Analysis of verbal disorders with the help of neural networks

Menshakou P.A.
Gomel State Technical University named after P.O. Suhoy
Gomel, Belarus
pmenshakov@gmail.com

***Abstract*—At present, voice identification and voice analysis are one of the foremost directions, both in the field of information security systems and in the definition of voice parameters [1]. It is proposed to consider a method for analysing the state of the voice and identifying its problems, such as fatigue of the vocal cords, damage or inflammation of the vocal tract with the help of the previously recorded voice of the speaker and neural network analysis of voice changes.**

***Keywords*—Voice, identification, neural networks.**

I. Introduction

Evaluation of voice impairment involves the use of the evaluation of multiple speech characteristics (degree of respiration, tension, roughness). One of the problems associated with the use of multidimensional data is their comparison. To perform the comparison and classification it is proposed to use the self-organizing map of Kohonen. In view of the possibility of learning without a teacher, it does not need a target vector for outputs and, consequently, does not require comparison with predetermined ideal answers, and the training set consists only of input vectors. The learning process, therefore, highlights the statistical properties of the learning set and groups similar vectors into classes. The input of a vector from a given class will give a certain output vector.

II. Hardware implementation

To implement the device you need a microphone, filter and analog-to-digital converter, for further work with digital voice recording. The circuit of the device is shown in Figure 1.

From the output of the microphone, the signal is fed to the input of the filtration unit. The next step is the passage of the ADC [2]. Further, the digitized signal enters the digital processing unit. In the digital processing block, the signal is filtered and converted into a vector, with which the microprocessor and the neural network processor will continue to operate. For the subsequent comparison with the previously saved vector of chalk-cepstral coefficients, the obtained vector is stored in non-volatile memory. After comparing the vector in memory with the resulting vector, the micro controller instructs the control unit of the external device, for example, on the magnetic door lock. The process of voice identification is not demanding of resources, and consists of two stages. The first stage is the receipt of the speech characteristics of the announcer and the conversion to a form in which it can be compared with others. The second step is to compare them with a trained neural network [3].

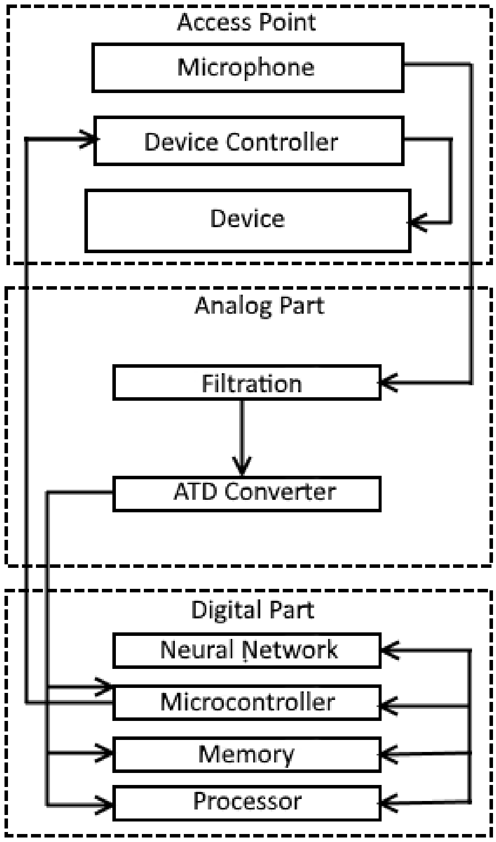

Figure 1. Hardware scheme

III. Dysphonia

Dysphonia is clearly defined as the underlying disorder of phonation, a consequence of diseases or pathology of the vocal cords. Since this deviation of the work of the vocal tract is accompanied by an audible change in the voice, it can be recorded and processed in relation to the sample of the voice of the announcer, before the appearance of deviations. There are two types of dysphonia - functional and damaging. Damaging dysphonia is divided into two types: congenital and acquired dysphonia. Damaging dysphonia in some cases may appear after functional dysphonia. Functional dysphonias are possible:

- hoarseness;

- laryngitis;
- inflammation of the larynx;
- hyperkinesis;
- mucous cyst or cyst shell;
- hypokinesis;
- a throat infection;
- glossoplegia;
- rhinopharyngitis.

These diseases have different degrees of severity, which significantly affects the quality of life. Thus, diagnosing fade disorders will help determine the level of quality of life. The scale of voice disorders is shown in Table 1.

Table I
SCALE OF VOICE DISORDERS

Gradation of violations	Degree of Violations	Description	Recommendations
No violations	0	-	-
Minor violations	1	Linguistic disorder is hardly felt or felt by the patient alone	Speech therapy is recommended
Moderate abnormalities	2	Decreased ease and speed of speaking	Speech therapy needed
Heavy Violations	3	The Talker needs the help of a listener. The patient often can not be understood, but understands himself	Speech therapy and help from the listener are necessary
Deep damage	4	Speaking with fragmentary expressions. The listener has to guess a lot. Information is small, and the listener	Need speech therapy and the study of sign language, consultation or synthesis of voice.

IV. Analysis of speech impairment

In a study conducted by Leinonen Et Al.[4], To replace direct listening to the voice, a scale of assessments of various degrees and forms of dysphonia was created. To compare the criteria, a neural network without a teacher was used, the training of which was conducted using a perceptual map of estimates of the normal and dysphonic voice [3]. The results of the experiment are shown in Fig. 2.

This approach has several disadvantages:

- Absence of comparative characteristics with the previous state of the voice. In view of this shortcoming, it is impossible to separate the congenital disorders from acquired ones;
- Lack of diagnosis of several disorders at the same time.

Classification of various forms and the degree of dysphonia can be made by using not only perceptual assessments of pathology, roughness, respiration, tension and asthenia, but also by comparing the estimates with the previous value, by

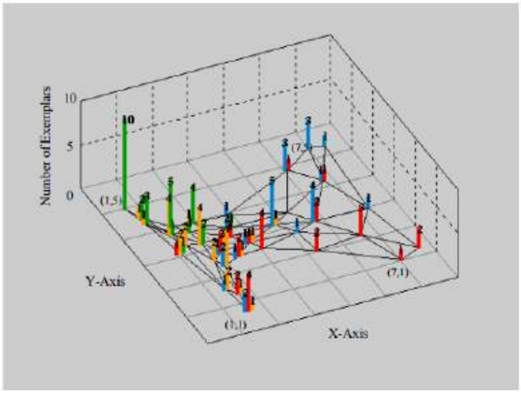

Figure 2. The results of determining the voice dysphonia (green - normal voice, yellow - hypotonic, red - hypertonic, blue - spasmodic)

including a voiceprint with a simulation of some degree of disease as input parameters. Thus, the voice print can be used not only for access control equipment, but also for assessing the speaker's voice deviations. The obvious advantage of this approach is the absence of direct contact with the speaker and the possibility of remote diagnostics, including the use of electronic means of communication. Also, this procedure has the possibility of full automation and undemanding resources.

V. Neural network comparison

The self-organizing map (SOM) (Kohonen, 1982) is one of the most important neural network architecture. Since its invention it has been applied to so many areas of Science and Engineering that it is virtually impossible to list all the applications available to date (van Hulle, 2010; Yin, 2008). In most of these applications, such as image compression (Amerijckx et al., 1998), time series prediction (Guillen et al., 2010; Lendasse et al., 2002), control systems (Cho et al., 2006; Barreto and Araújo, 2004), novelty detection (Frota et al., 2007), speech recognition and modeling (Gas et al., 2005), robotics (Barreto et al., 2003) and bioinformatics (Martin et al., 2008), the SOM is designed to be used by systems whose computational resources (e.g. memory space and CPU speed) are fully available. However, in applications where such resources are limited (e.g. embedded software systems, such as mobile phones), the SOM is rarely used, especially due to the cost of the best-matching unit (BMU) search (Sagheer et al., 2006). Essentially, the process of developing automatic speech recognition (ASR) systems is a challenging tasks due to many factors, such as variability of speaker accents, level of background noise, and large quantity of phonemes or words to deal with, voice coding and parameterization, among others. Concerning the development of ASR applications to mobile phones, to all the aforementioned problems, others are added, such as battery consumption requirements and low microphone quality. Despite those difficulties, with the significant growth of the information processing capacity of mobile phones, they are being used to perform tasks previously carried out only on personal computers. However, the standard user interface still limits their usability, since conventional keyboards are becoming smaller and smaller. A natural way to handle this new demand of embedded applications is through speech/voice commands. Since the neural phonetic typewriter (Kohonen, 1988), the SOM has been used in a standalone fashion for speech coding and recognition (see Kohonen, 2001, pp. 360-362). Hybrid architectures, such as SOM with MultiLayer Perceptrons (SOM-MLP) and SOM with Hidden Markov Models (SOM-HMM), have also been proposed (Gas et al., 2005; Somervuo, 2000). More specifically, studies involving speech recognition in mobile devices systems include those by Olsen et al. (2008); Alhonen et al. (2007) and Varga and Kiss (2008). It is worth noticing that Portuguese is the eighth, perhaps, the seventh most spoken language worldwide and the third among the Western countries, after English and Spanish. Despite that, few automatic speech recognition (ASR) systems, specially commercially available ones, have beendevelopedand it isavailable worldwidefor the Portugueselanguage. This scenario is particularly true for the Brazilian variant of the Portuguese language, due its large amount of accent variation within the country. Scanzio et al. (2010), for example, report experiments with a neural network based speech recognition system and include tests with the Brazilian Portuguese language. Their work is focused on a hardware-oriented implementation of the MLP network. In this context, the current paper addresses the application of self-organizing maps to the Brazilian Portuguese isolated spoken word recognition in embedded systems. For this purpose, we are particularly interested in evaluating several software strategies to speedup SOM computations in order to foster its use in real-time applications. The end-user application is a speaker-independent voice-driven software calculator which is embedded in smartphones.

We used learning without a teacher, because it is much more plausible model of learning in the biological system. Kohonen developed and many others, it does not need to output the target vector and therefore, does not require comparison with predetermined ideal responses, and learning set consists only of the input vectors. The training algorithm adjusts network weights so as to produce consistent output vectors, ie, to sufficiently close the presentation of input vectors produce the same outputs. The learning process, therefore, highlights the statistical properties of the training set and groups similar vectors in the classes. Presentation of the input vector of this class will give a certain output vector. The spread signal in such a network is as follows: input vector is normalized to 1.0 and applied to the input, which distributes it on through the matrix of weights W. Each neuron in layer Kohonen calculates the sum at its input and depending on the condition of the surrounding neurons becoming active layer or inactive (1.0 and 0.0). Neurons in this layer operate on the principle of competition, ie. E. As a result of a certain number of iterations is still an active one neuron or a small group. This mechanism is called lateral. Since testing of this mechanism requires significant computing resources, in my model it replaced by finding the maximum neuron activity and awarding him the activity 1.0, and 0.0 all other neurons. Thus, the neuron is activated for which the input vector closest to the vector of the weights. As a sigmoid activation function is used, which is as follows:

$$f(x) = 1/(1 + e^{-a*x}) \quad (1)$$

where a – slope parameter.
Geometrically, this rule shows next picture:

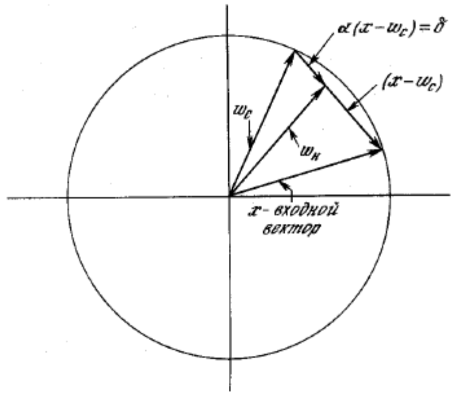

Figure 3. Correction weights of Kohonen neuron

Since the input vector x is normalized, ie. E. Is on a hypersphere of unit radius in the space of weights, then the correction weights on this rule is rotated vector weights toward the input that allows to produce statistical averaging of input vectors, which reacts active neuron. Thus, the study was replaced lateral approach leading to the activation of neurons.

VI. Conclusion

The result of this study was to modular application performing voice user authentication with analysis of user voice disorders. The program consists of three main parts. The first carries the addition of users, the second and the third carries the identification sends information to identify the user.

References

[1] GOST R 51241-2008. Sredstva i sistemy kontrolya i upravleniya dostupom. Klassifikatsiya. Obshchie tekhnicheskie trebovaniya. Metody ispytanii [State Standart R 51241-2008. Means and systems of control and access control. Classification. General technical requirements. Test methods]. Moscow, Standartinform, 2009. 10p.

[2] Raho, G.I. [et al.] Universal Developing of Persons Identification Based on RFID/ In Journal of Emerging Trends in Computing and Information Sciences, 2015. Vol. 6, No. 10. P. 592-597.

[3] Adeyemo, Z.K. Development of Hybrid Radio Frequency Identification and Biometric Security Attendance System. In International Journal of Applied Science and Technology, 2014. Vol. 4, No. 5. P. 190-197.

[4] Leinonen, L. Dysphonia detected by pattern recognition of spectral composition. In Journal of Speech, Language, and Hearing Research. April 1992, Vol. 35, 287-295. doi:10.1044/jshr.3502.287

[5] Biometrics, Computer Security Systems and Artificial Intelligence Applications / Ed. K. Saeed, J. Pejas, R. Mosdorf. Springer, 2006. 345 p.

[6] You, Y. Audio Coding: Theory and Applications. NY: Springer, 2010. 349 p.

[7] Herbig, T. [et al.] Self-Learning Speaker Identification: A System for Enhanced Speech Recognition. Berlin:Springer Verlag Gmbh, 2011.

[8] Al-Shayea Q.K., Speaker Identification: A Novel Fusion Samples Approach / Qeethara Kadhim Al-Shayea, Muzhir Shaban Al-Ani. In International Journal of Computer Science and Information Security (IJCSIS), 2016. Vol. 14, No 7. P.423-427.

[9] Ding Ing-Jr., Chih-Ta Yen, Yen-Ming Hsu. Developments of Machine Learning Schemes for Dynamic Time-Wrapping- Based Speech Recognition. In Mathematical Problems in Engineering, 2013. P. 5.

[10] Kohonen, T. Self-Organizing Maps. Berlin: Springer, 2001.

[11] Bosi, M., Goldberg, R.E. Introduction to digital audio coding and standards. Springer Science+Business, Media USA, 2010. 434 p.

[12] Menshakou, P.A. Voice user identification in access control systems / P.A. Menshakov, I.A. Murashko // Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems], 2016, pp. 411-414

[13] Sergienko, A. B. Cifrovaya obrabotka signalov. 2-e izdanie. SPb: Piter, 2006. 751 p.

[14] Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transforms. In Proceedings of the IEEE. Vol. 66, Jan 1978. P. 51-83.

[15] Azarov, I.S., Petrovskii, A.A. Primenenie mgnovennogo garmonicheskogo analiza dlya antropomorficheskoi obrabotki rechevih signalov. In Vestnik MGTU im. N.E. Baumana. Ser. “Priborostroenie”, 2012. p. 51.

[16] Using self-organizing maps to cluster music files based on lyrics and audio features. In Research Congress 2013 De La Salle University, Manila, March 7-9, 2013.

[17] Ghitza, O. Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition. In IEEE Transactions on Speech and Audio Processing, 1994. Vol. 2, No 1. P. 115–132.

[18] Azarov, I.S., Petrovskii, A.A. Primenenie mgnovennogo garmonicheskogo analiza dlya antropomorficheskoi obrabotki rechevih signalov. In Dokladi BGUIR, 2011. No 4. PP. 59–70.

[19] Mwasiagi, I. Self Organizing Maps – Applications and Novel Algorithm Design. In Josphat Igadwa Mwasiagi InTech. January 21, 2011. P. 91.

[20] Alejandro, C. Analysis of Kohonen's Neural Network with application to speech recognition. In Mexican International Conference on Artificial Intelligence. Mexico: Guanajuato, 2009. P. 8.

[21] Koikkalainen, P. Self-Organizing hierarchical feature maps/ P. Koikkalainen, E. Oja// Proceedings of 1990 International Joint Conference on Neural Networks vol. II, San Diego, 1990. pp. 279–284.

[22] Zagumenov, A. P. Komputernaya obrabotka zvuka. M.: DMK, 1999. 384 pp.

ИССЛЕДОВАНИЕ РЕЧЕВЫХ РАССТРОЙСТВ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

Меньшаков П. А.
Кафедра «Информационные технологии»,
Гомельский государственный технический университет имени П.О. Сухого
Гомель, Республика Беларусь

В настоящее время, голосовая идентификация и анализ голоса являются одним из передовых направлений, как в области систем защиты информации, так и в определении голосовых параметров [1]. Предлагается рассмотреть способ анализа состояния голоса и выявления его проблем, таких как усталость голосовых связок, повреждения или воспаления речевого тракта при помощи ранее записанного голоса диктора и нейросетевого анализа голосовых изменений.