Applying intelligent systems of speech recognition for optimizing the algorithm of noise reduction in audio records

A V Ivanov, P S Fazluktinov, V A Kolesnev
Novosibirsk State Technical University, Novosibirsk, Russia

Abstract. The effectiveness of applying intelligent systems for speech recognition in speech intelligibility estimation is considered. The experiment with employing adaptive filtering to reduce ambient noise in audio records for further analysis of speech intelligibility with applying intelligent systems for speech recognition. Analysis of intelligibility with applying computer speech recognition is found to provide the intelligibility level close to a subjective method and differ not more than 6% averagely. The architecture of a Web service for automated noise reduction in audio records and automatised analysis of intelligibility for each result is designed. The recommendations on applying the obtained results to improve the speech security level and noise reduction are given.

1. Introduction

Modern methods to estimate speech intelligibility (such as Pokrovsky approach based on formants [1]) have been evolved for estimating the quality of communication lines. Since the information security field differs from communication theory, it has led to several drawbacks. A number of authors have been discussing the drawbacks over the last time [3-7].

One of the drawbacks is determined by the fact that the approach does not consider the opportunity of applying methods for noise reduction.

The most popular methods of noise reduction are adaptive filtering, bandpass filtering, spectral subtraction, pseudo-stereo. Considered as the most effective, adaptive filtering is widely used for noise reduction in audio records. However, this method provides with various intelligibility levels if different signal-to-noise ratios are used. To analyse speech intelligibility, objective and subjective methods are used. Subjective methods allow predicting final intelligibility more accurately since objective methods do not take into account features of the ear structure. To automatise applying of subjective methods for intelligibility analysis, systems of intelligent speech recognition such as Yandex Speech Kit, Dragon Mobile, Google Speech Recognition, Microsoft Speech API may be employed.

2. Adaptive filtering of audio records based on the noise patterns

Adaptive electric and acoustic echo cancellers as well as electric and acoustic channel equalizers have been used widely. Nowadays adaptive devices are essential in equipment of radioengineering systems, since applying adaptive processing affects technical characteristics of radioengineering systems.

An adaptive filter is the most important part of the devices. Adaptive filters change their parameters permanently. Terms of applying adaptive filters influence on determination of their parameters.

In general, adaptive noise reduction is used to reduce different noises and interferences in records that affect system operation. In most cases, noise influences on signals as their spectra may match to spectra of useful signals. In addition, in some cases, parametric filters cannot be used if the frequency band is not known.

In adaptive data processing, transmission function parameters are connected with input, output, expected, predicted and other additional signals as well as with parameters of statistical relationship. It allows self-tuning for optimal signal processing. In simple cases, adaptive devices contain a programmable adaptive data filter and an adaptive unit with an algorithm. The algorithm generates a control signal based on the input, output or additional data analysis. The flowchart of an adaptive filter is presented in fig. 1. The impulse response of adaptive systems can be finite or infinite.

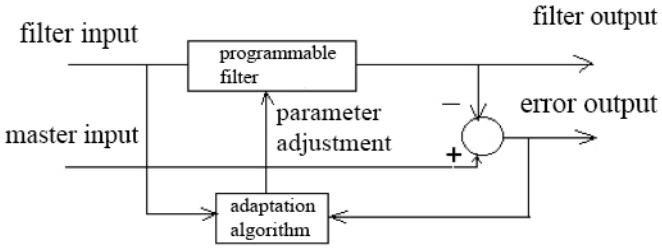

Figure 1. Flowchart of an adaptive filter

Adaptive devices are usually used for certain types of signals. The adaptive system structure and the adaptive algorithm are fully adjusted to the purpose and minimal input data. It leads to a variety of approaches to system development, significantly complicates their classification and development of general theoretical principles.

Adaptive filtering is based on suppressing the frequencies of an original signal according to the noise patterns [7]. First, noise frequency response called the noise profile is obtained. The noise profile is a set of A-weighted signal levels at each frequency.

The original signal is processed by bandpass filters at the frequencies corresponding the noise profile. The signal depression level (filter gain) is directly proportional to the A-weighted noise level at the corresponding frequencies.

The main principle of adaptive system difference from others is changing and self-tuning over time. Adaptive filter enhances and reduces the effect depending on the A-weighted noise level in a record. Self-tuning is important otherwise a producer must take into account all the possible conditions when developing a not adaptive system. The developed system must work in any conditions.

A producer also considers operation to be estimated by a certain criterion such as the average number of errors.

However, in practice, operation conditions are unknown and also may change.

In this case, adaptive systems adjusts to the conditions to find a better operation mode. This is a definite advantage of such systems.

The properties of adaptive systems are connected with external conditions and change over time as input data is changing.

Adaptive systems change the input signal, i.e. the transmission function should be adapted so the real signal is able to be transmitted through the system without interference and signal distortion, and the noises to have been reduced.

Systems can estimate the statistical parameters of a signal and adapt the transmission function to minimise some of the objective functions. This function is usually defined by a reference signal. The reference signal can be considered as the desired signal at the filter output.

A primary signal is applied to the input during adaptive filtering.

A primary signal contains all the necessary information including noise. In this regard, a second signal must be applied to the adaptive system. The second signal is an independent noise pattern. If the signal is applied to the filter input, the filter generates impulse response. The second signal is subtracted from the first one. As a result, the needed signal is obtained. If the noise signal cannot be obtained, the noise pattern extracted from the pauses between speaker's phrases is used as a reference

signal [10]. To analyse the pauses, the filter parameters are being changed according to the spectrum of the found noises.

.

To carry out the experiment, a free command line utility "SoX" is used. This utility allows generating the noise profile from a file and applying bandpass filtering according to the profile. The filtration is carried out by several bandpass filters with high Q-factor.

An example of an adaptive filter is shown in Fig. 2.

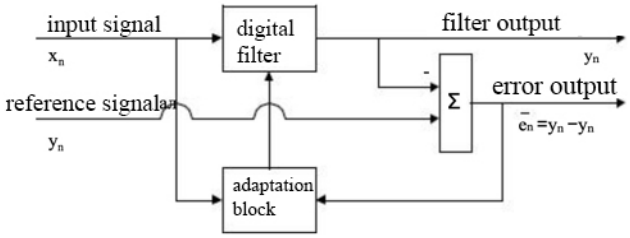

Figure 2. Implementation of an adaptive filter

An adaptive filter contains the adaptive algorithm to change parameters and ratios and a unit with separated digital filters. Two signals are applied to the input simultaneously. One of the signals is a signal-plus-noise mixture, and the second signal is used for creating a noise model.

3. The experiment design with participation of experts

To estimate speech intelligibility in audio records, a group of experts is used. Speech intelligibility is a ratio between the number of correctly recognised words to the total number of words in a text.

Speakers have read texts aloud, and their voices have been recorded with a microphone in a non-empty room. Several records of different voices are obtained.

To obtain a spectrogram, short-time Fourier transformation is used.

Spectrum is calculated based on the short-time signals (fig. 3). Each of the obtained spectra stands for the bar in a spectrogram.

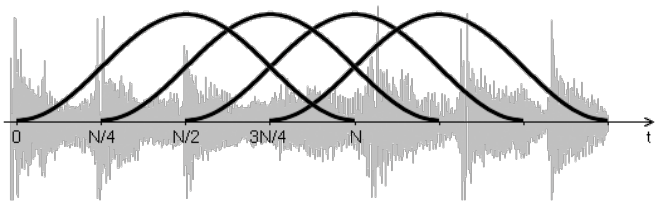

Figure 3. Short-time Fourier transformation

Time is plotted on the horizontal axis, and frequency is on the vertical. The amplitude is shown with brightness (or color). The spectrogram of a guitar note (fig. 4) shows the developing sound. High harmonics is observed to have a smaller amplitude and attenuates faster than the lower ones. In addition, noise in a record is observed in the spectrogram (navy blue color).

The dependence on the signal level (dB) is shown to the right from the frequency.

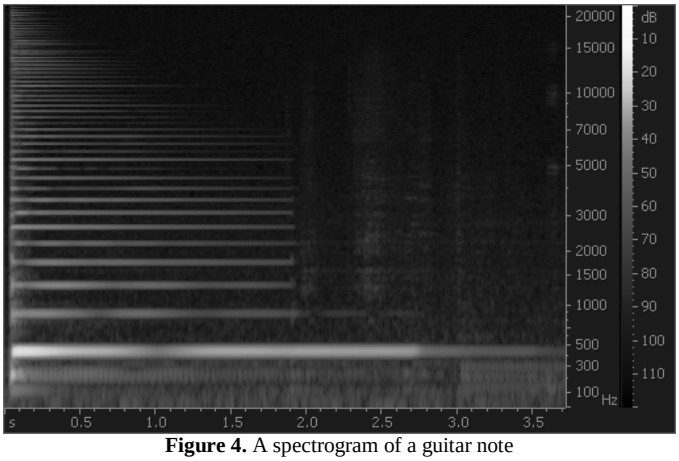

Figure 4. A spectrogram of a guitar note

Spectrum analysis can be applied for signals changing over time (such as a vibrato vocal technique).

A spectrogram of vibrato vocal is shown in fig. 5. Based on the spectrogram, the frequency, depth, shape and evenness of vibrato can be estimated.

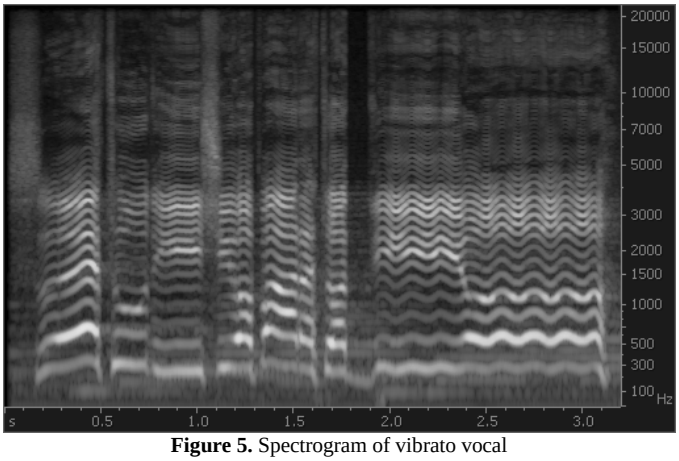

Figure 5. Spectrogram of vibrato vocal

To conduct the experiment, ambient noises have been recorded by the BC501 microphone connected with the ZET 110 noise level meter. The sampling rate is 50 kHz.

City street, factory, and appliance noises have been used as main noise. Their spectra are shown in fig. 6-8.

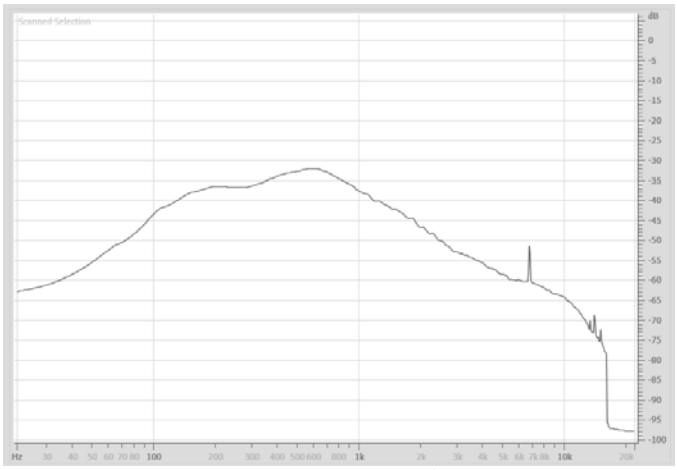

Figure 6. Frequency response of city noise

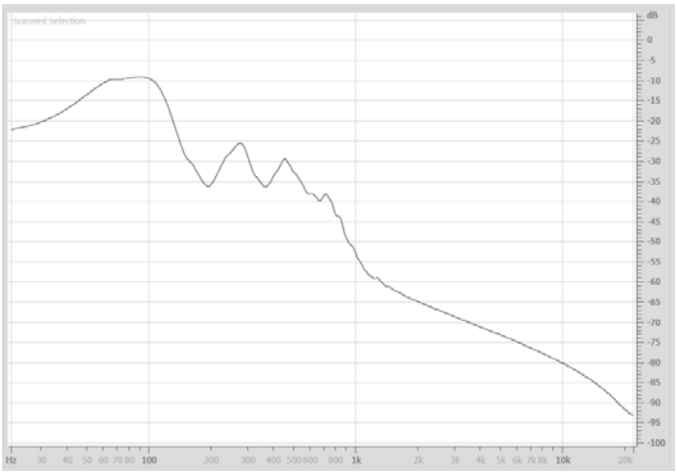

Figure 7. Frequency response of factory noise

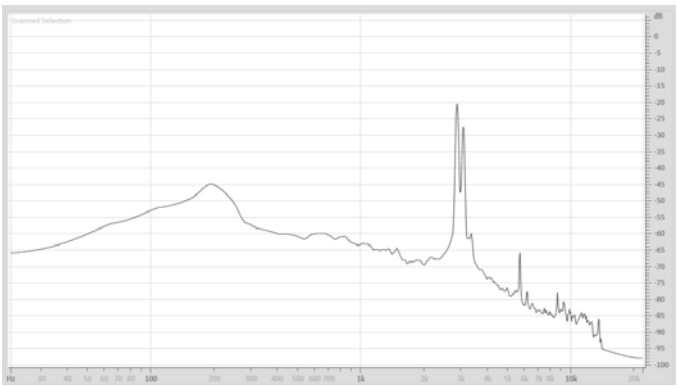

Figure 8. Frequency response of appliances

The noises are mixed with speaker's records. In this case, an audio record is a WAV-file with the sampling rate of 48 kHz and length of 2 minutes. Female and male voices have been used for the experiment. To add the noises, signal-to-noise ratios equal to -30 dB, -20 dB, -10 dB, 0 dB, 5 dB, 10 dB, 20 dB, 30dB are used.

An example of the noised voice spectrum is shown in fig. 9.

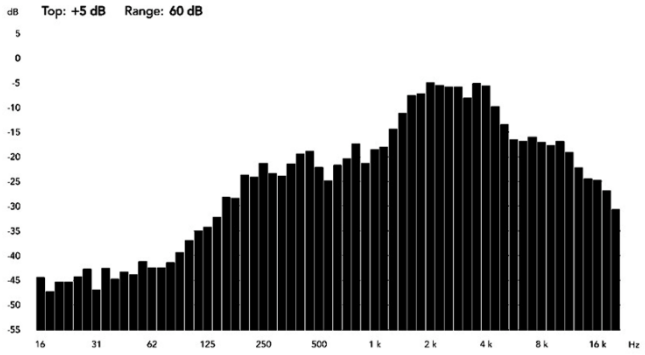

Figure 9. Voice spectrum after appliance noise is added

After the noises being added, a subjective method is applied to estimate intelligibility of audio records by experts.

Experts listen to the audio records and write down what they have heard. The number of correctly recognised words is counted and intelligibility is calculated (1):

$$W = \frac{k}{n}, \quad (1)$$

where k stands for the number of correctly recognised words, n – the total number of words in a text read by a speaker.

Noise in audio records is reduced applying the method of adaptive noise reduction for each noise. The resulted audio records are estimated by experts again and speech intelligibility is calculated (1).

Based on the results, the speech intelligibility dependence on a signal-to-noise ratio is obtained (fig. 10-12). The intelligibility increase dependence dW=W2-W1 on the signal-to-noise ratio q is shown in graphs (W2 – intelligibility after the noise reduction, W1 – intelligibility before the noise reduction).

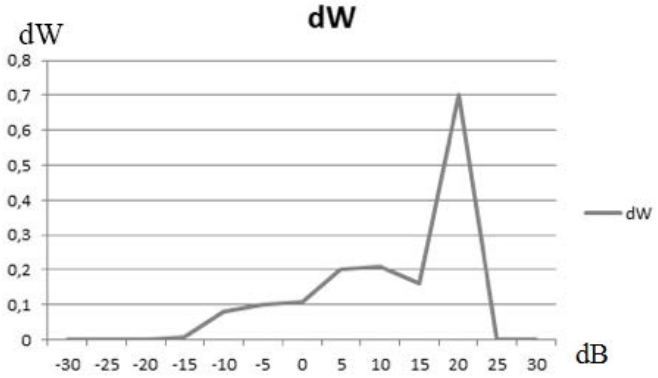

Figure 10. Speech intelligibility increase dependence on signal-to-noise ratios after city noise is reduced in audio records

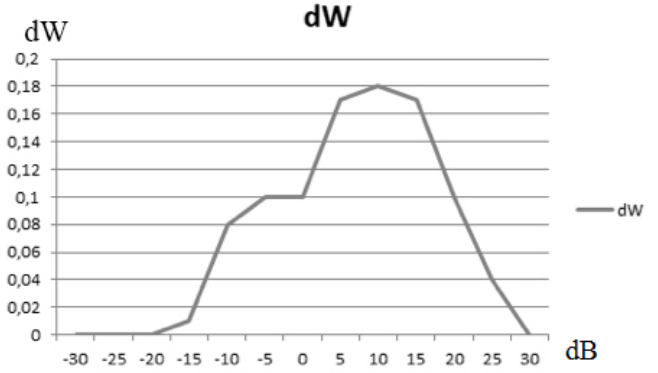

Figure 11. Speech intelligibility increase dependence on signal-to-noise ratios after factory noise is reduced in audio records

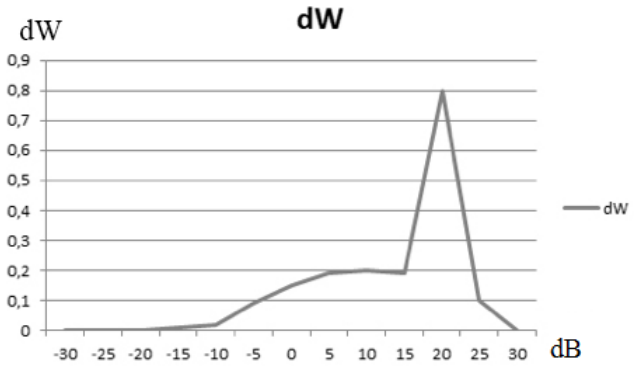

Figure 12. Speech intelligibility increase dependence on signal-to-noise ratios after appliance noise is reduced audio records

The conducted experiments allow concluding that adaptive filtering is the most effective for the approach to speech intelligibility estimation if the signal-to-noise ratio ranges from 15 dB up to 20 dB. The reduction of appliance noise is the most accurately.

4. Applying the intelligent systems for speech recognition in intelligibility analysis

The process of subjective speech recognition by experts does not allow noise reduction quality estimation to be increased and automated in audio records since the human is a key to estimate in this case [6].

To automatise speech intelligibility estimation, systems for speech recognition can be employed to convert audio information into text [9].

To analyse speech, the Yandex SpeechKit Web service has been chosen. The service is a free API for uploading files to the server and getting a response with recognised text via JSON [9]. To get a response from the service, a POST-request must be sent to the intelligent system's URL.

To send HTTP-requests, the free Postman software is used. The flowchart of the system in shown in fig. 13. An example of a request to send is shown fig. 16.

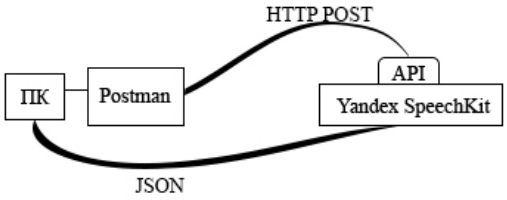

Figure 13. Flowchart of the system

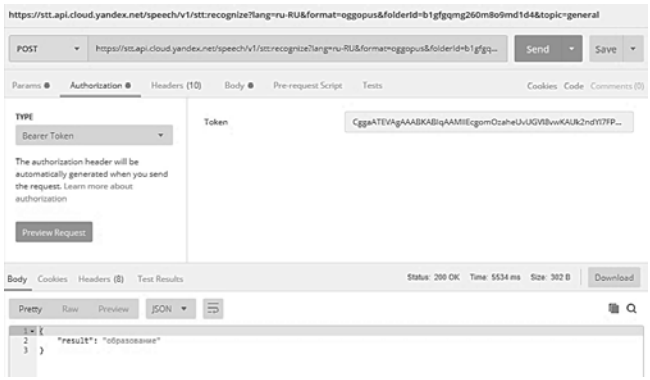

Figure 14. Example of sending a request to the Yandex SpeechKit Web service with using Postman

To estimate the operation quality of the intelligent system, the number of recognised words has been counted and related to the total number of words (1).

The number of recognised by AI words in audio records is equal to the total number of words in a text for city and factory noise with the signal-to-noise ratio ranged from -10 dB up to 30 dB, and for appliance noise with the ratio ranged between -5 dB and 30 dB.

It may be concluded that the ratio between the number of recognised words and the total number of words can be applied as speech intelligibility estimated by artificial intelligence. The total number of vowels in a text can be applied instead of the word number.

The dependence of intelligibility increase on the signal-to-noise ratio is shown in fig. 15-17.

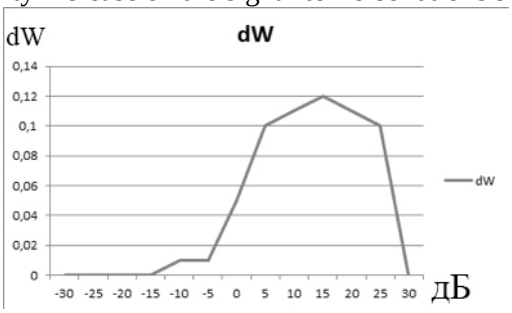

Figure 15. Speech intelligibility increase dependence on signal-to-noise ratios after city noise is reduced from audio records

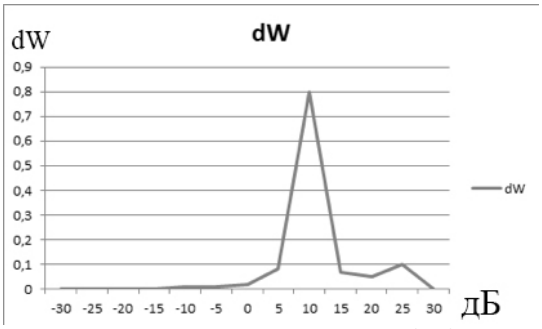

Figure 16. Speech intelligibility increase dependence on signal-to-noise ratios after factory noise is reduced from audio records

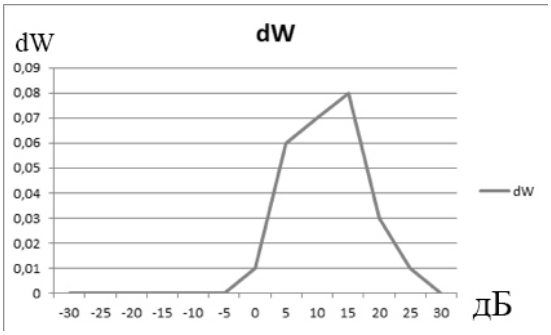

Figure 17. Speech intelligibility increase dependence on signal-to-noise ratios after appliance noise is reduced from audio records

The experiment allows concluding that Yandex SpeechKit is effective for speech intelligibility estimation if the signal-to-noise ratio ranges between 10 and 20 dB.

5. Designing the architecture of the web service for automatic noise reduction with the best intelligibility

The conducted research has shown that intelligent systems for speech recognition allows estimating speech intelligibility by counting the number of recognised words in a noise reduced audio record.

Therefore, it is possible to design a Web service for automatised ambient noise reduction in audio records.

Noises uploaded to the Web service database are used as noise patterns for adaptive filtering. The main goal of the Web service is to reduce noise one-at-a-time in uploaded audio records applying adaptive filtering for each noise from the database. After each filtering step, the reduced audio record is sent to the speech recognition service. Once a response is received, the main service counts the number of the words. The final audio record with the greatest number of recognised words is selected (fig. 18).

The audio record, characteristics of the filter as well as the intelligibility level is sent to a user. A user can make a conclusion based on the obtained results and apply own filters to the original audio record or to use the reduced record provided by the service.

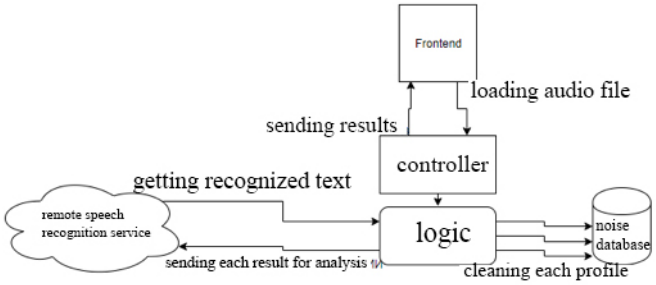

Figure 18. Architecture of the Web service

6. Conclusion

The experiments on applying adaptive filtering technique have been conducted to reduce ambient noises in audio records and to analyse speech intelligibility applying the subjective methods for speech intelligibility estimation and the intelligent system for speech recognition.

The effectiveness of applying intelligent systems for speech recognition to estimate speech intelligibility is considered for the problems of noise reduction.

Intelligibility analysis with applying computer speech recognition is found to provide with better results of intelligibility close to subjective methods and differ not more than 6% averagely.

The architecture of the Web service for automatic noise reduction in audio records and automatised analysis of intelligibility is designed. The recommendations are given to increase the quality of speech protection and noise reduction.

References

[1] Zheleznyak V K, Makarov Yu K and Khorev A A 2000 Some methodological approaches to evaluating the effectiveness of information security speech *Special Equipment* **4** pp 39–45

[2] Pokrovsky N B 1962 *Calculation and measurement of speech intelligibility* (Moscow: Svyazizdat) 390

[3] Avdeev V B 2013 About some directions of improvement of the methodical approaches applied at an assessment of efficiency of technical information security *Special Equipment* **2** pp 1–10

[4] Didkovskii V S, Didkovskii M L and Prodeus A N 2008 *Acoustic speech communication channel expertise* (Kiev: Imeks Publ.) 420

[5] Gerasimenko V G, Lavruhin Yu N and Tupota V I 2008 *Methods of protection of the acoustic speech information leakage via technical channels* (Moscow: RCIB «Fakel») 258

[6] Ivanov A V, Reva I L and Truschin V A 2010 *Proc. X Int. Sci. Tech. Conf. on Actual Problems of Electronics Instrument Engineering (Novosibirsk)* pp 133–136

[7] Khorev A A 2010 Monitoring the effectiveness of the protection of the premises allocated by the leakage of voice information through technical channels *Information Security. Inside* **1** pp 34–45

[8] Steeneken H J M, Houtgast T 1985 RASTI: A Tool for Evaluating Auditoria *Bruel & Kjaer Technical Review* **3** pp 13–39

[9] Cloud.yandex.ru (2019) Documents on Yandex SpeechKit [Online]. Available at: https://cloud.yandex.ru/docs/speechkit/ [Accessed 31 May 2019]
[10] Ivanov A V, Reva I L and Fazluktinov P S 2017 Application of sound cleaning methods for filtration of speech information *Dynamics of Systems, Mechanisms and Machines vol 5* **4** pp 70-73