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

SPEAKERS AND AUDITORS SELECTION TECHNIQUE IN ASSESSING SPEECH INFORMATION SECURITY

[1] YERZHAN N. SEITKULOV, [1] SEILKHAN N. BORANBAYEV, [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 requirements development for speakers and auditors in assessing speech information security. Speakers selection features and methods of their preparation for reading the text are considered. The choice of texts for articulation measurements method when assessing speech intelligibility and masking it with combined signals is substantiated. A technique for selecting auditors for auditory sensitivity and their training in speech perception in conditions of strong noise is proposed.

Keywords: *Speech Intelligibility; Voice Information Security; Announcers; Auditors; Auditory Sensitivity; Articulation Method.*

1. INTRODUCTION

The security of speech information from leakage through acoustic channels is determined by the values of speech intelligibility after speech passes through enclosing structures of premises, i.e. outside the premises. By speech intelligibility it is meant the ratio of the number of correctly taken elements of speech (words or syllables) to the total number of spoken [1–3]. Speech intelligibility is assessed by articulation measurements by a group of speakers and auditors. The complexity of this method is very high, since it requires a sufficiently large samples, both in textual material and in the number of speakers and auditors. At the same time, the question of the criteria of the selection and training of speakers and auditors for the assessment of speech intelligibility in the conditions of high levels of background noise still remains unresolved. To exclude subjective psychological factors on the test results and reduce the time for assessing the level of security of speech information, some instrumental and computational methods for assessing speech intelligibility based on experimental studies were developed [2–7]. These methods are based on measuring the ratio of the speech signal/masking noise in 1/3 octave frequency bands of the speech voice range and determining the articulation index. The determination of speech intelligibility is performed on the basis of its dependence on the articulation index, obtained experimentally. To assess the level of security of speech information, it is necessary to determine speech intelligibility at high levels of masking noise, when speech intelligibility is less than 10%. The methodological errors of instrumental and computational methods for assessing the intelligibility of Russian speech, based on the results of work [2], indicate the lack of consideration of the frequency dependence of auditory sensations and the stepwise change in the coefficients of speech perception during the transition from one octave band to another.

However, these methods are aimed at auditors with an average hearing sensitivity. To solve the problems of assessing the level of security of speech information on speech intelligibility values using computational and computational-instrumental methods, an orientation towards the average auditory sensitivity of auditors is unacceptable. Preliminary experiments have shown that at high levels of masking noise, when speech intelligibility is less than 10%, there is a large variation in intelligibility parameters of the protected speech depending on the hearing sensitivity of auditors. Therefore, it is necessary to change the approach to the selection of speakers and auditors in assessing speech intelligibility for solving problems of protecting speech information and supplement the methods of experimental studies of speech intelligibility with specific

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

methodological materials on the selection and training of auditors to perceive speech signals against masking acoustic noise.

The purpose of this work is to develop a procedure for selection speakers and auditors to assess speech intelligibility in conditions of high noise levels for auditors with increased hearing sensitivity, which will make it possible to determine speech intelligibility by an articulation index with a high degree of reliability and assess the security of speech information.

2. THEORETICAL PART

The fundamental documents on which experimental methods for assessing speech intelligibility are based are: State standard of the Russian Federation "Speech transmission via communication paths" [1]; American National Standard Methods for Calculation of Speech Intelligibility Index [2]; Acoustics – Speech articulation testing method [3].

Standard [1] is designed to assess speech intelligibility values, the quality of voice information transmission paths, both via wire and wireless communication lines, as well as the quality of speech synthesis and perception systems. According to [1], speech intelligibility is determined by the method of articulation measurements using text tables or incomplete syllable tables. The tables are formed taking into consideration the phonetic features of the Russian language. Articulation methods for measuring speech intelligibility are experimental. The experiment should involve at least three auditors and also at least three speakers. At the same time there must be at least two men and one woman aged from 18 to 30 years. Requirements for speakers are the absence of obvious speech defects (explanation is not provided how to determine it). Before reading the text, speakers should familiarize themselves with the texts, master the technique of pronunciation and reading of texts in an even voice without emphasizing individual sounds, as well as maintaining a constant rhythm of reading throughout the text. The training of auditors is carried out by listening to the head phones of speech material and complex articulation combinations from the base presented in [1]. Standards [1–3] are designed to solve a wide range of tasks, ranging from assessing the quality of voice information transmission over communication lines, evaluating the quality of speech information in large conference halls and ending with an assessment of the security of speech information from leakage through acoustic channels.

To solve the problems of assessing the security of speech information, it is proposed to additionally select speakers for clarity of pronunciation and auditors for hearing sensitivity [10, 11]. In addition, auditors should be trained to adapt to perception of speech, distorted in the test path of communications equipment or protected by masking signals.

The first works on the assessment of speech intelligibility by calculation were performed at Bell Telephone Laboratories to reduce long-term and expensive test articulation tests with the participation of speakers and auditors to assess quality while improving telephone communication lines [12]. A year later L. Beranek published a paper [13] on the study of speech intelligibility under noise conditions, in which Bell Telephone Laboratories studies [12] were quoted. These works became fundamental in the future for determining speech intelligibility using the articulation index. A significant step in the development of computational methods for assessing speech intelligibility was the work of K. Kryter (1962) [14, 15], which described and justified a comprehensive method for calculating the index of articulations for a wide range of conditions. Based on this method, the first edition of the American National Standard Methods for Calculation of Speech Intelligibility Index was developed [16].

Calculated methods for assessing the intelligibility of the Russian language are studied in the works of M.A. Sapozhkova [17] and N. B. Pokrovsky [4]. In [8], the graphical dependencies presented in [4] were translated into an analytical form for performing calculations using computer technology.

Calculated and instrumental methods for speech intelligibility assessing are based on the assumption that speech intelligibility is proportional to the average ratio between peak speech levels and masking noise levels made by French N.R. and Steinberg J.C. [12]. It was proposed to split the speech frequency range into 20 bands of equal intelligibility in table 1.

For Russian, the bands borders of equal intelligibility according to [17] are presented in table 2, and according to [4] – in table 3.

Differences in the bands borders for the English and Russian languages are due to the phonetic features of the languages. Differences for the Russian language, presented in different works, apparently, are associated with different methods

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

for determining frequency bands of equal intelligibility and different frequency ranges, for which frequency bands of equal intelligibility are presented. Comparative data band widths of equal intelligibility for English and Russian are presented in table 4.

In table 4, the frequency bands of equal intelligibility for the Russian language are represented by values in the form of two lines. The upper line corresponds to the data from [17], and the lower line corresponds to the data from [4].

However, later, to ensure unification, experimental studies began using standard octave or third octave bands with recalculation of the intelligibility coefficient for these frequency bands [6]. If for frequency bands of equal intelligibility the intelligibility coefficient is 0.05 for each frequency band, then for octave frequency bands it is unevenly distributed and is presented in table 5.

The results of experimental studies were mainly based on samples, which averaged 3–5 people, in some cases reached 10 people. These results should be used with a high degree of caution in determining of speech information security in terms of speech intelligibility. As the founder of the Soviet school of physiology of sensory systems G.V. Gershuni, perception of speech signals and other natural sound signals and speech information processing by the brain is a dynamically developing process, which poses completely new problems in the study of the physiology of hearing and psychoacoustics [18]. In Bradley S.J. paper [19] it is proposed to evaluate the security of speech information using the indicators of intelligibility, hearing and cadence (rhythm). Moreover, if for a intelligibility threshold, which is set at 25 %, the signal-to-noise ratio is estimated at –16 dB, then for a cadence threshold it is –20 dB, and for audibility threshold it is –22 dB. It is proposed to calculate speech intelligibility through the SNR or SPI indicator, using the signal-to-noise ratio for the 16th third of the octave frequency bands [19–22]. SNR is proposed to be determined from

$$SNR = \sum_{f=160}^{5000} \left[L_{ts}(f) - L_n(f) \right] / 16, \text{ dB}, \qquad (1)$$

where the sum is for each of the 1/3 octave band with the average frequency f;

Lts (f) is transmitted level of speech to the location of the offender;

Ln (f) is level of external noise at the location of the offender.

The number in square brackets should be cut so that it cannot be less than –32 dB.

If the signal-to-noise ratio in a specific band is less than –32 dB, then this value, significantly, is below the hearing threshold and such (extremely low) values will inappropriately exaggerate the degree of confidentiality of speech. Therefore, it is necessary to trim or limit the signal-to-noise difference values in each 1/3 octave frequency band to a value not lower than –32 dB.

In this case, the transition from the SNR parameter values to the speech intelligibility indicator is performed using the dependence presented in graphical form in Figure 1 [20]. However, it is very difficult to assess the accuracy of calculating speech intelligibility values, since there are no confidence areas. Figure 1 also shows the variation in speech intelligibility for various signal-to-noise ratios for different auditors, obtained experimentally, which to some extent characterizes the region of confidence intervals. However, with this method of speech intelligibility assessing, the contribution to the speech intelligibility of each of the 16th third octave frequency bands of the speech range is considered equal. This provision is not fully consistent with the French N.R. and Steinberg J.C. assumption [12] on the presence of 20 frequency bands of equal intelligibility.

Table 1: Bands Borders Of Equal Speech Intelligibility For The English Language.

Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz
1	250 – 375	6	955 – 1130	11	1930 – 2140	16	3255 – 3680
2	375 – 505	7	1130 – 1315	12	2140 – 2355	17	3680 – 4200
3	505 – 645	8	1315 – 1515	13	2355 – 2600	18	4200 – 4860
4	645 – 795	9	1515 – 1720	14	2600 – 2900	19	4860 – 5720
5	795 – 955	10	1720 – 1930	15	2900 – 3255	20	5720 – 7000

Journal of Theoretical and Applied Information Technology
30th June 2019. Vol.97. No 12

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

Table 2: Bands Borders Of Equal Speech Intelligibility For The English Language (According To The M.A. Sapozhkov Data).

Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz
1	200 – 330	6	900 – 1060	11	1800 – 2020	16	3200 – 3630
2	330 – 465	7	1060 – 1230	12	2020 – 2260	17	3630 – 4150
3	465 – 605	8	1230 – 1410	13	2260 – 2530	18	4150 – 4790
4	605 – 750	9	1410 – 1600	14	2530 – 2840	19	4790 – 5640
5	750 – 900	10	1600 – 1800	15	2840 – 3200	20	5640 – 7000

Table 3: Bands Borders Of Equal Speech Intelligibility For The English Language (According To The N.B. Pokrovsky Data).

Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz	Band number	Bands borders, Hz
1	100 – 420	6	1030 – 1220	11	1960 – 2140	16	3300 – 3660
2	420 – 570	7	1220 – 1410	12	2140 – 2320	17	3660 – 4050
3	570 – 710	8	1410 – 1600	13	2320 – 2550	18	4050 – 5010
4	710 – 865	9	1600 – 1780	14	2550 – 2900	19	5010 – 7250
5	865 – 1030	10	1780 – 1960	15	2900 – 3300	20	7250 – 10000

Table 4: Band Width Of Equal Intelligibility For The Russian And English Language.

Band number	1	2	3	4	5	6	7	8	9	10
Band width for the English language	125	130	140	150	160	175	185	200	205	210
Band width for the Russian language	130	135	140	145	150	160	170	180	190	200
	320	150	140	155	165	190	190	190	180	180
Band number	11	12	13	14	15	16	17	18	19	20
Band width for the English language	210	215	245	300	355	425	520	660	860	1280
Band width for the Russian language	220	240	270	310	360	430	520	640	850	1360
	180	180	230	350	400	360	390	960	2240	2750

Table 5: Intelligibility Coefficient For Octave Frequency Bands.

Frequency band, Hz	87,5–175	175–350	350–700	700–1400	1400–2800	2800–5600
Intelligibility coefficient	0,067	0,125	0,212	0,294	0,250	0,052

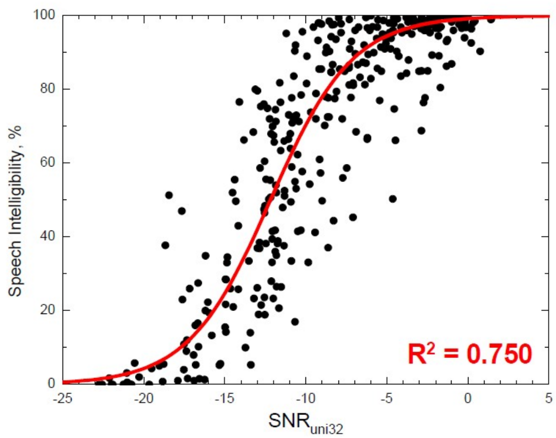

Figure 1: Speech Intelligibility Depending On The Signal-To-Noise Ratio

Speech intelligibility at low signal-to-noise ratios was experimentally studied only in some papers [20, 22] because of the main task of ensuring speech intelligibility in communication channels, and extrapolating speech intelligibility characteristics in areas with values up to 10 % can lead to significant errors.

3. REQUIREMENTS FOR SPEAKERS

Requirements for speakers is ability to clearly read a coherent text at a speed of 70–80 words per minute and with a difference between the average amplitude of 10 maximum values of the speech signal during the reading of 200 words of connected text and the rms value of the speech signal for this period is not less than 18 dB. When selecting speakers, it is necessary to pay attention to the pronunciation of difficult for articulation consonant sounds p, l, s, z, sh, zh, ch, sch [23]. If, during articulation of vowel sounds, the speech apparatus is open and freely passes the air and vibrations of the vocal cords are excited, then during articulation of the consonants, the speech apparatus forms a coordinated position of the tongue, lips, mouth and nose and has rather strong muscular tension. Phonetic disturbance of pronunciation is called dyslalia (violation of sound pronunciation during normal hearing and intact innervation of the speech apparatus according to B.M. Grynshpun). It should be noted that the articulation apparatus of a person is formed in childhood and the training of speakers can be carried out after the selection of persons with a well-developed articulation apparatus. The training of speakers includes the formulation of correct breathing, a detailed introduction to the text.

In the paper [24], it is proposed at experimental evaluation of speech intelligibility in the conditions of noise the speakers number and the auditors number should be the same and include both male and female speakers. The room for recording the tests should be quiet and without an echo, and the recorded speech should not be so loud that the amplitude is not limited and not so quiet that extraneous noises are heard. In normal quiet reading, the speech recording level should be about 12 dB lower than the full scale span [25].

4. REQUIREMENTS FOR AUDITORS

The auditors selection is usually performed using an assessment test and excluding extreme emissions in test results [24]. According to [25, 26], auditors are audited by audiometry. At the same time, auditors should have a hearing threshold of no more than 20 dB at any of their frequencies from 125 Hz to 8 kHz. When using students as auditors, it should be noted that candidates for the role of auditors who arrived to participate in the selection with in-ear headphones or use them quite often are unlikely to be suitable for participating in the assessment of intelligibility of noisy speech. In [27], it is indicated that eight auditors selected are sufficient to assess the intelligibility of noisy speech, but the author usually used 20 auditors.

To assess the security of speech information on speech intelligibility indicators, calculations should be carried out on the limiting states of auditory sensitivity of auditors. Therefore, it is necessary to change the approach to the selection of auditors and to supplement the calculation methods with specific results of experimental studies, taking into account the auditors' ability to contrast sensitivity and readiness to perceive acoustic signals against the background of masking noises. The selection of auditors is recommended for the following characteristics of the sensitivity of hearing.

Auditors should have high auditory sensitivity with a perception threshold of pure tones of 0–5 dB in the frequency range from 500 to 2000 Hz. Sensitivity with a threshold of 6–10 dB is medium, and with a threshold of 11 dB or more – reduced [27].

Secondly, to have a high differential hearing sensitivity, i.e. the ability to perceive changes in

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

sound intensity according to Luscher. In this case, measurements are carried out at an average sound intensity of 40 dB above the hearing threshold and for each of the frequencies 500, 1000, 2000, and 4000 Hz. If the auditor is able to distinguish, under such conditions, changes in sound intensity from 0.5 to 0.9 dB, then he has a high differential hearing sensitivity. Differential hearing sensitivity from 0.9 to 1.5 dB is considered to be moderate, and more than 1.5 dB is reduced. The differential acoustic sensitivity of the sound pressure depends on the sound pressure level. As the sound pressure level rises, the differential hearing sensitivity of the sound pressure rises, and at sound pressure levels of 80 dB, its average value is about 0.6 dB, and at sound pressure levels of 55 dB, the average value is 1 dB.

In addition to the differential hearing sensitivity, it is necessary to determine the differential sensitivity of hearing to changes in the frequency of the sound according to the sound intensity. The maximum differential sensitivity of hearing is noted at frequencies in the region of 1000 Hz. When the auditor catches a change in the frequency of a tone of 1000 Hz at 5 Hz, it is highly sensitive, from 6 to 10 Hz – medium, to 11 Hz or more – low (unsatisfactory).

Contrast sensitivity is the ability to perceive an acoustic signal against the background of more powerful masking acoustic signals. Determining the contrast sensitivity of hearing is performed by assessing the auditor's ability to hear a sound with a frequency of 1000 Hz against the background of a sound with a frequency of 400 Hz and a sound pressure of 40 dB. If in such conditions a sound with a frequency of 1000 Hz and a sound pressure of 15–20 dB is heard by the auditor, then the auditor is considered to have a high contrast sensitivity. When sound is perceived with levels of 21–30 dB, the contrast sensitivity of hearing is average, and sound perception with 31 dB or more indicates a low contrast sensitivity of hearing [27].

Auditors should have good binaural hearing – the ability to determine by ear the location of the sound source. For these studies, it's possible to use Perekalin's lateral meter. Binaural hearing is rated as good if the differential threshold is 3 to 10 degrees.

To study the rhythmic hearing and memory for the rhythm, it's possible to use the rhythmography apparatus. The rhythmic motive recorded by telegraphic signs (dots and dashes) is fed through headphones to the auditor, who memorizes it and writes it in a convenient form. The rhythmic pattern is repeated three times. The rhythmic pattern transfer speed is about 40 characters (dots and dashes) per minute. If the auditor is able to accept and reproduce 15 to 20 characters, then he has a good rhythmic ear. Usually the sequence and type of rhythmic pattern is formed from the letters of the telegraph alphabet. Studies of memorizing the rhythmic pattern of long duration and the effect of the speed of its transmission on the ability of auditors to memorize is described in [28].

5. EXPERIMENTAL TECHNIQUE

5.1 Selection Of Phonetically Balanced Texts

Phonetically balanced (representative) text is understood as a textual material in which the frequency distribution of phonetic units (phonemes, allophones, syllables) corresponds to the common language distribution obtained from statistical analysis of the reference text corpus [29]. At the same time, a phonetically balanced text should also take into account the frequency distribution of the length of words, sentences and paragraphs characteristic of the given language. For different languages, the frequency distribution of phonetic units will be different, taking into account the peculiarities of the language. So for the Kazakh language, the frequency distribution of phonetic speech units is presented in [30, 31]. When forming texts for the study of intelligibility of speech in Chinese, it is necessary to follow the standard [3].

For the Belarusian language, ready-made phonetically balanced texts need to be prepared independently, as they are not found in publications. Experimental studies of the frequency distribution of phonemes for the Belarusian language are presented in table 6 [32]

The probability distribution of the letters number in a word for the Belarusian language is presented in table 7.

For the Russian language, it is proposed to use the texts presented in the standard [1] (Appendix D). These texts are separate sentences that are not related to a common theme. This requirement for the texts preparation for the determination of noisy speech intelligibility is very important for solving problems, namely, the speech information security. Individual words or syllables for solving these problems should not be used. This is due to the fact

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

that a person possesses associative thinking and, recognizing one of the words of a sentence, he is able to recognize the subsequent word, although it is not clearly heard. Sentences should not be long (no more than 5–6 words). The total number of words in the text should be about 200. This is due to the fact that during the text reading the announcer does not accumulate fatigue and the text will be read with exactly one volume level and one tempo. On the other hand, with such text length, the auditor's fatigue will not come either. These requirements must be considered when determining the noisy speech intelligibility in the Belarusian, Russian and Kazakh languages. With regard to assessing the speech intelligibility in Chinese, the individual syllables presented in [3] (Appendix A) can be used for this language. This is because associative thinking is not as effective for recognizing noisy speech in Chinese, since the language is for the same acoustic sounding of the word corresponds to a large number of its semantic meanings. The number of texts from [1] to study the intelligibility of Russian speech should be at least 3 and not more than 5 in order to be able to exclude accidents of receiving information about the text content by the auditor.

5.2 Speakers Selection

Speakers should be selected from people between the ages of 18 and 30 with good articulation. The number of announcers should be at least 10 people, 5 males and 5 females. The pre-selection of speakers may include a greater amount, because after recording and processing phonograms, some recordings will be rejected according to the ratio of the maximum sound pressure levels of speech to the root-mean-square value, which must be at least 18 dB.

5.3 Training Announcers

The narrator's training consists in the formulation of correct breathing when reading a text. With proper breathing, the speech apparatus is not overworked and the strength of the voice, the richness of the dynamic hues and the melody of speech are ensured. The system of exercises for voice and correct breathing is detailed in [33, 34].

5.4 Record Phonogram Text

Text phonograms recording must be performed in an acoustically muffled camera or in a recording studio. The microphone should have a uniform amplitude-frequency characteristic in the frequency range from 50 to 10,000 Hz with an irregularity of no more than 8 dB. The self-noise of the microphone should not exceed 20 dB with a scale of A, and the sensitivity should be above 18 mV / Pa. Phonograms must be digitized with a sampling frequency of at least 22 kHz.

5.5 Recorded Phonograms Selection

Recorded phonograms selection is performed taking into account the requirements for legibility and speed of reading the text. Phonograms with the ratio of the maximum sound pressure levels of speech to the rms value less than 13 dB should be discarded. Phonograms, which have a reading speed of less than 50 words per minute and more than 100 words per minute, are also rejected.

5.6 Overlaying Noise On Phonograms And Their Valuation

Before imposing noise on phonograms, it is necessary to find a speech spectrum for each phonogram in 1/3 octave and octave frequency bands. For 1/3 octave bands, the frequency range should be from a band with a geometric mean frequency of 63 Hz to a band with a geometric mean frequency of 10000 Hz. For octave frequency bands, the range should be from a band with a geometric mean frequency of 63 Hz to a band with a geometric mean frequency of 8000 Hz. In addition, the rms sound pressure level should be calculated for each of the phonograms. All these data are recorded in the passport phonogram. Phonograms should be superimposed with «white» noise with a uniform power spectral density in the frequency range from 80 to 10000 Hz and attenuation outside the specified band of 30 dB per octave. The ratio between the rms value of the signal and the rms noise value should be –12 dB, –14 dB, –16 dB, –18 dB, –20 dB, –22 dB, –24 dB, –26 dB, –28 dB and -30 dB . For each of the phonograms, it is necessary to determine the signal-to-noise ratio in 1/3 octave and octave frequency bands and record the phonogram in the passport.

5.7 Auditors Selection

The auditors selection should be carried out taking into account the requirements set out above. Speakers should be selected from people between the ages of 18 and 30 with good auditory sensitivity. The number of auditors should be at least 20 people, 10 male and 10 female. Pre-selection of auditors may include more.

All selected auditors should have the following characteristics of auditory sensitivity:

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

a) the auditory sensitivity should be with a pure-tone perception threshold of no more than 0–5 dB in the frequency range from 500 to 2000 Hz;

b) differential hearing sensitivity in terms of sound level is not more than 0.5 to 0.9 dB with an average sound intensity of 70 dB at frequencies of 500, 1000, 2000, 4000 Hz;

c) the differential sensitivity of hearing to a change in the frequency of the sound should be no more than 5 Hz from the center frequency of a tone of 1000 Hz and a sound pressure level of 70 dB;

d) contrast sensitivity to sound at a frequency of 1000 Hz relative to a sound with a frequency of 400 Hz should be no more than 45 to 50 dB, with a sound pressure at a frequency of 1000 Hz in 70 dB;

e) binaural hearing should be characterized by an angle of no more than 3 to 10 degrees at a frequency of 1000 Hz with a sound pressure of 70 dB;

f) the ability to memorize and reproduce a rhythmic pattern with the number of characters not less than 15–20.

Special attention should be paid to such characteristics as contrast sensitivity of hearing.

The equipment for determining the auditory sensitivity included a control computer, headphones, and software allowing the generation of signals and their changes. The generated signals from the output of the sound card were fed to the headphones. Signal changes were performed after a period of time equal to 3 seconds when the signal passes through zero. Studies of binaural hearing should be conducted in an acoustically muffled chamber.

5.8 Training

Auditor training should be aimed at developing the ability to perceive speech messages against the background of broadband noise masking speech. At the first stage, exercises aimed at adapting the auditor to the voice, rhythm and character of the pronunciation of a particular speaker are performed. The second stage of training includes the auditor adaptation to the masking noise by repeatedly listening to the headsets of the phonograms of masking noise. The third stage is the auditors training – this is listening to the headphones of the phonograms of the speaker's speech with masking signals, adaptation to the voice of which was performed at the first stage of training. This begins listening to phonograms with signal-to-noise ratios –12 dB and further with a decrease in this ratio. The fourth stage of training includes the implementation of the first three stages, but for two more speakers.

5.9 The Study Of The Intelligibility Of Audible Noisy Phonograms

The study of the intelligibility of noisy phonograms begins with the exercises on the adaptation of the auditor to the masking noise by listening to them on headphones for 3 to 5 minutes. Then begins listening to noisy phonograms, starting with the phonogram with the highest signal-to-noise ratio – this is 30 dB. The auditor records the number of recognized words. At the same time, it is possible to repeatedly listen to this phonogram until the auditor is convinced that the number of recognized words cannot be more. The recommended number of auditions of a noisy phonogram is not more than 10 times, however, the auditor himself determines the necessary number of auditions. Further, the auditor proceeds to determine the intelligibility of the next noisy sound track with a signal-to-noise ratio 2 dB more – 28 dB. Experiments continue further with an increase in the signal-to-noise ratio until the auditor can recognize 30 % percent of words from a phonogram – that is 60 words with an average number of words in a phonogram of about 200 words. The total research time should not exceed 4 hours per day, as the announcer's fatigue further comes. In this case, one day is not allowed to listen to the soundtracks of another speaker or noisy soundtracks of the same speaker, but with different text. Legibility is determined verbal as recommended in [1].

Table 6: The Probability Distribution Of Phonemes Occurrence In Speech In The Belarusian Language.

Allophone	Probability	Allophone	Probability	Allophone	Probability
a	0,1642	m	0,0274	ch	0,0158
b	0,0211	n	0,0571	sh	0,0143
v	0,0287	o	0,0413	y	0,0405
g	0,0198	p	0,0289	soft sign (')	0,0149
d	0,0339	r	0,0393	Ə	0,0076
e	0,0395	s	0,0404	ju	0,0055
zh	0,0085	t	0,0336	ja	0,0376
z	0,0303	u	0,0338	jo	0,0065
j	0,0092	f	0,0005	i	0,0508

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

k	0,0412	kh	0,0125	short u	0,0257
l	0,0436	cz	0,0260		

Table 7: The Probability Distribution Of The Letters Number In A Word (Word Length).

Letters number in the word	1	2	3	4	5	6	7	8
Probability	0,103	0,126	0,103	0,108	0,130	0,128	0,115	0,077
Letters number in the word	9	10	11	12	13	14	15	
Probability	0,051	0,031	0,016	0,007	0,003	0,001	0,001	

6. EXPERIMENTAL RESULTS

During experimental studies the effect of selected texts on the results is examined. Speaker 10 speech spectra obtained for various tables D50 and D70 [1] (Appendix D) are shown in figure 2.

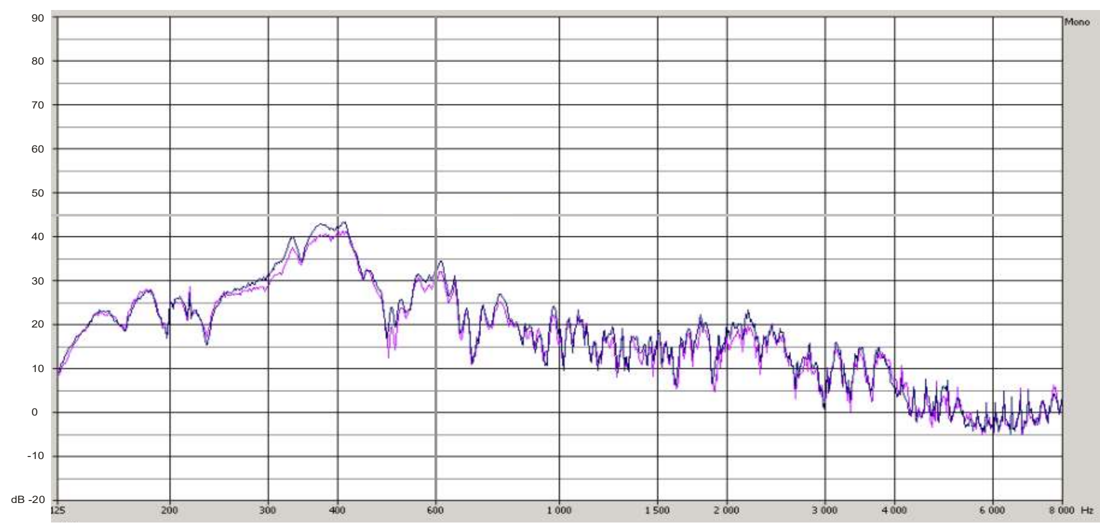

Figure 2: Speaker 10 Speech Spectra For Text From Tables D50 And D70

From figure 2 it is clear that for speaker 10 the text influence does not cause significant changes in the spectrum of his speech. However, for speaker 20, there are significant differences in the spectrum of speech depending on the text, which is presented in figure 3.

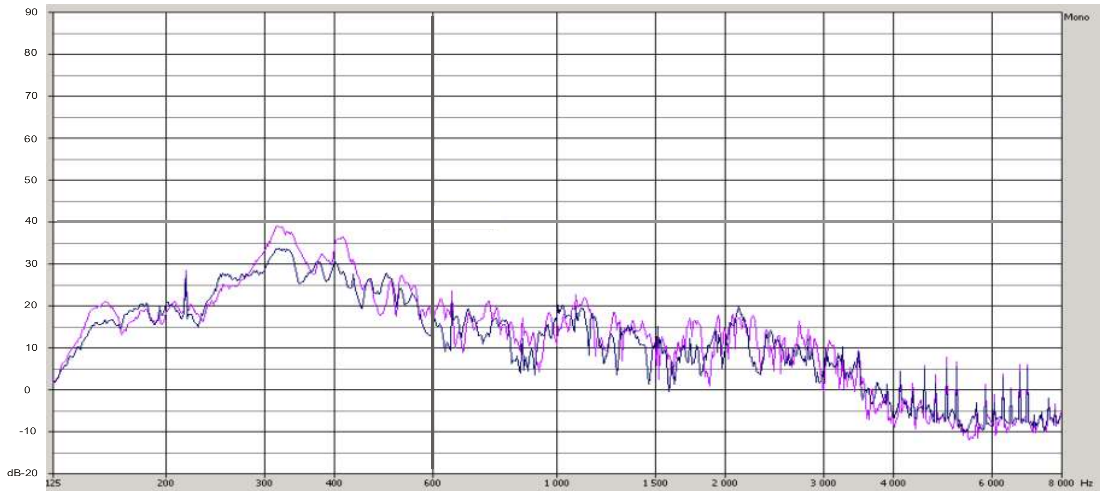

Figure 3: Speaker 20 speech spectra for text from tables D50 and D70

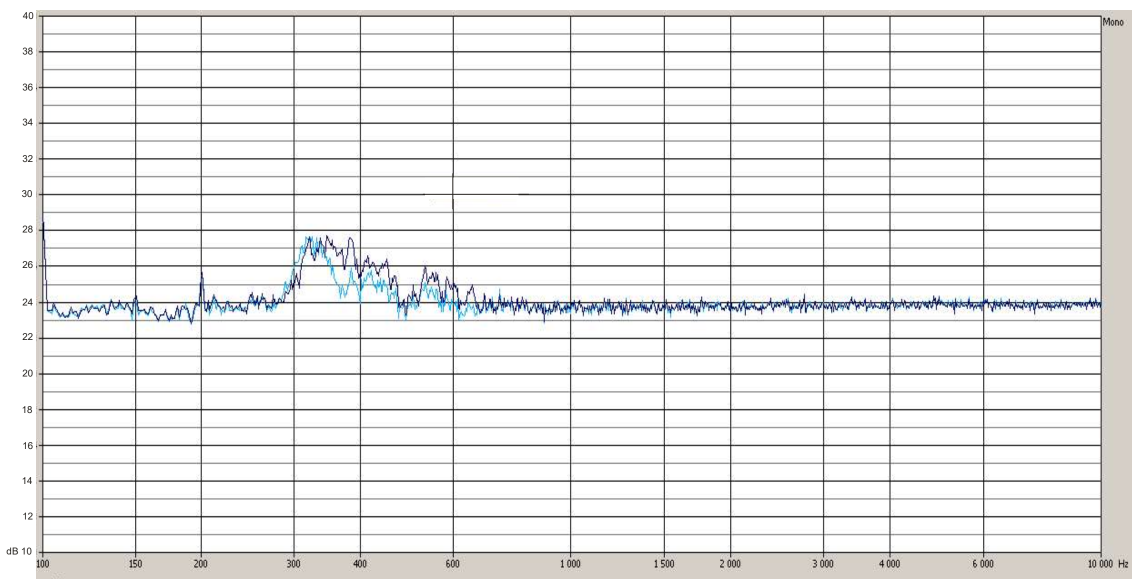

Figure 4. Speaker No. 1 and No. 2 speech spectrum masked by "white" noise with a signal-to-noise ratio of 20 dB.

If for speaker 10 the ratio of the maximum sound pressure levels of speech to the rms value for text D50 and text D70 was 16.1 dB and 16.4 dB, respectively, then for speaker 20 the ratio of the maximum sound pressure levels of speech to rms value for text D50 and text D70 were respectively 21.1 dB and 22.2 dB. In this case, speaker 20 read the text of the D50 by 1.2 dB with greater sound pressure than the text of the D70, which also affected the differences in the spectra of speech by the announcer when reading different texts.

The feature of speech signals is that they have the formant's character in terms of energy. Formant is a frequency range domain in which the main energy is concentrated when pronouncing a certain vowel phoneme. The number of formant can be from 3 to 5 for each of the vowel phonemes. While consonant sounds have an energy distribution throughout a frequency range, vowel sounds are characterized by an energy concentration in certain domains of the frequency range.

Experimental studies of the energy characteristics of vowels, breath consonants and sonant in Russian have demonstrated that the energy indices of vowels are about 70–78 dB with a root-mean-square value of sound pressure for the whole text 70 dB. The probability of vowel sounds in the Russian language is 0.486, and for the text D50 it was 0.421, for the text D70 it was 0.433. In this, the vowels under the stress are pronounced with a sound pressure of 73 - 78 dB. Sound pressure for hissing and whistling consonants is in 58 - 63 dB without pronounced formants in the spectrum. The probability of hissing and whistling sounds in the Russian language is 0.423 and for the text D50 it was 0.113, for the text D70 - 0.1129. In this, speech intelligibility will be determined by the ratio between the information speech signals and the level of masking noise.

Speech intelligibility studies with a signal-to-noise ratio of 20 dB for table D50 and speakers No. 1 and No. 2 were conducted by two groups of auditors. The first group of auditors included persons aged 20 to 30 years without special selection and training. The second group of auditors consisted of individuals who were selected and trained. While for the first group of auditors the average verbal intelligibility was 1.5%, for another group of auditors it was 4.5%. Figure 4 shows the speech spectrum of speakers No. 1 and No. 2, masked by “white” noise with a signal-to-noise ratio of 20 dB.

From figure 4 it can be seen that the energy of the speech signal is concentrated in the frequency range from 260 to 600 Hz. This makes it possible to see that, the spectral density of the sound pressure rises in this frequency range, on average, by 3–4 dB.

7. CONCLUSION

The procedure for selection of speakers and auditors to determine speech intelligibility is proposed for conditions of high noise levels when verbal speech intelligibility does not exceed 20%. These experimental data are important in finding the correspondence between the articulation index and the meanings of the verbal intelligibility of speech.

Proposed technique was tested on 5 texts from [1] (Appendix D) of the table 50, 70, 76, 94, 100, the audio recordings of which were masked by «white» noise with the signal-to-noise ratio –12 dB, –14 dB, –16 dB, –18 dB, –20 dB, –22 dB, –24 dB, –26 dB, –28 dB and –30 dB. The selected speakers number was 28: 20 males and 8 females. The auditors number was 24: 16 males and 8 females. All announcers and auditors were between the ages of 18 and 30. The technique has confirmed that the speech intelligibility when masking it with noise is influenced by factors related to the selection of announcers and auditors.

The future development of this technique will be performed for the Kazakh language, taking into consideration its phonetic features and the vowel harmony law.

8. ACKNOWLEDGEMENT

This work was supported by grant funding from the CS MES RK, №AR05130293.

REFERENCES

[1] STB GOST R 50840-2000 (State Standard System). Transmission of speech through communication paths. Methods for assessing quality, intelligibility and recognizability. Minsk, 2000, 366 p. (in Russian)

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

[3] GB/T 15508 – 1995. Acoustics – Speech articulation testing method, 1995, 15 p.

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

[5] Masking Speech in Open-Plan Offices with Simulated Ventilation Noise: Noise Level and Spectral Composition Effects on Acoustic Satisfaction by Jennifer A. Veitch, John S. Bradley, Louise M. Legault, Scott Norcross, & Jana M. Svec Internal Report No. IRC-IR-846 Date of issue: April 2002.

[6] Architectural speech security of offices and meeting rooms Bradley, J.S. NRCC-48326 Proceedings of the International Conference on Building Acoustics, Chonnam University, Gwangju, South Korea, April 2004, pp. 16–34.

[7] Trushin V.A, Reva I.L., Ivanov A.V. About methodological errors in the assessment of verbal speech intelligibility in information protection tasks. Reports TUSURa, vol. 1 (25), June 2012, pp. 180–184. (in Russian)

[8] 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)

[9] Rashevskij Ja.I., Kargashin V.L. Review of foreign methods for determining speech intelligibility. Special equipment, 2002, no. 3–6; 2003, no. 1. (in Russian)

[10] Seitkulov Y.N., Davydov G.V., Patapovich A.V. Requirements for auditors and factors in assessing the security of voice information. Report Abstracts of the XII Belarusian-Russian Scientific and Technical Conference Technical means of information protection, Minsk, BGUIR, 2014, pp. 11–12. (in Russian)
[11] Gotovko M.A., Davydov G.V., Seitkulov Y.N. Evaluation of voice information security. Information technology and systems 2013 (ITS 2013): materials of the international scientific conference, BGUIR, Minsk, Belarus', 23 October 2013. (in Russian)
[12] French N, Steinberg J. Factors Governing the Intelligibility of Speech Sounds. Acoust. Soc. Am, 1947, Vol. 19, pp. 90–119.
[13] Beranek, L. Desing of Speech Communication Systems, L. Beranek, Proceedings of the IRE – 1947, Vol. 35, Issue 9, pp. 880 – 890.
[14] Kryter, K.D. Methods for the calculation and use of the Articulation Index, K.D. Kryter, Acoust. Soc. Am, 1962, Vol. 34(11), pp. 1689–1697.
[15] Kryter, K.D. Validation of the Articulation Index, Acoust. Soc. Am., 1962, Vol. 34(11), pp. 1698–1702.
[16] American National Standard Methods for Calculation of the Speech Intelligibility Index. American Nationals Standards Institute: ANSI S3.5 1997, New York, 1997, 35 p.
[17] A.P. Efimov, A.V. Nikonov, M.A. Sapozhkov, V.I. Shorov, Acoustics: Handbook, M. Radio and communication, 1989, 336 p.
[18] Chistovich L.A., Speech physiology. Human perception of speech, Guide to Physiology, L., 1976, 334 p.
[19] Bradley S.J., Designing and Assessing the Architectural Speech Security of Meeting Rooms and Offices: IRC Research Report, RR – 187, August, 2006, 45 p.
[20] 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.
[21] 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.
[22] 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.
[23] Felicheva T.B., Cheveleva N.A., Chirkina G.V. Basics of Speech Therapy: A manual for students, teachers interns on specials. Pedagogy and psychology.. Moscow, Education, 1989, 223 p.
[24] Voiers W.D. Evaluating processed speech using the diagnostic rhyme test. Speech Technology, Jan. 1983, pp. 30 – 39.
[25] McLoughlin J. Subjective intelligibility testing of Chinese speech. IEEE Transactions on Audio Speech and Language Processing 16 (1), 23–33 February 2008. DOI: 10.1109, TASL.2007.909450.
[26] ANSI., ANSI S3.2. Method for Measuring the Intelligibility of Speech over. Communication Systems, 1989.
[27] Soldatov I.B. Lectures on otorhinolaryngology. Moscow, The medicine, 1990, 288 p. (in Russian)
[28] Warren R.M. Auditory temporal discrimination by trained listeners. Cong. Psychol., 1974, pp. 237 – 256.
[29] Smirnova N.S., Hitrov M.V. Phonetically representative text for fundamental and applied studies of Russian language. Proceedings of higher educational institutions, Instrumentation, 2013, vol. 59, no. 2, pp. 5 – 10. (in Russian)
[30] 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.
[31] 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.
[32] 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)
[33] Nazarova L.V., Setting voice. The system of practical exercises: Methodical recommendations. Chelyabinsk: Publisher JuUrGU, 2004, 27 p. (in Russian)
[34] Njensi Zi The Art of Breathing: Six easy lessons for success, health, and prosperity. Sofija, 2003, 271 p. (in Russian)