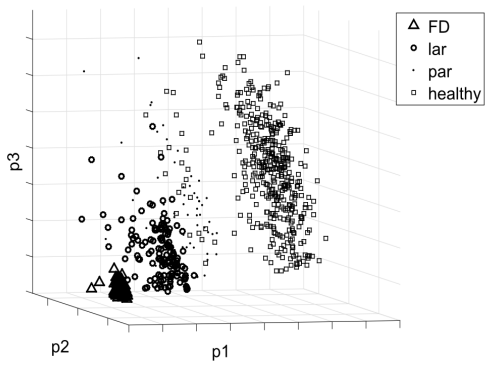

Fig. 2. Presentation of the selected three parameters from ANN of vowel /a/ for women for 3 voice pathologies (hyperfunctional dysphonia - HF, laryngitis - lar, recurrent laryngeal nerve paralysis - par and healthy group)

TABLE III
CONFUSION MATRIX [%] FOR MALE CLASSIFICATION, HF - HYPERFUNCTIONAL DYSPHONIA, LAR - LARYNGITIS, PAR- RECURRENT LARYNGEAL NERVE PARALYSIS

		output results			
		HF	lar	par	healthy
training data	HF	70,00	7,50	22,50	0,00
	lar	15,00	75,00	10,00	0,00
	par	13,89	9,72	76,39	0,00
	healthy	0,00	0,00	0,00	100,00

To assess the generalization capabilities of the system, a 8-fold cross-validation was performed in this experiment. The dataset was randomly split in 8 different subsets and the classification process was repeated 8 times. At each time a different subset was used for testing the performance. When cross-validation is finished the results are averaged across all the repetitions. The validation set was divided into two subsets: a training set and a test set. The validation set do not take part in learning process of the algorithm and enables estimation of the parameters. In order to evaluate the performance of classification F1 score have been taken into account:

$$F1 = 2\frac{p \cdot r}{p + r} \cdot 100[\%] \tag{2}$$

where: p- precision, is the ratio of a number of correctly classified positive examples and sum of all cases classified as positive and r- recall, which is defined as the ratio of all correctly classified positive cases to the sum of cases that were actual positive in the test set. F1 score is the harmonic mean of precision and recall.

Tables II and III present the percentage of patients, who were classified to each voice pathology or to healthy cases. Performance evaluation was measured by a confusion matrix being the result of 8-fold cross validation that confirmed the superiority of the used feature set. The mean F1 values (after cross validation) for each group is presented in table IV.

The mean F1 values of proposed classification system for each analysed group of voice pathologies and group of healthy

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE IV
MEAN F1 VALUES AFTER CLASSIFICATION [%].

name of the pathology	women	men
hyperfunctional dysphonia	77±0, 06	61±0, 16
laryngitis	63±0, 10	79±0, 08
recurrent laryngeal nerve paralysis	83±0, 10	77±0, 08
healthy	100±0, 00	100±0, 00

candidates are between 63-100% for women and 61-100% for men. From the experimental results, it is observed that auto-associative neural network enables 100% correct binary classification and point if patient has a voice abnormalities or the voice do not indicate any irregularities, both for female and male recordings. The overall system accuracy is 87,5%, the same for women and men samples.

VI. CONCLUSION

Proper preparation of representative learning data set is a fundamental task, both in the educational process of neural networks as well as their usage. Discussed above preparation procedure of a feature set describing the input acoustic signal to educate chosen neural network can be treated as a preprocessing step (preliminary data preparation to be acceptable by the network). The technique used for non-linear data compression with a usage of ANN with the structure of multilayer perceptron is efficient method that allows a significant learning vector dimension reduction without loss of any signal information. From the experimental results, it is observed that the binary classification to give the information if the voice pathology exists is possible and ensures 100% correct results. To identify the type of the pathology and the healthy state the F1 value was cpl alculated for each of analysed groups and its result is balanced between 61-100 %. Through the use of described processing method it was possible to effectively generate neural network topology for efficient identification of selected voice diseases. The proposed system can be used as a valuable tool for speech pathologists to help in decision making system and detect specific type of voice pathology.

ACKNOWLEDGMENT

This work was funded by the Ministry of Science and Higher Education in Poland under the Diamond Grant program, no. 0136/DIA/2013/42 (AGH 68.68.120.364).

REFERENCES

[1] S. Jothilakshmi, *Automatic system to detect the type of voice pathology*, Applied Soft Computing, 21, 244-249, 2014
[2] C. D. Maciel, J.C. Pereira, D. Stewart, *Identifying healthy and pathologically affected voice signals*, IEEE Signal Processing Magazine, 27(1), 2010
[3] K. Werth, D. Voigt, M. Dllinger, U. Eysholdt, J. Lohscheller, Clinical value of acoustic voice measures: a retrospective study. European Archives of Oto-Rhino-Laryngology, 267(8), 1261-1271, 2010
[4] B. R. Kumar, J. S. Bhat, N. Prasad, *Cepstral analysis of voice in persons with vocal nodules*, Journal of Voice, 24(6), 651-653, 2010
[5] A. Akbari, M.K. Arjmandi, *An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features*, Biomedical Signal Processing and Control, 10, 209-223, 2014
[6] S.V. Gangashetty,C.C Sekhar, B. Yegnanarayana, *Spotting consonant-vowel units in continuous speech using alitoassociative neural networks and support vector machines*, In Machine Learning for Signal Processing,Proceedings of the 2004 14th IEEE Signal Processing Society Workshop, 401-410, 2004
[7] S. Hadjitodorov, B. Boyanov, B. Teston, *Laryngeal pathology detection by means of class-specific neural maps. IEEE Transactions on Information Technology in Biomedicine*, 4(1), 68-73, 2000
[8] D. Hemmerling, A. Skalski, J. Gajda, *Voice data mining for laryngeal pathology assessment*, Computers in biology and medicine, 69, 270-276, 2016
[9] Z. Ali, I. Elamvazuthi, M. Alsulaiman, G. Muhammad, *Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals*, Journal of medical systems, 40(1), 20, 2016
[10] J.I. Godino-Llorente, P. Gmez-Vilda, *Automatic detection of voice impairments by means of short-term Cepstral parameters and neural network based detectors*, IEEE Transactions on Biomedical Engineering, 02, 51(2), 380-384, 2004
[11] J.I. Godino-Llorente, P. Gmez-Vilda, *Blanco-Velasco, M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term Cepstral parameters* IEEE Transactions on Biomedical Engineering, 53(10), 1943-1953, 2006
[12] L. Salhi, T. Mourad, A. Cherif, *Voice disorders identification using multilayer neural network*, Int. Arab J. Inf. Technol. 7(2), 8, 177-185, 2010
[13] X. Wang, J. Zhang, Y. Yan, *Discrimination between pathological and normal voices using GMM-SVM approach*, Journal of Voice, 25(1), 38-43, 2011
[14] V. Srinivasan, V. Ramalingam, P. Arulmozhi, *Artificial Neural Network Based Pathological Voice Classification Using MFCC Features*, Int. J. Science, Environment Technology, 3(1), 291-302, 2014
[15] D. Panek, A. Skalski, J. Gajda, *Quantification of linear and non-linear acoustic analysis applied to voice pathology detection*, Information Technologies in Biomedicine, 4, 355-364, 2014
[16] A. Akbari, M.K. Arjmandi, *An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features*, Biomedical Signal Processing and Control, 10, 209-223, 2014
[17] A. A. Dibazar, T. W. Berger,S.S. Narayanan, *Pathological voice assessment*, Engineering in Medicine and Biology Society, 28th Annual International Conference of the IEEE, 1669-1673, 2006
[18] http://www.stimmdatenbank.coli.uni-saarland.de/
[19] P. Henrquez, J.B. Alonso,M.A. Ferrer, C.M. Travieso, J.I. Godino-Llorente, F. Daz-de-Mara, *Characterization of healthy and pathological voice through measures based on nonlinear dynamics*, IEEE transactions on audio, speech, and language processing, 17(6), 1186-1195, 2009
[20] B.S. Aghazadeh, H. Khadivi, M. Nikkhah-Bahrami, *Nonlinear analysis and classification of vocal disorders*, In Engineering in Medicine and Biology Society, EMBS 2007, , pp. 6199-6202, 2007
[21] S. Hadjitodorov, P. Mitev, *A computer system for acoustic analysis of pathological voices and laryngeal diseases screening*, Medical engineering & physics, 24(6), 419-429, 2002
[22] P. Dhanalakshmi, S. Palanivel, and Vennila Ramalingam, *Classification of audio signals using AANN and GMM*, Applied Soft Computing 11(1), 716-723, 2011
[23] M. Bianchini, P. Frasconi, M. Gori, *Learning in multilayered networks used as autoassociators*, IEEE Transactions on Neural Networks, 6(2), 512-515, 1995
[24] J. Sangeetha, S. Jothilakshmi, *A novel spoken document retrieval system using Auto Associative Neural Network based keyword spotting*, Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on. IEEE, 2015.
[25] S. Jothilakshmi, *Spoken keyword detection using autoassociative neural networks*, International Journal of Speech Technology, 17(1), 83-89, 2014
[26] B. Yegnanarayana, S.P. Kishore, *AANN: an alternative to GMM for pattern recognition*, Neural Networks, 15(3), 459-469, 2002
[27] S.F. Gurgen, K. Aikawa, K. Shikano. *On the training strategies of neural networks for speech recognition*, Neural Networks, International Joint Conference on. vol. 4. IEEE, 1992.