ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

METHOD FOR SPEECH INTELLIGIBILITY ASSESSMENT WITH COMBINED MASKING SIGNALS

[1] YERZHAN N. SEITKULOV, [1] SEILKHAN N. BORANBAYEV [1] BANU B. YERGALIYEVA
[2]HENADZI V. DAVYDAU, [2] ALEKSANDR V. PATAPOVICH

[1] L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan

[2] Belarusian state university of informatics and radioelectronics, Minsk, Belarus

E-mail: [1] Seitkulov_y@enu.kz, [2] nil53@bsuir.edu.by

ABSTRACT

The article is devoted to the method for speech intelligibility assessment while protecting it from leakage through acoustic channels by masking it with combined acoustic signals, including white noise and speech-like signals. Difficulties in solving the tasks of voice information protection, as well as the tasks of information protection in general, are caused by uncertainties associated with difficulties in the mathematical formulation of protection problems on the one hand and a large number of factors affecting on the speech information security on the other hand. The method for speech intelligibility assessment is proposed for estimation the speech information security when it used in voice information protection systems according to limit states. For a correct assessment of the speech information security by its intelligibility indicators, it is necessary to make a number of assumptions and limitations that can be adopted on the basis of experience in the practical implementation of the speech information protection by known technical means and a set of organizational measures.

Keywords: *Speech Intelligibility; Combined Masking Signals; Security Of Voice Information; "White" Noise; Speech-Like Signals.*

1. INTRODUCTION

Historically, the development of methods for the speech intelligibility assessment in a noise environment was caused by the need to control the good conditions of intelligibility provision when transmitting information over communication lines, and primarily for use in aviation. The development of the methods went in two directions. The first approach was based on the formant structure of the speech signal, i.e. the concentration of the speech signal energy for certain formants in a number of areas of the speech frequency range. A formative method for speech intelligibility assessment was developed for the Russian language and was primarily aimed at ensuring the quality of speech transmission over communication channels [1–3].

The essence of the formant method for speech intelligibility assessment is to find the sum for the entire frequency range of speech of the product of the probability of the location of formants in a given frequency band by the perception coefficient of formants in this frequency band for a given level of external noise. Such summation over the entire frequency range of speech is possible if the signal and the masking noise at these frequencies are independent. Formant speech intelligibility is calculated from the expression

$$A = \sum_{k=1}^{k} p_k \cdot w(E_k), \tag{1}$$

where k – frequency band number for which formant speech intelligibility is calculated; p_k – the probability of the formants location in the k-th frequency band; $w(E_k)$ – speech perception coefficient as a function of the signal-masking noise ratio in the k-th frequency band; E_k – the ratio of the level of the speech signal in the k-th band to the level of the masking signal in this frequency band.

The probability of finding formants in a band is calculated by the distribution function of formants in the speech frequency range and is determined from the expression [1]

$$p_k = F(f_{hk}) - F(f_{lk}), \tag{2}$$

where $F(f_{hk})$ и $F(f_{lk})$ – frequency distribution function of formants at the highest frequency of the k-th frequency band and, accordingly, the lower frequency of the same frequency band.

The dependences of the speech perception on the signal-to-noise ratio and the dependence of syllabic intelligibility on the integral level of articulation necessary for speech intelligibility assessment by the formant method were obtained by experimental studies in [1].

The second approach is also to some extent a formant method for the speech intelligibility assessment using the articulation index and was developed for the English language [5–7].

The articulation method for speech intelligibility assessment by the articulation index was developed at Bell's laboratory to ensure the quality of communication in aviation technology and was focused on the English language. Both formant and articulatory methods for the speech intelligibility assessment were developed for areas of speech intelligibility above 50 % and only subsequently with significant improvements, they found application for areas of speech intelligibility of several percent in solving problems of speech information protection.

The articulation method for speech intelligibility assessment is based on the calculation of the articulation index for a given frequency range.

In the works [6, 7] it's proposed to evaluate the speech information security using indicators of intelligibility, audibility and cadence (rhythm). It's proposed that speech intelligibility be calculated through the SNR or SPI indicator using signal-to-noise ratios for 16 third of the octave frequency bands [6–10]. SPI is proposed to be determined from the expression

$$SPI=\sum_{f=160}^{5000}\left[L_{ts}(f)-L_{n}(f)\right]/16, \qquad (3)$$

where the sum is for each of 1/3 octave bands with an average frequency f; $L_{ts}(f)$ – transmitted speech level to the position of the offender; $L_n(f)$ – level of external noises at the intruder's position.

The number in the square brackets should be limited so that it cannot have values less than –32 dB in one of the frequency bands. If the signal-to-noise ratio in a particular band is less than –32 dB, then this value is significantly lower than the auditory threshold and such (extremely low) values will inappropriately exaggerate the speech confidentiality degree. Therefore, it's necessary to trim or limit the difference in signal-to-noise levels in each 1/3 octave frequency band to a value of at least –32 dB.

In this case, the transition from the values of the SPI parameter to the indicator of speech intelligibility is performed using the dependence presented in graphical form [6] in Figure 1.

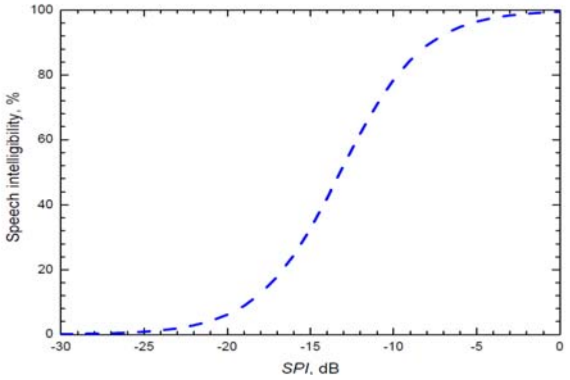

Figure 1. Speech intelligibility dependence from SPI articulation index.

However, this dependence is characteristic of English speech and the application of this dependence for other languages, including the Kazakh language, with its phonetic specificity, is very problematic.

At the same time, it should be noted that all the methods for speech intelligibility assessment basically contain experimentally obtained dependences of speech intelligibility on the index of articulation or on some other parameter. And this other parameter or articulation index, one way or another, is associated with the signal-to-noise ratio in individual frequency bands (octave, 1/3 octave or 20 equal intelligibility bands) or in the entire frequency range of the speech signal. In publications, these dependences are usually given for the integral sound pressure level, i.e. for the entire frequency range. Known methods for the intelligibility assessment of the speech masked by noise suggest that if in one or more octave bands the signal-to-noise ratio is greater than the average over the entire frequency range, then in other bands it should be as much smaller than the average over the entire frequency range, i.e. we apply the principle of additivity.

(3)

If in the initial period of development of methods for the speech intelligibility assessment it was proposed to divide the speech frequency range into 20 equal intelligibility bands, then in the future when determining the articulation index, which was determined by the results of experimental studies, there were difficulties with the lack of hardware to determine the signal–noise in these bands frequencies. For these purposes, began to use acoustic sound level meters with octave or 1/3 octave frequency bands. Moreover, in [1–3] for the Russian language it was proposed to use the recalculated values of the weighting coefficients for octave frequency bands. If for 20 bands of equal speech intelligibility, the weight coefficient was 0.05, then for octave frequency bands with geometric mean frequencies of 250, 500, 1000, 2000 and 4000 Hz, it was 0.03, 0.12, 0.20, 0.30, respectively and 0.26. Bands of equal intelligibility were determined from the condition of dividing the entire frequency range of the speech signal into bands in which the probability of the appearance of speech formants was the same. At the same time, in recent works [7–9] on intelligibility of English speech, weights are not used for 1/3 octave frequency bands.

The security of speech information assessment by the parameter speech intelligibility for languages other than Russian and English using the methods considered can introduce a system error due to differences in the spectra of speech for different languages. In addition, different distributions of phonemes by frequency in languages will also have intelligibility. Differences in the spectra of speech were studied in [10] for 12 languages and significant differences were shown both at low and high frequencies.

Research and comparison of formant properties of Ukrainian and Russian speech was performed in [11, 12]. It was found that with small signal-to-noise ratios and high levels of sound pressure of the speech signal, intelligibility of Russian and Ukrainian speech is almost the same, but with large signal-to-noise ratios, intelligibility of Ukrainian speech is noticeably lower. In this case, the formant method for the speech intelligibility assessment was used. The same intelligibility for Ukrainian and Russian speech with small signal-to-noise ratios is due to the influence of the fact that the speech apparatus of the speakers was formed in the conditions of bilingualism and they equally easily knew each language.

If Ukrainian and Russian speech are close in phonetic structure and for them intelligibility indicators are close, then for the Kazakh language, when assessing speech intelligibility in noise conditions, it is necessary to take into account the phonetic features of the Kazakh language. In addition, the existing methods for the speech intelligibility assessment, discussed above and used in voice information protection systems, are focused on a masking signal – it is “white” or “pink” or another type of noise. The use of combined masking signals in modern speech information protection systems imposes its own characteristics on the assessment of speech intelligibility as an indicator of the protection of speech information [13–19], which has been reflected in recent years in publications.

The aim of the work is to analyze well-known methods for the speech intelligibility assessment and applying these methods to assess speech intelligibility in the Kazakh language, taking into account masking by combined signals. It should be noted that no studies of speech intelligibility when masking it with combined signals have been found in the literature. In this paper, we try to fill this gap in the research of speech intelligibility for the Kazakh language using the hypothesis that speech intelligibility when masking it with combined signals is significantly lower than when applying "white" or "pink" noise with the same signal-to-noise ratio. This can be explained by the fact that in evaluating intelligibility of speech being masked with combined signals, the auditor involuntarily has a psychological need to recognize more powerful speech-like signals and only then recognize weaker masked speech signals. Due to the fact that the use of the articulatory method for the Kazakh speech intelligibility assessment requires a dependence of intelligibility on the index of articulation for a given language, it is necessary to conduct experimental studies to obtain this dependence for the Kazakh language in order to create a finished method. Research should be carried out with combined masking signals and the necessary measures for the preparation of textual material, as well as the selection and training of broadcasters and auditors.

2. COMBINED MASKING SIGNALS

The considered methods for the speech intelligibility assessment when it is protected by masking signals are applicable when separately “white” or “pink” noise or speech-like interference act as masking signals. This is due to the fact that only for masking interference in the form of “white”, “pink” noise or speech-like interference, experimental dependences of speech intelligibility on articulation index were obtained. There are no methods for the speech intelligibility assessment when masking it with combined signals. To eliminate this gap, this work was done.

According to the structural composition, combined masking signals designed to protect speech information from leakage through technical channels usually contain a noise component in the form of “white” noise and speech-like signals formed on the basis of structural units of speech taking into account the probability distribution of their appearance in this language [15, 17–19].

Quite often, it is recommended to use “pink” noise as a noise component, whose spectral density decreases with increasing frequency according to f_0 / f_c, where f_0 – value of low frequency of the noise; f_c – current frequency value [20, 21].

The dependencies of verbal intelligibility presented in a number of publications [22] for various types of masking signals were obtained, as a rule, by recalculating the signal-to-noise ratio in octave or 1/3 octave frequency bands for a given masking signal and calculating the obtained signal-to-noise ratios of the articulation index. Moreover, verbal intelligibility is determined by the same dependence of speech intelligibility on the articulation index, which was obtained experimentally for white noise in [1], published in 1962, and is further approximated by an analytical expression in [3].

The speech-like interference in [21, 22] and the Baron device is formed by randomly playing sonograms of the speakers' speech. At the same time, there are sections of a coherent text and auditors evaluating intelligibility do not psychologically cause great difficulties in restoring sections of speech signals that act as masking signals. Due to the fact that the auditor does not experience psychological stress in recognizing such masking speech-like signals, all attention can be focused on the recognition of the information signal and, as a result, low levels of security of speech information with this method of masking speech signals.

An important requirement in the formation of masking signals is the requirements of their random nature. For these purposes, it is recommended to use “white” noise generated due to thermal noise of semiconductor devices or other nature of physical noise. This requirement is due to the need to exclude any possibility of cleaning the noise of intercepted acoustic signals. The use of noise generated by digital methods raises concerns about the possibility of cleaning phonograms from such masking signals.

In [13–19], a different approach was proposed for the formation of speech-like signals not from separate sections of a connected text, but from phonetic structural units of speech, for example, allophones or polyphons. The base of phonetic structural units of speech for the synthesis of speech-like signals in the Kazakh language consists of 263 allophones and 80 structural units in the form of suffixes and endings [17]. For the Belarusian language, the phonetic base consists of 432 allophones and 44 sounds characteristic of the Belarusian language and consonants with a soft sign [18]. The base of the phonetic structural units of the Chinese language for the synthesis of speech-like signals was created by analyzing the dictionaries of the modern Chinese language. Based on the analysis results, a database of 406 phonemes of the same transcription in English was compiled, and taking into account the tonality of the language, the phoneme database was 1239. For the Russian language, the minimum practically used base should have at least 256 allophones [23]. For practical implementation, a base of 336 allophones was used [24].

Speech-like signals intended for masking speech information are similar in their formal properties to continuous speech, however, there are temporary areas where the speech-like signal is absent (as well as in natural speech), so these gaps must be filled with a noise signal so that there are no gaps and The information signal has got to empty temporary sections not filled with noise.

This approach to the formation of speech masking combined signals provides higher levels of security of speech information, since the isolation and processing of any signal is all the more difficult, the closer the interference (combined masking signals, including speech-like signals) in shape and frequency to the protected signal [25]. Therefore, one of the promising options for the formation of masking speech-like signals is their formation on the basis of the structural units of the speech of the speakers, whose speech signals require an increased degree of security. Moreover, the formants of the protected speech signals and the formants of the masking speech-like signals will be difficult to distinguish.

This is due to the fact that the formants of masking and speech signals will appear with a high probability at the same frequencies. In this case, for the same phonemes of speech-like signals and speech masked signals, their frequencies will match. In addition, the pitch frequencies for both speech-like signals and masked speech signals will be identical, since they belong to the same speaker.

For a clearer understanding of the description of combined masking signals figure 2 shows the spectra of white noise, combined masking signals, and the spectrum of the information signal masked by combined signals.

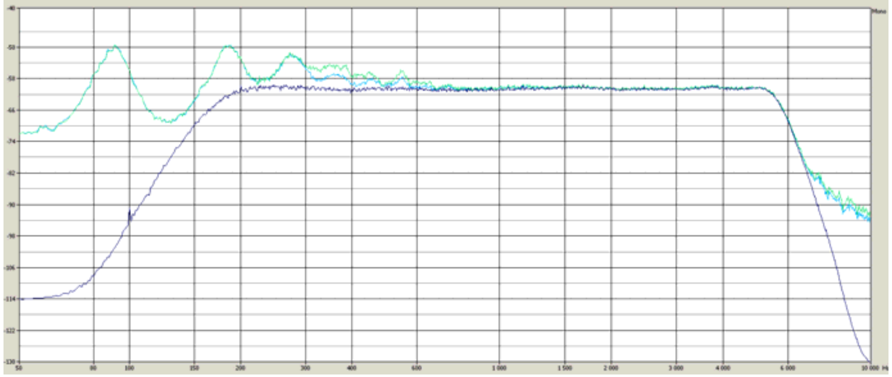

Figure 2. Spectra of the signals: 1 – "white" noise; 2 – combined masking signals, including "white" noise and speech-like signals; 3 – a spectrum of an information signal masked by combined signals.

Table 1 shows the distribution of the combined masking signals levels with speech-like signals in the Kazakh language in octave bands.

Table 1 – Distribution of the masking signals levels with speech-like signals in the Kazakh language in octave bands

The geometric mean frequency of the octave band, Hz	Sound pressure levels in octave frequency bands, dB			
	For "white" noise	For speech-like signals	For "white" noise and speech-like signals	For informational speech acoustic signal
125	50,6	55	56,4	47
250	53,6	61	61,7	53
500	56,6	57	59,8	49
1000	59,6	53,5	60,6	45,5
2000	62,6	50	62,8	42
4000	65,6	47	65,7	39
8000	68,6	44	68,6	36
In the band 89–11280	70	64	72,3	56

In a graphical form, these distributions are presented in Figure 3. The ordinate axis represents sound pressure levels in octave bands with the notation: 1 – for the "white" noise component; 2 – for the components of the speech-like signal; 3 – for the combined masking signal; 4 – for the protected speech signal.

Speech -like signals can be formed in the form of a monologue by a single speaker, or in the form of a dialogue between participants in negotiations. In this case, the ratio of the speech-like signal to the masking "white" noise should be – 6 dB, which allows to ensure that the level of consonant speech-like signals exceeds the vowels of the speech information signal [24].

A functional diagram of the formation of combined masking signals is shown in Figure 4.

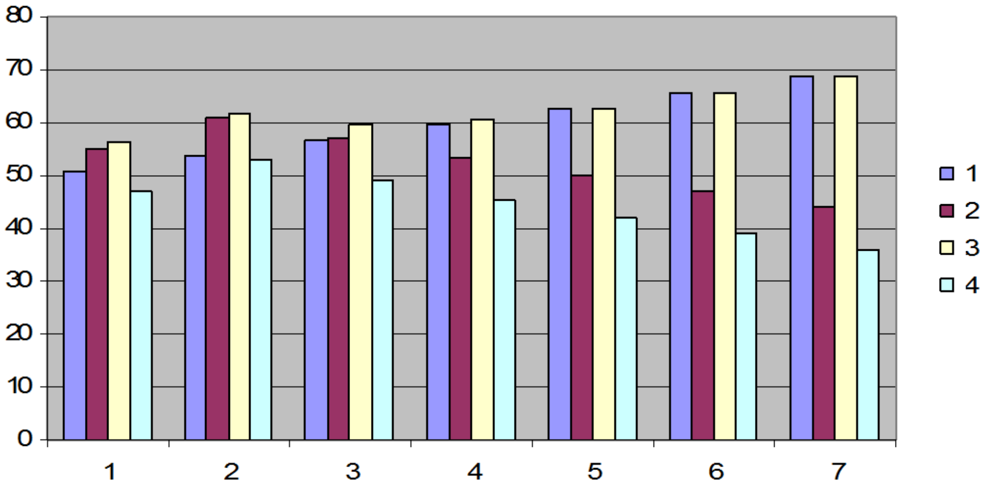

Figure 3. Distribution of the combined masking signals levels in octave frequency bands from 125 to 8000 Hz for the Kazakh language.

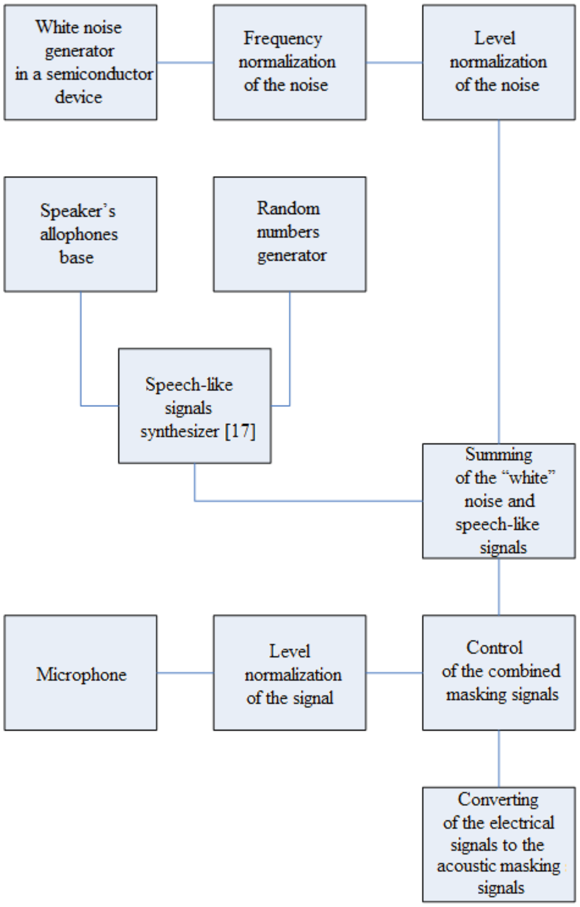
White noise generator in a semiconductor device
Frequency normalization of the noise
Level normalization of the noise
Speaker's allophones base
Random numbers generator
Speech-like signals synthesizer [17]
Summing of the "white" noise and speech-like signals
Microphone
Level normalization of the signal
Control of the combined masking signals
Converting of the electrical signals to the acoustic masking signals

Figure 4. Functional diagram of the combined masking signals formation

To generate “white” noise, it’s necessary to use physical noise sources, for example, on semiconductor devices, which makes it impossible to clear the speech information signal from masking noise. This noise must be normalized both in frequency and level before the operation of summing with speech-like signals. Speech-like signals are formed according to the algorithm described in [17], based on the speaker allophones base and using a random number generator. The level control of the combined masking signals is carried out in accordance with the sound pressure level of the protected speech information. For this, a microphone is used, the signal from which is normalized by level and then goes to the level control device of the combined masking signals. Further, the electric combined signals are converted into acoustic, masking speech signals.

It is analytically difficult to evaluate the intelligibility of speech masked by combined signals, since the ratio of the speech signal to the combined noise will change over time within small limits and it is necessary to take into account the probability of coincidence of the formants of the speech signal and the formants of the speech-like interference. Therefore, the most acceptable solution is to use the limit state method in calculations and to take into account the specific features of the phonetics of the language.

Besides, we should mention that the limit state method takes into account the hearing abilities of auditors and it focuses on the auditors with good auditory function who are specially trained to recognize signals against the background of noise or other more powerful signals. When selecting auditors, you should also keep in mind the fact that the human hearing aid is formed since childhood and is focused on the structural units of speech of the native language.

So, the proposed method for evaluating speech intelligibility will combine both the use of an experimental research tool and analytical dependencies for the speech signal spectrum, as well as the principle of additivity for the frequency bands of the speech signal.

3. METHOD FOR THE SPEECH INTELLIGIBILITY ASSESSMENT

The features of the Kazakh language phonetics, which can affect to the speech intelligibility in the Kazakh language compared to Russian, English and other languages, are as follows.

The law of syngarmonism of the Kazakh language. The essence of this law is that the vowels of the Kazakh language can be hard or soft. In one word, the vowels can be either hard or soft. Experimental studies have shown that words spoken with soft vowels are 1 dB lower in sound pressure level than words with hard vowels. At the same time, the number of words with hard vowels in the Kazakh language is 59 %, and the number of words with soft vowels is 41 % (data obtained from the analysis of texts).

The peculiarity of speech signals is that from an energetic point of view they have a formant character. Formant - this region of the frequency range in which the main energy is concentrated when pronouncing a certain vowel phoneme. For each vowel phoneme, the number of formants can be from 3 to 5. If consonant sounds have an energy distribution over a frequency range, then vowel sounds are characterized by a concentration of energy in certain areas of the frequency range.

Experimental studies of the energy characteristics of vowels and consonants, deaf and voiced, hard and soft showed that the energy performance of vowels is about 70–78 dB with a rms sound pressure of 70 dB. In this case, stressed vowels are pronounced at a sound pressure of 73–78 dB. For hissing and whistling sound pressure values of 58–63 dB and without clearly expressed formants in the spectrum are characteristic. Speech intelligibility will be determined by the relationship between informational speech signals and the level of masking noise with speech-like signals.

Speech intelligibility is strongly influenced by the combination of vowels and consonants pronounced with an increased level of sound pressure. Vowel sounds are formed on the basis of vibrations of the vocal cords and have a greater power than consonants, which are formed by modulating the air stream. The intelligibility of consonants is not the same. The intelligibility of sonor consonants is higher than hissing ones, and the intelligibility of solids is higher than soft ones. To take into account the above mentioned phonetic features of the Kazakh language, experimental studies were conducted of the amplitude spectrum of speech in the range from 100 to 8000 Hz, performed for a sample of 14 people. Figure 5 shows the averaged amplitude spectra of speech in the Kazakh language (lighter dependence for a female voice, darker dependence for a male voice).

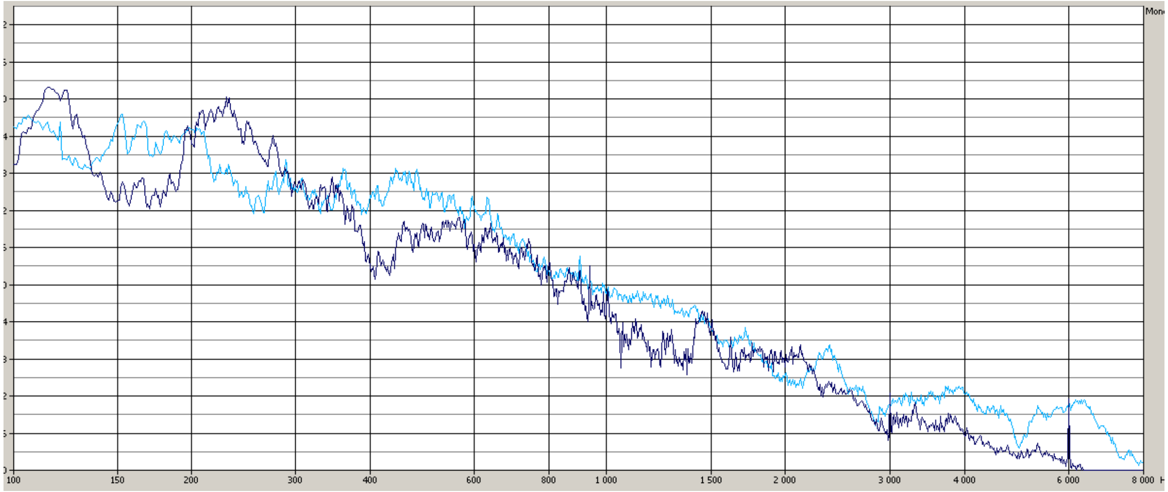

Figure 5. Speech spectrum of Kazakh speakers while reading a text in the Kazakh language.

However, using such dependencies when performing calculations is not entirely convenient. So in [26] the spectral density of the Russian speech in the frequency range from 200 to 5000 Hz is proposed to be approximated by the dependence

$$S_\xi(\omega)=\frac{\rho\cdot\sigma_\xi^2}{\pi}\left[\frac{1}{\rho^2+(\omega_0-\omega)^2}+\frac{1}{\rho^2+(\omega_0+\omega)^2}\right], \quad (4)$$

where $\rho = 1{,}14\cdot10^3$ s^{-1}, $\omega_0 = 2{,}98$ s^{-1} ($f_0 = 210$ Hz), σ_ζ^2 – speech variance.

At the same time $\omega_0 = 2{,}98\cdot10^3$ s^{-1} physically characterizes the frequency close to which there is a maximum at the spectral density of the sound pressure of the speaker's speech, and the first term, taking into account the coefficient $\rho = 1{,}14\cdot10^3$ s^{-1} – the maximum severity. The second term characterizes the decrease in the spectral density of the sound pressure of speech with increasing frequency. As for the choice of specific approximations, their choice is determined by the nature of the dependencies close in appearance to the class of widely used functions, and on the other hand, the approximation coefficients had their own physical interpretation.

Comparison of this dependence with the speech spectrum presented in Figure 5 has showed that these approximating dependencies cannot be used for the Kazakh language. In this regard, the amplitude spectra of speech in the Kazakh language were averaged for 14 speakers and this averaged dependence is shown in Figure 6.

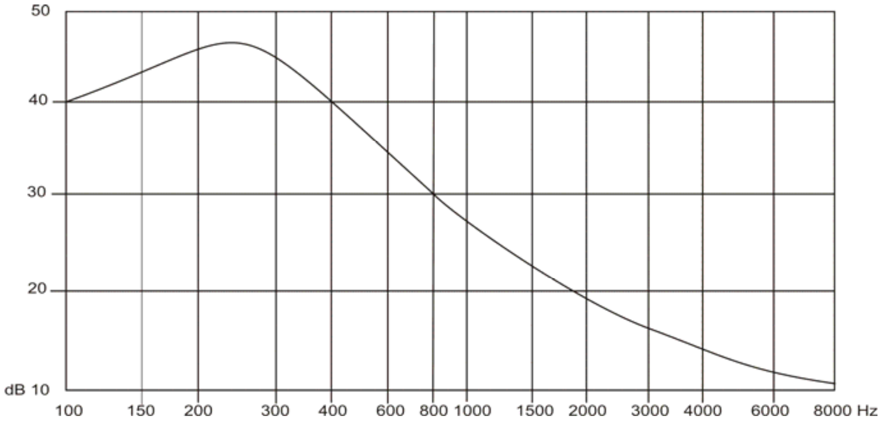

Figure 6. The averaged amplitude spectrum of the Kazakh speakers speech.

According to the experimental studies results performed for a sample of 14 people, the amplitude spectrum of the Kazakh speech, presented in Figure 6, can be approximated for the frequency range from 100 to 8000 Hz and for the integral sound pressure level from 6,3·10^{-3} Pa to 0,36 Pa (from 50 to 85 dB) in this frequency range by the expression

$$S(f)=\rho\cdot P\cdot K\left[\frac{1}{\rho+|(f_0-f)|}+\frac{1}{\rho+(f_0+f)}\right], \quad (5)$$

where $\rho = 100$ Hz, P – the sound pressure level of the speech in the frequency band from 100 to 8000 Hz expressed in Pa, $f_0 = 225$ Hz, K – proportionality coefficient.

At the same time $f_0 = 225$ Hz physically characterizes the frequency close to which there is a maximum at the spectral density of the sound pressure of the speaker's speech. Coefficient ρ close to the value of the frequency of the fundamental tone. The proportionality coefficient has a dimension s$^{-1/2}$ and equaled to 0,0585 s$^{-1/2}$. The first term taking into account the coefficient $\rho = 80$ Hz characterizes the severity of the maximum in the speech spectrum. The second term characterizes the decrease in the spectral density of the sound pressure of speech with increasing frequency.

The spectral density of the speech signal at the places of the speech intelligibility assessment outside the premises is determined from the expression

$$S_r(f)=\rho\cdot P\cdot K\left[\frac{1}{\rho+|(f_0-f)|}+\frac{1}{\rho+(f_0+f)}\right]\cdot K_r(f), \quad (6)$$

where $K_r(f)$ – speech transmission coefficient as a function of frequency (room soundproofing).

The spectral density of the combined masking signals is determined from the expression

$$S_{ms}(f)=\rho\cdot P_{sl}\cdot K\left[\frac{1}{\rho+|(f_0-f)|}+\frac{1}{\rho+(f_0+f)}\right]+S_{wn}, \quad (7)$$

where S_{wn} – spectral density of "white" noise in the speech frequency range and this value is constant; P_{sl} – sound pressure level of speech-like signals in the frequency band from 100 to 8000 Hz.

In expressions 6 and 7, the affiliation indices of the spectral density of speech and speech-like signals are not used, since they are the same in nature depending on the frequency, but differ in amplitude due to different levels of the speech signal and the speech-like signal P and P_{sl}. The spectral density of the masking noise at all control points remains unchanged in time for the white noise component. The spectral component of the speech-like signals at the control points changes over time and is redistributed in frequency in accordance with expression (5). In this case, the spectral components of speech-like signals in amplitude significantly exceed the spectral components of "white" noise (by 6–12 dB in the frequency range up to 500 Hz), but they will be short-term at a given frequency. It should be noted that if the speech-like signals are formed on the basis of the allophone of the speaker whose speech is necessary to protect, then the probability of overlapping of the frequency components of the information signal by the frequency components of the speech-like signal is much higher, since the formants of a certain phoneme of the information signal, for example, phonemes **a**, will exactly coincide with the formants masking speech-like phoneme signal **a**, because the first and the second phonemes belong to the same speaker.

The averaged spectral components of the speech-like masking signal and the protected speech signal will have approximately the same value if the sound insulation is uniform in frequency and the generated noise is close to dependence (5).

Formal speech intelligibility is determined from the expression

$$A=\sum_{k=1}^{k} p_k \cdot w\left(S_r \big/ S_{ms}\right), \quad (8)$$

where p_k is the probability of finding formants in the k-th frequency band in band-frequency analysis; w(S_r / S_{ms}) – speech perception coefficient for a given frequency band and the ratio of the speech signal and the masking signal in a given frequency band.

Verbal speech intelligibility can be determined by formant speech intelligibility using the expressions presented in [2].

However, in a number of works [11, 29, 31, 32] doubts are expressed about the adequacy of the main provisions of the formant theory of speech intelligibility. This primarily relates to the exclusion of the frequency dependence of the coefficient of effective perception of speech formants [11, 29], as well as to the assertions that the amplitude spectral power density of formants and their relationships in frequency bands is the only determining parameter of the information component of a speech signal [29, 32].

The method for the speech intelligibility assessment for information protection systems should be performed according to the limiting states. In [27] it is indicated that the intelligibility limit is –18.5 dB, and to ensure complete security of speech information on speech intelligibility, the signal-to-noise ratio should be –27 dB (taking into account the burst nature of speech).

Recent studies of speech intelligibility for information protection tasks that have been performed recently indicate a number of methodological errors [28–30] in well-known techniques for evaluating speech intelligibility. First of all, this is the absence of taking into account the dependence of the perception coefficient on the frequency and on the sound pressure level of masking noise in excess of 40 dB. In [31], it is pointed out that there is a subjectivity factor in the security of speech information assessment due to the possibilities of a delayed analysis of phonograms of an information speech signal masked by various kinds of interference and the use of various tools. In addition, in [31] it was established that the contribution of voiced phonemes to the intelligibility of isolated words is equivalent to the contribution of consonant (consonants), and for words in sentences, the weight of voiced phonemes in speech intelligibility increases by an additional 40 %.

The essence of the method for the speech intelligibility assessment is that it is first necessary to determine the signal-to-noise ratio in each of the octave frequency bands of the speech signal. This ratio must be determined for those places where it is possible to find the offender. Therefore, the technique is an experimental calculation process: on the one hand, it is necessary to carry out experimental measurements at a given location of sound pressure levels of acoustic production noise and sound pressure levels together signal+noise, since production noise is in no way correlated with the information signal. A noise signal with an envelope corresponding to the spectrum of speech is used as an information signal simulating a speech signal. If in the calculations, intelligibility is determined for an average speech sound pressure level of 70 dB, then for experimental studies the sound pressure level of a speech-modeling signal should be increased by 20 dB to ensure reliable signal reception in a noise environment. Then, the measurement results are converted to a sound pressure level of the modeling signal of 70 dB.

When the acoustic vibrational information signal is excited by the bending vibrations of the building envelope and this information is transmitted outside the premises, the assessment of the security of speech information in this case is determined by the ratio of the level of vibration accelerations caused by the acoustic speech signal and the level of vibration accelerations caused by industrial noise. As in acoustic calculations, the level of acoustic, speech-modeling, influence can be increased by 20 dB, and the level of vibration accelerations caused by this effect should be reduced by 20 dB when converted to the sound pressure level of the modeling signal of 70 dB.

The technique is based on experimental measurements of the information signal-masking noise ratios in the places where the intruder is possible. Using the obtained signal-to-noise ratio, speech intelligibility is determined using a graphical, experimentally obtained dependence of verbal intelligibility on the integral signal-to-noise ratio for the frequency range of the speech signal, and not for octave frequency bands, and for a given type of masking signals. For some types of masking signals and an information signal in Russian, the graphical dependence is approximated by an analytical expression [3, 22].

At the same time, as Bradley points out, when protecting voice information, one must take into account intelligibility, recognition (coding) and audibility. Recognition is when speech intelligibility is absent, but can be determined by the speaker's timbre if the auditor is familiar with the recordings of this speaker. The auditor will hear what the given speaker says, but it's not clear whether this is an information signal or a speech-like masking noise. Audibility is when the auditor can confidently assert that a speech signal is present in a given phonogram record.

S.J. Bradley in [4] has showed that the probability of increasing the signal-to-noise ratio depends on the level of ambient noise at different times of the day. Since speech and noise level vary from moment to moment, therefore, the actual intelligibility of speech will similarly change over time.

4. EXPERIMENTAL RESULTS

The auditors The experimental studies of the spectral density of speech in the Kazakh language made it possible to approximate its dependence on the frequency and take into account the phonetic features of Kazakh speech when assessing the security of speech information using the formant method. This made it possible to create a model of a speech signal in the form of noise with the envelope of the speech spectrum characteristic of the Kazakh language.

For experimental studies, combined masking signals were generated, as described above, and they were superimposed on the phonograms of informational speech signals in the form of a coherent text lasting about 2 to 3 minutes and a volume of 198 to 202 words. Soundtracks were voiced by selected and trained speakers. The signal-to-combined masking noise ratios were –14, –12, –10, –9 dB.

To assess the security of speech information with combined masking signals, tests were carried out, for which 5 auditors were selected aged 20 to 30 years and with a differential sensitivity of hearing to a change in sound frequency of not more than 5 Hz at a frequency of 1000 Hz and high differential auditory sensitivity i.e. the ability to perceive changes in sound intensity from 0.5 to 0.9 dB according to Luscher. In this case, measurements are carried out at an average sound intensity of 40 dB above the auditory threshold and for each of the frequencies 500, 1000, 2000, 4000 Hz.

The results of experimental studies of verbal speech intelligibility when using combined masking signals by trained auditors by repeatedly listening to phonograms by various auditors are presented in Figure 7. Figure 7 also shows the dependence of verbal intelligibility on the signal-to-noise ratio for masking signals in the form of “white” noise.

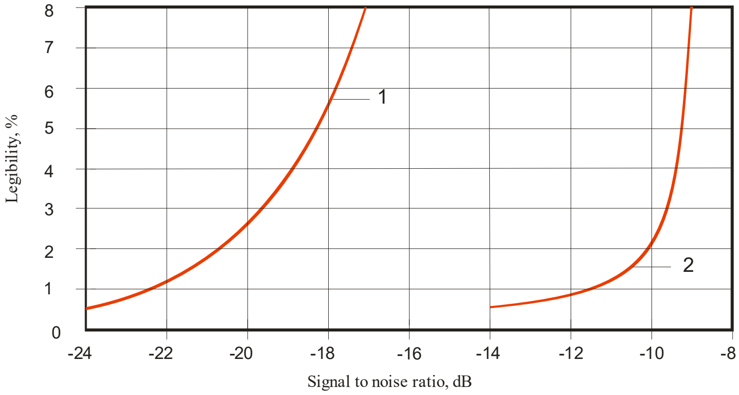

Figure 7. The dependence of the verbal speech intelligibility on the levels of masking signals for the Kazakh language: 1 – with the masking signal in the form of “white” noise; 2 – for the combined masking signals.

Experimental studies of speech intelligibility masked by only “white” noise by selected and trained auditors yielded the following results: with a signal-to-noise ratio of 24 dB, speech intelligibility was less than 1 %. In general, the dependence coincides with the results presented in [6, 30].

The intelligibility of speech masked by combined signals, including white noise and speech-like signals, is less on 6 dB than white noise, for the same selected and trained auditors it was less than 1 % for the signal / combined masking noise ratio of 14 dB. On the dependence of verbal speech intelligibility on the signal / combined masking noise ratio, there is a sharp increase in verbal speech intelligibility after a signal to noise ratio of more than –11 dB. This is explained by the fact that the differences in levels between the signal and the speech-like signals included in the combined masking signals become smaller. In the study of speech intelligibility masked by combined signals, auditors first of all want to recognize more powerful signals – these are speech-like signals against a background of “white” noise. Information speech signals are less in terms of both speech-like signals and “white” noise, and the auditor has to recognize them against the background of speech-like signals and “white” noise.

A sharp increase in speech intelligibility with an increase in the signal-to-noise ratio of more than -11dB is due to the fact that the speech-like interference and the speech signal under protection belong to different speakers. When the signal-to-noise ratio is more than -11dB, a speech signal with a different tone is heard against the background of masking speech-like interference, and a slight his increase of 1-3 dB leads to a sharp increase in intelligibility by 5-6 %. If the speech-like interference and the information speech signal being protected belong to the same speaker, then such a sharp increase in speech intelligibility with an increase in the signal-to-noise ratio of more than -11db will not be observed.

5. CONCLUSION

For the first time there have been obtained experimental dependences of speech intelligibility on the signal-to-noise ratio for combined masking signals that include "white" noise and speech-like signals formed from the base of allophone. Experimental studies were performed using the limit state method by selecting the speakers with high auditory sensitivity and speech recognition abilities against the noise background.

The conducted studies have shown that the use of the combined masking signals, including “white” noise and speech-like signals, to protect speech information is a very effective protection mean and can reduce uncomfortable noise in the room by 8–10 dB compared with the mode when as masking Signals use white noise. An even higher degree of protection of the speaker’s speech information can be achieved when masking speech-like signals are formed on the basis of allophones of the same speaker and then there is a very high probability of the overlap of the formants of the speech-like masking signals on the formants of the protected speech message for given frequencies and frequencies of the fundamental tone. The use of speech-like signals in combination with white noise made it possible to reduce speech intelligibility, as well as the influence of context [33].

The experimental studies of the spectral density of speech in the Kazakh language made it possible to approximate its dependence on the frequency and take into account the phonetic features of Kazakh speech when assessing the security of speech information using the formant method.

6. ACKNOWLEDGEMENT

The work was supported by the grant funding from the Ministry of Education and Science of the Republic of Kazakhstan, No. AP05130293.

REFERENCES

[1] Pokrovskij N.B. Calculation and measurement of speech intelligibility. Moscow, Svyazizdat, 1962, 392p. (in Russian).

[2] Sapozhkov M. A. Speech signal in cybernetics and communication. Moscow, Svyazizdat, 1963, 452 p. (in Russian).

[3] Zheleznjak V.K., Makarov Ju.K., Horev A.A. Some methodological approaches to assessing the effectiveness of voice information protection. Special equipment, 2000, no. 4, pp. 39–45. (in Russian).

[4] Didkovskij V. S, Prodeus A. N. Comparison of formant properties of Ukrainian and Russian speech Electronics and communications. Thematic issue "Electronics And Nanotechnology", p.2, 2009, pp. 88 – 94. (in Russian).

[5] French, Steinberg J. J. . Factors Governing the Intelligibility of Speech Sounds, J. Acoust. Soc. Am., vol. 19, 1947, pp. 90–119.

[6] Bradley S.J., Cover B.N. Designing and Assessing the Architectural Speech Security of Meeting Rooms and Offices: IRC Research Report, RR – 187, August, 2006, 45 p.

[7] American National Standard Methods for Calculation of the Speech Intelligibility Index. American Nationals Standards Institute: ANSI S3.5 1997, New York, 1997, 35 p.

[8] Bradley S.J., Cover B.N. A new system of speech privacy criteria in temps of Speech Privacy Class (SPC) values. Proceeding of 20th International Congress on Acoustics. ICA 2010. Sydney, Australia, 23–27 August 2010, 4 p.

[9] Bradley S.J., Cover B.N. Speech Levels in Meeting Rooms and the Probability of Speech Privacy Problems. J. Acoust. Soc. Am., vol. 127(2), 2010, pp. 815–822.

[10] Byrne, D. et al. An international comparison of long-term average speech spectra. J. Acoust. Soc. Am., Vol. 96 (4), 1994, pp. 2108 – 2120.

[11] Gavrilenko O.V, Didkovskij .V. S., Prodeus A.N. Calculation and measurement of speech intelligibility at small signal-to-noise ratios. Part 1. Correct measurement of the speech distribution function. Electronics and communications. Thematic issue "Problems of Electronics", part 1. 2007, pp. 137–141. (in Russian).

[12] Gavrilenko O.V, Didkovskij V. S., Prodeus A.N. Calculation and measurement of speech intelligibility at small signal-to-noise ratios. Part 2. Correction of perception coefficients Correct measurement of the speech distribution function. Electronics and communications. Thematic issue "Problems

of Electronics", part 1. 2007, pp. 142–147. (in Russian).

[13] Davydau H.V., Patapovich A.V., Seitkulov Y.N. Method for the formation of combined speech masking signals A. International scientific and technical conference dedicated to the 50th anniversary of MRTI-BSUIR (Minsk, March 18–19, 2014): materials conf. V.2, part. 1, Minsk, 2014, pp.344–345. (in Russian).

[14] Davydov G. V., Papou V.A., Patapovich A.V. , Seitkulov Y. N. Li Ye, Fan Yanhong, Jiang Jingsai, Bi Xiaoyan. Method for protecting speech information Reports BSUIR, No.8 (94), 2015, p. 107.

[15] Seitkulov Y.N., Davydou G.V., Potapovich A.V., Justification of the method of forming combined speech masking signals. Herald KazNT, 2014, № 2 (102). (in Russian).

[16] Seitkulov Y.N., Davydau H.V., Patapovich A.V. The base of speech structural units of Kasakh language for the synthesis of speech-like signals. Proceeding of the IEEE 12th International Conference on Application of Information and Communication Technologies, Almaty, 17 – 19 October 2018.

[17] Seitkulov Y.N., Boranbayev S.N., Davydau H.V., Patapovich A.V. Algoritym of forming speech base units using the method of dynamic programming // Journal of Theoretical and Applied Information Technology, 15th December 2018, vol. 96, no. 23, pp.7928–7941.

[18] Davydov G.V., Popov V.A., Patapovich A.V., Seitkulov Y.N., Savchenko I.V. Synthesis of speech-like signals in the Belarusian language. Reports BGUIR, 2015, no. 4 (90), pp. 27–32. (in Russian).

[19] Seitkulov Y.N., Davydov G.V., Potapovich A.V., Justification of the method of forming combined speech masking signals. Herald KazNTU, 2014, № 2 (102). (in Russian).

[20] Zajcev A. P., Shelupanov A. A., Meshherjakov. Technical means and methods of information protection: Textbook for high schools. Moscow, Publishing Engineering, 2009, 508 p.(in Russian).

[21]Horev, A. A. Technical protection of info rmation: textbook. manual for university students, Vol. 1. Technical channels of information leakage. Moskow, SPC "Analytics", 2008, 436 p. (in Russian).

[22] Horev, A. A., Carev N.V Method and algorithm for the formation of speech-like interference. Series: System Analysis and Information Technology. Harold VGU, 2017, № 1, pp. 57–67. (in Russian).

[23] Lobanov, B. M., P'orkovska B, Rafalko Ja., Cirul'nik L. I., Shpilevskij Je. Phonetic-acoustic database for multilingual speech synthesis according to the text in Slavic languages. Computational linguistics and intellectual technologies: proceedings of the international. conf. Dialog'2006, Bekasovo. Moscow, The Science, 2006, pp. 357–363. (in Russian).

[24] Seitkulov Y.N., Davydau H.V., Patapovich A.V. Allophone database for compilation synthesis of speech-like signals in Russian. Modern communications: materials of the XX International Scientific and Technical Conference, Minsk, 14–15 October, VGKS. Minsk, 2014, pp. 193–195. (in Russian).

[25] Vorob'ev V. I., Davydov A. G., Davydov G. V, Ivonin A. I., Leshhenko D. V., Lobanov B. M., Lyn'kov L. M., Popov V. A., Potapovich A. V. Device for protecting speech information from leakage through vibration and acoustic channels: pat. Resp. Belarus' №3053. MPK7 H 04K 3/00, G 10K 11/00. Ofic. bjul. National Center Of Intellectual Property, № 5, 2006, 184p. (in Russian).

[26] Velichkin, A.I. Amplitude speech limitation. Acoustic magazine, 1962, Vol.8, issue 2, pp. 168–174. (in Russian).

[27] Bradley J.S., Gover B.N. Developing a new measure for assessing architectural speech security. Canadian Acoustics, Vol. 31, No. 3, 2003, pp. 50 – 51.

[28] Bucula, A.P., Ivanov A. V., Reva I. L., Trushin V. A. On the reliability of the assessment of the security of speech information from leakage through technical channels. TUSUR reports, № 1 (21), 2010, pp. 89–92. (in Russian).

[29] Trushin V. A., Reva I. L., Ivanov A. V. On methodological errors in evaluating verbal intelligibility of speech in information protection problems. TUSUR reports, № 1(25), 2012, pp. 180–185. (in Russian).

[30] Trushin V. A., Reva I. L., Ivanov A. V. Improving the methodology for assessing speech intelligibility in information security tasks. Polzunovsky Bulletin, № 3/2, 2012, pp. 180–185. (in Russian).

[31] Kozlachkov S. B., Dvorjankin S. V., Bonch-Bruevich A. M. Problems and prospects of protection of acoustic speech information.

Special technique. № 6, 2016, pp. 22–29. (in Russian).

[32] Kozlachkov S. B., Bonch-Bruevich A. M., Dvorjankin S. V., Vasil'evskaja N. V., Selenina A. L. Some features of the formation of an acoustoelectric channel for acoustic speech information leakage. The security of information technology № 4, 2017, pp. 60-70. (in Russian).

[33] Kozlachkov S. B., Dvorjankin S. V., Bonch-Bruevich A. M. Principles of the formation of test rechevy signals in evaluating the efficiency of shumoochistki technologies. Cybersecurity issues, № 3(27), 2018, pp. 9-15. (in Russian).