UDC 004.777

SPEAKER IDENTIFICATION FOR SPEECH INFORMATION PROTECTION SYSTEMS

Shakin K.P.[1], Makarenya E.A.[2], Dai Junyi [2]

[1] Joint Institute of Informatics Problems of the National Academy of Sciences of Belarus
Minsk, Republic of Belarus
[2] Belarusian State University of Informatics and Radio Electronics
Minsk, Republic of Belarus

O.B. Zelmansky – PhD, associate professor

A structural diagram of the process of identifying an announcer using a neural network is considered. A method of identifying the voice of an announcer based on a neural network with a three-layer perceptron architecture is disclosed.

Today, one of the priority tasks for ensuring information security is to protect the speaker's speech. The solution of this problem is carried out using passive and active methods of information protection

In cases of protection of premises using active methods, white, pink and speech-like noise generators are used. In a number of experiments, it has been proven that the best masking effect is obtained using vibrations similar in spectral composition to the speech signal. Accordingly, in order to protect voice information, it is advantageous to give preference to speech-like interference protection systems.

In order to increase the efficiency of voice information protection, and thereby reduce the speaker speech intelligibility, it is proposed to use an announcer identification module for reconnaissance equipment using a neural network, which will allow generating speech-like interference in the future in accordance with the voice features of the announcer.

The essence of identifying a speech announcer is to select, classify, and then respond to human speech from the input audio stream [1].

Announcer identification is the process of identifying a person from a voice pattern by comparing a given sample with patterns stored in the announcer database.

All speech recognition systems can be divided into two classes: ·

1. narrator-dependent systems; ·
2. systems independent of the announcer.

Given that human speech is highly variable, high-efficiency real-time identification of announcers requires high-speed computing. One way to solve speech identification problems is to use neural networks.

In any speech recognition system, there is always a step of comparing the input signal with the available standards. Regardless of the presence or absence of pre-processing of the signal (highlighting the main features, converting to another form in the new parametric space, etc.), the signal is a vector in the established parametric space, which will later be compared with vectors stored in memory to determine its belonging to a certain class.

Main stages of speech signal classification:

1. Extracting features from the input speech signal.
2. Build the narrator model (template) based on the feature vectors obtained in the previous step [2].

The procedure for identifying the announcer taken into account in the system by the input voice signal in all methods consists in selecting the most suitable saved model based on any criteria.

At the moment, there are various methods for classifying the speech signal, which allow solving the problem of identifying the announcer by voice. The most common are:

- dynamic programming method;
- vector quantization method;
- method of Gaussian mixtures;
- support vector method;
- hidden Markov model.

It is proposed to identify the announcer according to the scheme indicated in the figure below (Figure 1).

On an entrance of the module of voice identification of the announcer the audio signal arrives record of which happens via the microphone connected to an entrance of the sound card of the computer or already earlier written down signal is used. The acoustic component will transform a signal to a digital form with the set sampling frequency parameters. Then cleaning of a signal from noise, removal of the sites which are not bearing information is carried out, normalization of a signal and its splitting into the fixed intervals in a time domain on whom characteristics will be defined is carried out. In the block of allocation of characteristic signs there is an allocation of characteristic signs of a signal by means of Gilbert's transformation - Huanga. On an entrance of the "neural network" block the result from the previous block gets, further actions depend on the choice of an operating mode.

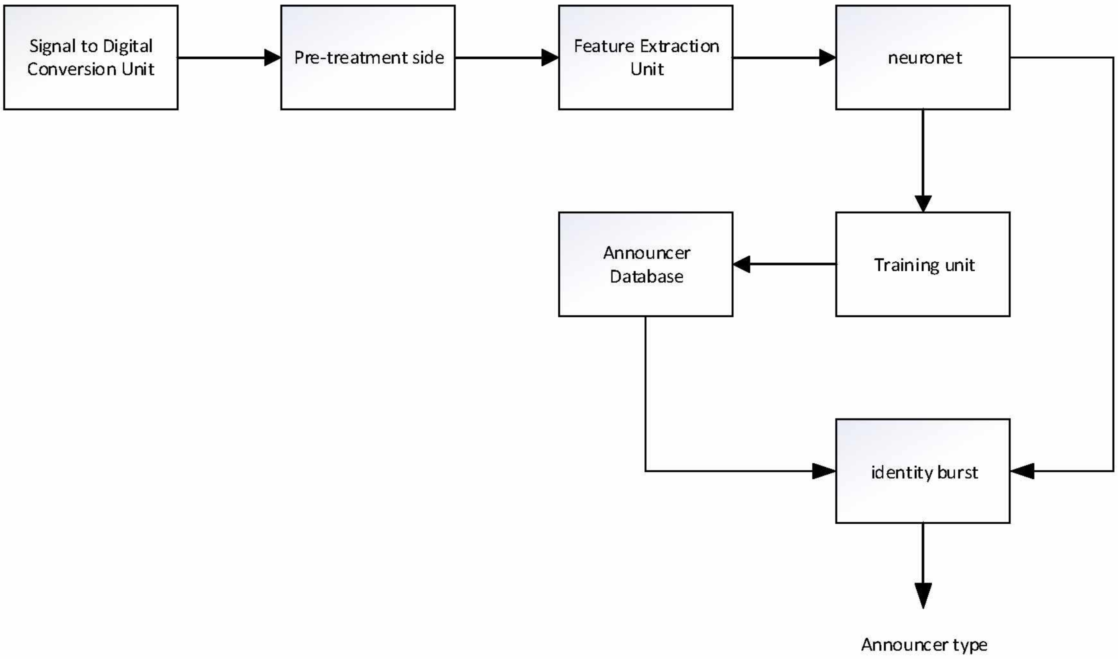

Figure 1. Block diagram of the narrator identification process

When choosing "training," data is transmitted to the training unit of the neural network, where it is trained and transmitted to the database. When choosing the "recognition" mode, the data goes to the identification unit, and data from the database unit with already stored data also comes to this unit. Then there is identification and classification (determination of belonging to a certain class), then the result of the announcer type is sent.

It is worth considering that before starting work, the user must add a database of announcers and train a neural network.

A neural network is a computational system or machine created to simulate analytical actions performed by a human brain and is a structure of neurons connected to each other and characterized by their internal properties, individual topology (architecture), as well as learning rules for obtaining the desired output signal.

Neural networks are models based on machine learning, that is, they acquire the necessary properties in the learning process, which consists in iteratively adjusting the network weights according to some rule called the learning algorithm.

One of the stages of the neural network functioning is training, during which data from the training set is alternately received at its input in order to adjust the weight coefficients of synaptic connections to obtain the most adequate signal at the output of the neural network. Training is carried out by sequential presentation of input vectors with simultaneous adjustment of weights in accordance with a certain procedure. During training, the weights of the network gradually become such that each input vector produces an output vector.

There are three learning algorithms: "with teacher," "without teacher" (self-study) and mixed. In the first case, the neural network has the correct responses (network outputs) to each input example. Weights are adjusted so that the network produces answers as close as possible to the known ones. Teaching without a teacher is associated with such a concept as a pattern - by processing significant amounts of data, the algorithm must first independently identify patterns. At the next stage, based on the identified patterns, the machine interprets and systematizes the data. In mixed learning, some of the weights are determined through teaching with a teacher, while the rest is obtained through self-learning.

To solve the classification problem, it is proposed to use a neural network with a three-layer perceptron architecture. Its advantages include a comparative simplicity of analysis and a fairly high classification efficiency. By using the continuous excitation function, such networks are capable of generalizing the training set.

In conclusion, it is worth noting that the proposed method of identifying an announcer using a neural network with a three-layer perceptron architecture will reduce the time and resources spent, which in turn will increase the efficiency of the voice protection system.

List of sources used:

1. Herbig T., Gerl F., Minker W. Self-Learning Speaker Identification: A System for Enhanced Speech Recognition. Berlin: Springer, 2011. 172 p

2. Speaker-by-voice-text identification [Electronic resource] – access mode: http://seminar.at.ispras.ru/wp-content/uploads/.

3. Koval, S. L. Comprehensive Voice and Speech Announcer Identification Methodology // S. L. Koval. Informatization and information security of law enforcement agencies: proceedings of the XX International Scientific Conference. Moscow.: Academy of Management of the Ministry of Internal Affairs of Russia, 2011. C. 364-370.