Speechlike Signal Synthesis Module For Information Security Systems

PetrovSergeiNikolaevich – Faculty Of Info-communications, Belarusian State University Of Informatics And Radio-electronics,
e-mail meinchest@inbox.ru
Zelmanski OlegBorisovich– Faculty Of Info-communications, Belarusian State University Of Informatics And Radio-electronics,
e-mail 7650772@rambler.ru
PulkoTataynaAlexandrovna – Faculty Of Info-communications, Belarusian State University Of Informatics And Radio-electronics,
e-mail pulko@bsuir.by

ملخص:

تحليل مقارن لفعالية إشارات الضوضاء من أنواع مختلفة من أجلال حد من وضوح الكلام المقنع في حالة التسرب من خلال القنوات الصوتية والصوتية الاهتزازية المباشرة. كإشارة اختبار تستخدم جداول التعبير اللفظي من STB GOST R 50840-2000 القياسية. لقد تم اقتراح تقنية لاختبار فعالية أنواع مختلفة من إشارات الضوضاء. تتمثل إحدى نتائج العمل في تطوير طريقة لتشكيل إشارة تقني فعالة للحماية الفعالة للمعلومات الصوتية،وهي عبارة عن ضوضاء تشبه الكلام،تتشكل مباشرة من إشارة الكلام المقنّع، والتي يمكن أن يكون مستواها أقل من 7- 9 ديسيبل تحت مستوى الضوضاء البيضاء لضمان القيمة المطلوبة من وضوح الكلام. تشتمل طريقة توليف التداخل التي تشبه الكلام على المراحل التالية: الكشف عن الكلام،والتحقق من السماعة،وتقسيم الكلام،وتصنيف الكلام،وتوليف التداخل الذي يشبه الكلام. يعتمد توليف التداخل المشابه للكلام على تجميع التغيرات في الصوت لحرف معين بواسطة نص صوت يتم إنشاؤه مسبقا مع مراعاة إحصائيات اللغة. انها تسمح لتشكي لشبه الكلام في اللغات الروسية أو الإنجليزية أو العربية.

Abstract

A comparative analysis of the effectiveness of noise signals of different types in order to reduce the intelligibility of masked speech in the case of leakage through direct acoustic and vibro-acoustic channels. As a test signal used articulation phrase tables from the standard STB GOST R 50840-2000.

A technique for testing the effectiveness of various types of noise signals is proposed. One of the results of the work is the development of a method for the formation of an effective masking signal for active protection of acoustic information, which is a correlated speech-like noise, formed directly from the masked speech signal, the level of which can be 7-9 dB below the white noise level to ensure the required value of speech intelligibility. Speech-like interference synthesis method includes the following stages: speech detection, speaker verification, speech segmentation, speech classification, speech-like interference synthesis. Synthesis of speech-like interference is based on compilation of allophones by phonemic text generated taking into account language statistics. It allows to form speech-like interference in Russian, English or Arabic languages.

Keywords*: speech intelligibility, speech-like interference, articulation tests, speech leakage channels, compilation synthesis of speech-like interference.*

Introduction

Speech data protection is one of the most important tasks in general set of activities on data security of a facility or an establishment. One of the most efficient ways of speech data protection from unauthorized audio interception is the active one which provides creation of masking noise in critical points of the premises. Active protection is implemented by different types of jamming devices. Speech is a noise-like process with complicated amplitude and frequency modulation and consequently the best form of masking interfering signal is also noise process with normal law of distribution of density of instantaneous values probability (i.e. white or pink noise). Speech-like interference is especially effective in terms of features similarity. The goal of this work is to study the efficiency of different types of interfering signals. Noise efficiency will be evaluated through speech intelligibility. The method include the following steps: test articulation tables recording; speech-like interference generation; noise masking of test signals with interfering one; noisy signal recording; determination of signal-to-noise ratio of test and interfering signals; calculation of errors and measurement confidence interval; determination of intelligibility and evaluation of speech signal security. Security evaluation is carried out using speech intelligibility value.

1 Record of articulation tables

Test acoustic signals were generated based on articulation tables from standard STB GOST R 50840-2000 (Speech transmission over varies communication channels. Techniques for measurements of speech quality, intelligibility and voice identification). Speech intelligibility value is the criterion of speech data security for speech signals (Zhelezniak, Makarov, Horev, 2000, pp. 39-45). Initial data necessary for the analysis of acoustic speech signals security are sound pressure level of speech signals, sound pressure level of background acoustic noise. It is obvious that the speech intelligibility index can also be used to assess the effectiveness of the protection of speech information from leakage through technical channels.

Experimental evaluation of speech intelligibility should be carried out in compliance with STB GOST R 50840-2000 using phrase articulation tables. The team of speakersconsisting of five men and five women was put together. Recording of acoustic signals representing test phrases from the STB GOST R 50840-2000 was performed using the microphoneAKG P120, audio interface Focusrite Scarlett 2i2 and personal computer with installed software Sound Forge 9.0 in acoustically muffled room. Processing of the audio files obtained was carried out in Sound Forge 9.0 software environment. Not less than 40 sound records are presented for evaluation. Each record should contain 50 phrases masked with interfering signal. One half of total number of records is made with female voice and another half – with male voice. Duration of preliminary articulation training of articulation team (speakers and auditors) 4 hours. Five listenings were done for every articulation table, every auditor analyzed 20 articulation tables 50 phrases each. Validity of phrase intelligibility calculation made 0.95 in this case.

2 Comparative analysis of speech intelligibility evaluation methods

Speech intelligibility represents integral estimation of speech signal and is defined as “the extent to which speech can be understood by the listeners”. Thus, it is the extent to which the listeners can understand the meaning of a phrase, identify words, syllables and phonemes. Accordingly, intelligibility is divided into different types: phonemic, syllabic, word and phrasal, which are interconnected and can be converted one into another. Due to the fact that the degree of predictability during phrase listening is higher than during listening to separate words or syllables, phrasal intelligibility is higher than word intelligibility, word – higher than syllabic, syllabic – higher than phonemic.

The following factors can be singled out from those multiple influencing speech intelligibility first of all: masking with other sounds including noise, reverberation, sound propagation path. Different expert methods and standards such as GOST 25902-83, GOST 51061-97, IEC 268-16, ANSI S3.2-1989, etc. are practically applied to determine speech intelligibility, in particular, during estimation of acoustic properties of lecture halls, theaters, concert halls, studios and other rooms. The following methods of speech intelligibility determination can be referred to as expert: attenuation equivalent method, selection tables method, articulation tables method (Mihailov, Zlatoustova, 1987). Attenuation equivalent method consists in measurement of sound intelligibility dependence on attenuation for tested and standard paths. The disadvantage of such a method is the necessity of standard path. Moreover, presence of amplifier in the path limits usage of this method. Intelligibility measurement according to selection tables consists in measurement of mistakes number during transmission of separate words from the group of phonetically similar words through the tested path. Low training requirements to operators and small measurement duration can be referred to method advantages. The method using articulation tables represents measurement of a relative number of correctly transmitted words, syllables and sounds through the tested path. Usage of syllable tables has a disadvantage connected to the fact that during syllables reading naturalness is lost to a significant extent as well as intonation overtone of oral speech. Limited number of word tables and their high memorability make their frequent usage quite undesirable. Phrase tables are almost not used due to their great excessiveness and consequently low sensitivity to distortions. Digital tables are used in cases of extremely low intelligibility.

Besides, three groups of objective methods of speech intelligibility determination can be singled out: formant, modulation, empirical (Gavrilenko, Didkovsky, Prodeus, 2007, pp. 54-65). The following formant methods can be pointed out among the foreign ones: Articulation Index (AI), Speech Intelligibility Index (SII). It is considered within the framework of AI version that speech intelligibility is proportional to average difference between peak level of speech and effective level of masking noise. AI improvement resulted in appearance of SII standardized into ANSI S3.5-1997. Formant methods to which the methods of Pokrovsky(1962), Bykov (1959), Sapozhkov(1963) can be referred are significantly based on foreign works. The methods that can be referred to modulation ones are STI (Speech Transmission Index), STIr (revised Speech Transmission Index), RASTI (Rapid Speech Transmission Index), STITEL (STI for telecommunication systems), STIPA (STI for Public Address). Usage of STI method makes it possible to take account of noise and reverberation interference simultaneously which is provided by a special choice of test signal in the form of noise with the spectrum identical with the specter of long lasting speech. This noise is modulated by periodic signal in each octave frequency band in such a manner that the envelope of momentary signal power would be sinusoid-shaped (Gavrilenko, Didkovsky, Prodeus, 2007, pp. 54-65). According to “full” version of STI method, also called STIr or STI-14, there are 98 values of STI index that are obtained for 14 values of modulation frequencies which are averaged thereafter by means of special methods. RASTI method represents cut version of STI method. Both STI and RASTI method allow taking account of reverberation interference. However, the account of background noise with irregular spectrum and of nonlinearity distortions is not taken correctly The most popular among empirical methods is %Alcons – the method of measurement of consonants articulation loss value expressed as a percentage (Sapozhkov, 1979). %Alcons method is widely used, especially in the USA, for approximate evaluation of speech intelligibility and reflects the loss of voiced consonants caused by indoor reverberation and sound absorption. It seems reasonable to use expert method for speech intelligibility evaluation when it is influenced by interfering signal while objective method will not allow finding the difference in speech intelligibility calculations caused by the type of interfering signal influencing speech. It is also reasonable to apply articulation phrase tables due to the fact that they are the closest to natural speech.

3 Development of an algorithm of speech-like interference synthesis for talks protection by means of speakers' allophones compilation

Nowadays the methods of speech-like signal synthesis are divided into the following categories: parametric methods (speech path model, formant model); compilation methods (microwave model, allophonic model). The authors have developed a method for the formation of speech-like interference in real time, based on the compilation synthesis. It can be concluded in consequence of the carried out analysis of speech-like interference (SLI) synthesis methods that the method of sections compilation of natural speech flow according to a phonemic text, which is generated by probability methods (Zelmansky, Davydov, 2010, p. 118), is the most efficient and flexible method. However, such an approach requires previously created bases of speech sections, for example, allophone bases. Process of such bases creation is quite labor-consuming. Besides, individual base must be created for every speaker and periodically updated, e.g. if the speaker catches cold or due to hoarseness, what makes common usage of the systems implementing this method more complicated. Nevertheless, it is SLI synthesis that is suggested. It is based on allophones compilation according to a phonemic text formed taking into account language statistics and allows generating SLI to protect talks in different languages, for example, in Russian, English or Arabic as well as in several languages at a time, what will make it possible to use suggested synthesis for speech data (SD) protection from leakage through technical channels when holding talks between multilingual participants (Zelmansky, Petrov, Al-Khatmi, Lynkov, 2011). The chart of suggested synthesis is shown in Figure 1.

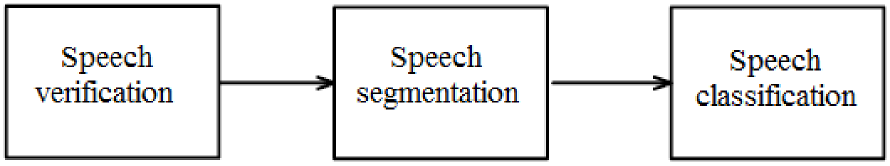


Figure 1: Chart of speech-like interference synthesis.

This synthesis supposes formation of phonemic texts in different languages according to statistical regularities of these languages and their further vocalization by means of compilation of allophones previously separated from the speakers', (native speakers of respective languages) speech, and allophone bases are created. It allows generating SLI for talks protection in one language or several languages at a time while SLI will present a mixture of several SLI in respective languages. Suggested synthesis uses speaker verification by speech to detect the moment of language change. Verification consists in recognizing the speaker and confirming his identity on the basis of unique voice properties. All the participants are registered before talks start and the samples of their speech, are included into database. It is individual characteristics as well as information about the language that each of the speakers will speak that are selected from their speech. Consequently, during verification process, the speaker's identity is established and proved, the language, the statistics of which will be used to create phonemic text, is determined, and allophone base, based on which SLI generation starts, is selected according to the language determined. However, this synthesis requires previously formed allophone bases of native speakers of Russian, English and Arabic, and it also requires statistical regularities of these languages. Thus, suggested synthesis allows masking the speech of multilingual participants of talks using SLI generation to protect talks in different languages.

3.2 Algorithm of automatic creation of allophone bases

Algorithm of automatic creation of allophone bases directly from the speech of multilingual talks participants consists in segmentation, which was discovered during speech detection for phonetic units and their allophonic classification according to minimum distance criterion, for speech-like interferences synthesis. Speech detection consists in discovering and separating speech from acoustic environment. The methods of speech detection consist in comparing classification parameters of the signal with respective threshold values, while tracking of statistical minimum of the given values is applied to determine threshold values. Parameters both in time and frequency domain are used as classification parameters. Time characteristics of the signal are short-term mean-square value of the signal and short-term function of average number of signal transmission through zero. And signal spectral estimation, amount of calculation of which decreases when fast Fourier transform (FFT) and symmetry properties of

Fourier transform are used, is a frequency characteristic. Verification of the speaker allows separating speech of one talk participant from speech of another one to create speaker's allophone base out of the speech uttered exactly by him what also allows dividing speech uttered in different languages. Besides, it can be established as a result of verification that there is previously created allophone base for this speaker already and there is no need to create a new one.

In the process of segmentation, speech separated during detection is divided into the sections of uniform vibrations corresponding to phonetic units by means of finding interphonemic transitions. It is the analysis of spectrum change function as a measure of correlation between consecutive windows of the signal analyzed that is used in the methods of speech segmentation. At the same time, distance between parameter vectors circumscribing signal windows are used as correlation measure. It is establish during phonetic units classification to which class of phonemes each of them belongs to. Speech classification methods consist in calculation of coefficient of difference between the analyzed phonetic unit and the base of samples of phoneme realization options, which is conditioned by certain phonetic environment of these phonemes represented by cepstral coefficients, on the basis of correlation matrix.

Decision-making about analyzed phonetic unit belonging to one or another group of phonemes and classification of this unit as a specific phoneme is carried out by finding a sample, duration of which is closest to the analyzed phonetic unit and to which corresponds the smallest value of difference coefficient. As a result, allophone base of the speaker of a certain language containing utterances of phonetic units and information about them is created.

3.3 Design of the system of speech-like interferences for protection of talks in different languages

Main task of SLI generation system is masking of confidential talk. Consequently, this system should be activated right at the moment the talk begins. Thus, speech detection module analyzing acoustic environment is the first system module. Main task of this module is determination of the moment the talk starts and launch of SLI generator. Only those sections of input signal that contain speech go to module input. Suggested system of SD protection should provide SLI generation on the basis of talks participants' voices in such a way so that it will be impossible to separate their speech and generated noise. To this end, noise is generated from the section of the uttered speech. This process is implemented in the modules of speech segmentation, speech classification and synthesis. Segmentation represents the process of speech division into sections of uniform vibrations corresponding to different types of phonemes: vowel-like, nasal, fricative, occlusive. Had speech been segmented into phonetic units, it is necessary to divide them into classes what is done in speech classification module. Classification task needs special classifying characteristics for its solution. Every section obtained is characterized by certain spectral-time parameters which can be divided into classes according to spectrum shape, energy change, periodicity, etc. There is no parameter which would allow accurately identifying all segment types. Each type possesses its own peculiarities that distinguish it from the second type, for example, but not from the third what requires usage of other peculiarities. Thus, it is required to use a set of parameters and solve classification task separately for each segment type. Direct synthesis of SLI generated randomly and according to formal features (splash nature, presence of words and intervals between them, frequency range), the most similar to speech signal (SS) but containing no semantic information is carried out by speech synthesis module. Main task of this module is conversion of a phonemic text, which is created taking into account statistical regularity of the chosen language, into acoustic vibrations of audio range of frequencies which come to the output of this module. Allophonic model of speech synthesizer and the bases of speaker's allophones are used to this end. It should be noted that SLI synthesis system contains the module of speaker verification by voice, which allows confirming speaker's identity on the basis of unique voice characteristics. Information on speaker's identity can be used to choose previously created base of this speaker's allophones and to use it later in speech synthesis module for SLI generation. Moreover, speaker verification is necessary for separation of one speaker's speech from that of another to create allophone bases for each of them and to separate speech in different languages. Module of speaker verification by voice (Zelmansky, Davydov,

Davydov, Lynkov, 2010) is connected to the output of speech detection module and provides confirmation of speaking person’s identity on the basis of his unique voice characteristics. Previously created allophone base corresponding to this speaker and representing a set of natural speech wave segments, which correspond to phonetic units of speaker’s speech, is chosen based on the results of procedure of speaker’s identity confirmation. In case if speaker’s identity is not confirmed, it is necessary to create a new allophone base for this speaker, and to this end speech is transmitted to speech segmentation module to divide it into phonetic units (allophones) which afterwards are classified in speech classification module and go to speaker’s allophone base. Structural flow-chart of the developed system of SLI synthesis (Zelmansky, Ganiyev, Kubankova, Munster, 2011, pp. 609-613) is shown in Figure 2. Blocks developed by the authors of this article are shown with dashed line.

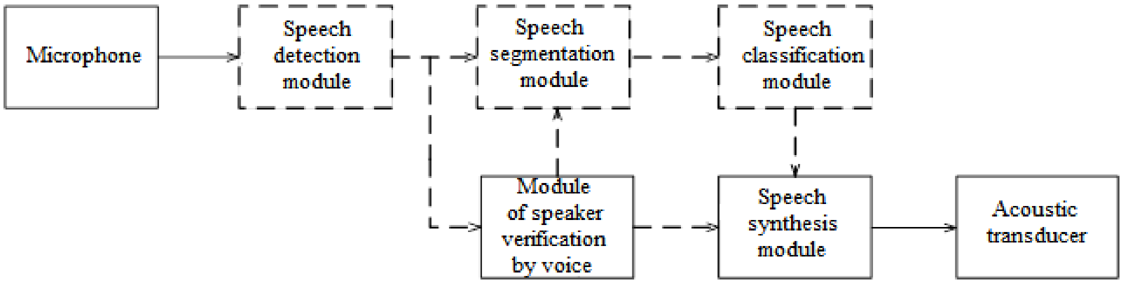

Microphone
Speech detection module
Speech segmentation module
Speech classification module
Module of speaker verification by voice
Speech synthesis module
Acoustic transducer

Figure 2: Structural flow-chart of SLI synthesis system.

Signal generated in such a way is overlaid on the speech of confidential talk participants. The obtained mixture of information and masking signals can be picked up by secret voice recorders or microphones and other intercepting devices but it will be impossible to restore the meaning of talk.

3.4 Results of speech detection module test

Test of the developed speech detection module was carried out using known speech corpus TIMIT initially intended for development and assessment of automatic speech recognition systems and using the bases of noise records Noisex-92 and Series 6000 The General Sound Effect Library

(40 CD)\SOUND IDEAS 6028 – Office. Acoustic-phonetic corpus TIMIT for American English consists of 2342 records of separate sentences of 630 speakers from 8 regional dialect areas of the USA. Correlation of the speakers makes about 70% of men-speakers and 30% of women-speakers. The bases of noise records contain background noises typical for office. Noise with signal-to-noise ratio of 10dB, 0 dB, 10 dB were overlaid on speakers' utterances records to test speech detection module. Error frequency of the first and the second kinds is determined using the following formulae:

$$\Pi = \frac{N_n}{N_{tn}} 100\,\%,$$

(3.1)

$$B = \frac{N_s}{N_{ts}} 100\,\%,$$

(3.2)

where Π – error frequency of the first kind;

N_n – number of non-speech sections taken for speech ones;

N_{tn} – total number of non-speech sections;

B – error frequency of the second kind;

N_s – number of speech sections taken for non-speech ones;

N_{ts} – total number of speech sections.

Thus, the developed speech detection module provides results sufficient for its application in SLI synthesis systems for separation of speech sections from audio signal analyzed.

3.5 Results of speech segmentation module test

Test of the developed speech segmentation module (Zelmansky, Davydov, 2012) was carried out using known speech corpus TIMIT initially intended for development and assessment of automatic speech recognition systems. Module test results are presented in Table 1.

Table 1: Results of speech segmentation module test

Number of excessive boundaries, %	Number of missed boundaries, %	Accuracy of boundaries identification, %
17.5	27.9	98.7

As it is seen from Table 1, the suggested speech segmentation module allows identifying up to 72% of boundaries between the phonemes with relative accuracy up to 98%. Herewith, number of excessive boundaries makes up to 17.5%. As a comparison, known algorithm of segmentation RASTA-SVF identifies 70.2 % of boundaries with accuracy of 95.5% (Petek, Andersen, Dalsgaard, 1996, pp. 913-916). Thus, the developed speech segmentation module provides the results sufficient for application of this module in SLI synthesis systems to create allophone bases of the speakers participating in protected talks.

3.6 Design of speech synthesis module

The task of speech synthesis module is SLI generation randomly and according to spectral and time parameters as well as auditory perception of the most similar SS not carrying semantic load. While the module is based on the method of phonemic text synthesis and allophonic model of speech synthesizer, module operation consists in conversion of phonemic text formed taking into account statistical regularities into acoustic vibrations of frequencies audio range. Allophone base is sent to module input and SLI synthesized in the form of audio file is generated at the output. Structural scheme of speech synthesis module is similar to the structure of allophonic system of SS synthesis. Nevertheless, functional use of structural units is different.

1. Text formation unit of speech synthesis module forms “words”, which are united into sytagmatictic, phrases and phono-passages, on the basis of allophones in compliance with statistical regularities, while oral speech represents sound flow with the following elements hierarchy: phoneme (allophone) – syllable – word – sytagmatictic – phrase – phono-passage, according to (Lobanov, 2000, pp. 57-76). Thus, every element placed higher in the hierarchy of speech elements is composed of the elements placed below, the number of which corresponds to a certain

probability distribution law. Phonemic text formed taking into account statistical regularities goes to the output of text formation unit. It should be noted that, unlike speech synthesis module, the tasks of this unit in allophonic system of SS synthesis according to the text include text division into phono-passage, phrase and sytagmatictic.

2. Database of Russian, Arabic and English grammar dictionaries of speech synthesis module contains statistical data used for phonemic text formation while database of allophonic synthesis system includes an array of all the words possible. Statistical data of Russian, Arabic and English speech elements are presented in Table 2. Sytagmatictic, phrases and phono-passages are obligatory divided with pauses. Pause duration is determined by average statistical values. Average statistical values of pauses for Russian, Arabic and English are presented in Table 3.

Table 2: Statistical characteristics of Russian, Arabic and English speech elements

Number of speech elements	Occurrence probability, %								
	Words in a sytagmatictic			Sytagmatics in a phrase			Phrases in a phonopassage		
	Russian lang.	Arabic lang.	English lang.	Russian lang.	Arabic lang.	English lang.	Russian lang.	Arabic lang.	English lang.
1	5	5	4	5	5	4	5	5	4
2	10	10	12	10	10	12	10	10	12
3	30	32	21	30	32	21	15	17	22
4	25	27	19	25	27	19	30	30	21
5	15	15	22	15	15	22	25	27	19
6	10	11	12	10	11	12	10	11	12
7	5	-	10	5	-	10	5	-	10

Table 3: Average statistical values of pauses

	Phonopassage	Phrase	Sytagmatic
Duration for the Russian language, s	1,5	0,8	0,3
Duration for the Arabic language, s	1,5	0,8	0,3
Duration for the English language, s	1,6	1,2	0,5

3. The tasks of fundamental pitch formation unit include determination of word FPF depending on its place in a phono-passage, phrase, syntagmaticwhat is conditioned by the necessity of application of narrative style of text enunciation during SLI synthesis. Statistical regularities of

syllable number in a word for Russian, Arabic and English are presented in Table 4.

Table 4: Distribution of Russian, Arabic and English word length

Occurrence probability	Syllable number in a word								
	1	2	3	4	5	6	7	8	9
For Russian, %	16,31	36,67	25,52	13,60	5,16	1,40	0,29	0,034	0,004
For Arabic, %	31,63	42,75	19,82	4,87	-	-	-	-	-
For English, %	71,52	19,40	6,80	1,60	0,56	0,12	-	-	-

Word length influences stress position, regularity of which is determined basing on statistical data obtained experimentally and contained in Tables 5, 6.Distribution of probability of stress position in English word according to (Alcantara, 1998, 356 p.) is presented in Table 7.

4. Allophone separation unit determines which allophones will be needed for its conversion into acoustic vibrations by means of analysis of a phonemic text obtained in text formation unit. Thus, this unit carries out text markup of a phonemic text and generation of positional and combinatory allophones.

Table 5: Distribution of probability of stress position in Russian word

Syllable number in a word	Probability of stress on n-syllable, %								
	1	2	3	4	5	6	7	8	9
1	100,0								
2	49,83	50,17							
3	27,90	48,64	23,45						
4	7,07	43,31	44,72	4,89					
5	2,83	19,80	59,40	16,89	1,09				
6	0,32	9,11	33,00	49,42	7,81	0,33			
7	0,11	4,45	17,15	34,82	36,73	6,68	0,15		
8	0,01	2,28	9,13	21,51	39,09	25,15	2,83	0,00	
9	0,00	0,00	7,55	3,93	28,99	47,70	11,47	0,33	0,00

Table 6: Distribution of probability of stress position in Arabic word

Syllable number in a word	Probability of stress on n-syllable, %			
	1	2	3	4
1	100,0			
2	57,1	42,9		
3	31,5	37	31,5	
4	20,8	20,8	20,8	37,6

Table 7: Distribution of probability of stress position in English word

Syllable number in a word	Probability of stress on n-syllable, %				
	1	2	3	4	5
1	100,0				
2	75	25			
3	51	30	18		
4	91	5	4	0	
5	89	5	3	1	2

5. Acoustic unit synthesizes SLI by means of compilation of sections of natural sound waves of respective allophones stored in the base of allophone sound waves when it receives data from the units of allophone separation and fundamental pitch formation on what allophones and with which prosodic characteristics are needed to be synthesized to “vocalize” the formed phonemic text.

6. Database of allophone sound waves provides compliance of spectral SLI parameters with SS parameters while it is created out of real speakers’ SS. Database capacity influences the quality of natural SS simulation and it should be within the limits of 400-1600 allophones for the Russian language.

Application of conditional probabilities of the syllables allows taking into account phonetic peculiarities of speech uttered in Russian, Arabic and English. In case of using unconditional probabilities of syllable occurrence the results of phonemic text formation appear to be worse than in case of using unconditional probabilities of initial, pre-stressed, stressed and final syllables or conditional probabilities of the syllables that demonstrate almost the same results.

4 Study of interconnection of signal-to-noise ratio and speech intelligibility

Direct acoustic and vibration channels of speech data leakage were studied.During the working process, the comparison of speech signal intelligibility in the conditions of influence by different types of interference was made. To this end, the recorded articulation phrase tables were played by the acoustic system and were recorded at certain points of the room. Work on the stage was carried out in sound-proof room using the following equipment: microphone AKG P120; electronic stethoscope, audio interface Focusrite Scarlett 2i2; acoustic system Edifier R1900 T3;laptops with Sound Forge 9.0 software installed;acoustic noise analyzer MANOM-4/2;microphone preamplifier VPM-101;condenser microphone capsule M-101;acoustic transducers included into scope of delivery of speech protection device "Priboi". White noise and speech-like interference with signal-to-noise ratio from -5dB to -40dB were used as interfering signal. During study of direct acoustic channel of speech data leakage, a mixture of test and interfering signals was recorded in different points of room space. Choice of such points is conditioned, firstly, by nearness to building envelope components with minimum values of their own sound proofing; secondly, by room geometry. Standing waves create a set of peaks and drops in the room and at the same time volume in certain areas can be higher than that of reproduced by the source. Equipment placement scheme is shown in Figure 3 where acoustic system 1 is the system for test signal reproduction, acoustic system 2 is the system for interfering signal reproduction. Test and interfering signals are reproduced from PC1 and PC2, respectively. Recording of test and interfering signals mixture is carried out at PC3. Noise Analyzer is used to determine the level of test and interfering signals.

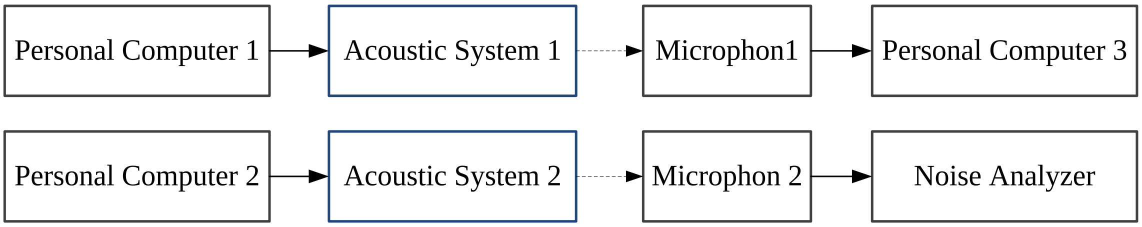


Figure 3: Equipment placement scheme (direct acoustic channel)

A set of noisy test signals was presented to each of the auditors.

A phrase is considered as incorrectly received if at least one word was received incorrectly, omitted or added. Phrase intelligibility *J* is found as average value for measurement cycle according to the formula:

$$J = \frac{1}{N}\sum_{i=1}^{N} Ji \qquad (1)$$

where:

J – phrase intelligibility at normal utterance speed, %;

J_i– single measurement result, %; is calculated as a percentage of correctly received phrases;

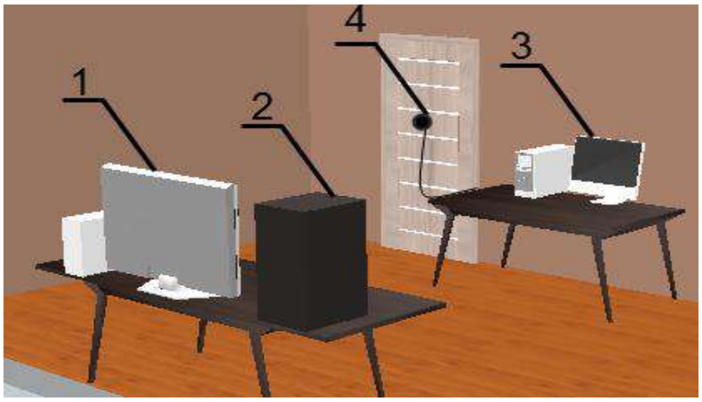
N – number of single measurements.

The results of phrasal speech intelligibility calculation using white and interfering noises as an interference for the indicated signal-to-noise ratios (q) in case of spatial noise masking (direct acoustic leakage channel) are presented in Table 8.

Table 8: Results of phrasal speech intelligibility calculation

Signal-to-noise ratio, *q*,dB	Phrase intelligibility, *J*, %	
	White noise	Speech-like noise
-15	89	87
-20	69	56
-25	18	10

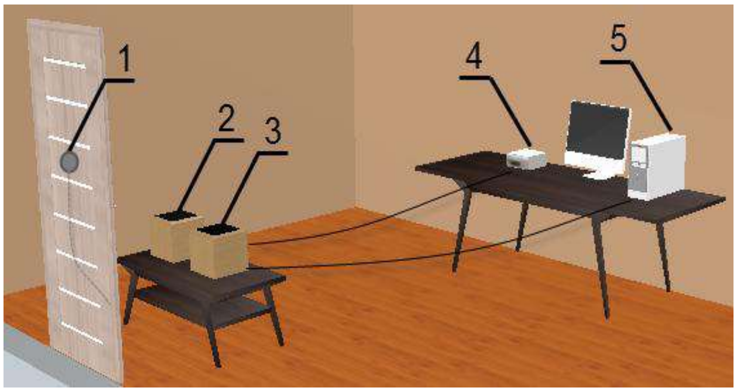
Study of vibration channel of speech data leakage was carried out in compliance with the scheme presented in Figures 4 and 5 in two rooms divided with a partition with a door. Vibration oscillator was fixed on the door dividing the rooms in different positions for interfering signal reproduction.



1, 2 – PCand acoustic system for test speech signal reproduction;
3 – PC for interfering signal reproduction; 4 – electronic stethoscope

Figure 4: Scheme of equipment placement in the room for test signal reproduction

The results of phrasal speech intelligibility calculation using white and speech-like interferences as an interference for the indicated signal-to-noise ratios (q) in case of vibration channel of speech data leakage are presented in Table 9.



1 – projection of vibration transducer, 2 – microphone for signals level control, 3 – microphone for noisy test signal recording; 4 – noise analyzer; 5 – PC.

Figure 5: Scheme of equipment placement in the room for control and recording of noisy test signal

Table 9: Speech intelligibility values (vibration channel)

Signal-to-noise ratio, *q*,dB	Phrase intelligibility, *J, %*	
	White noise	Speech-like interference
-15	60	48
-20	25	14
-25	13	4

5 The analysis of experimental data

The result of speech intelligibility evaluation by the auditors are shown in Figure 6.

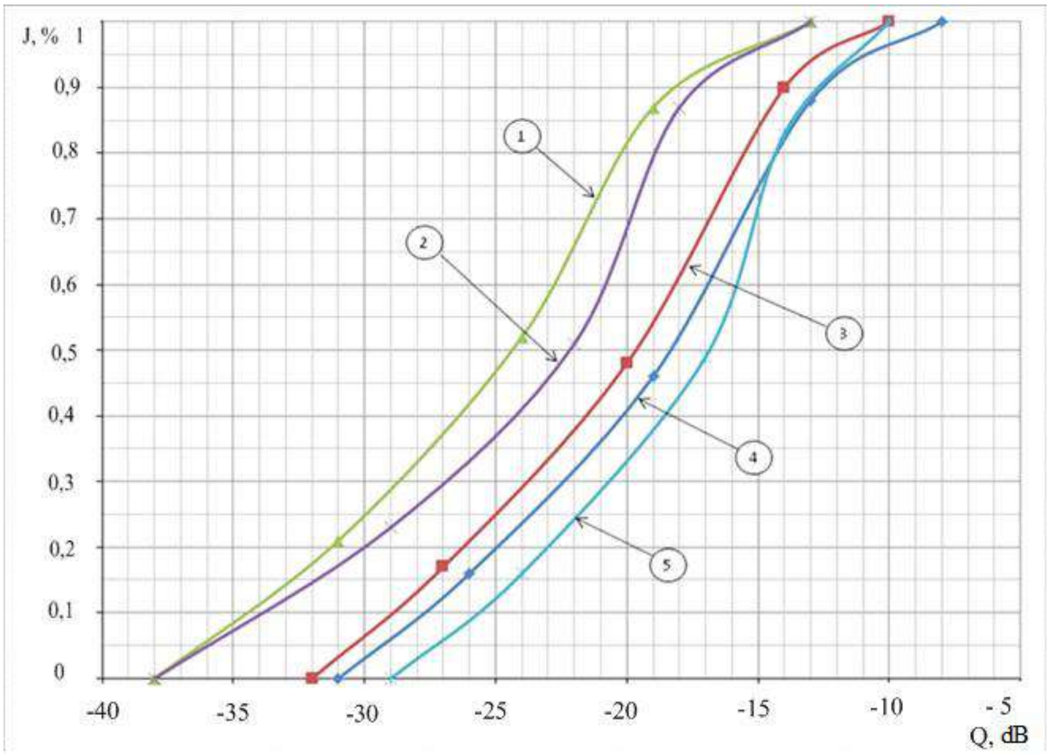


Figure 6: Dependence of phrase intelligibility J on integral signal-to-noise ratio in frequency band 63 ...8000 Hz

Line 1 shows efficiency estimate of non-correlated speech-like interference overlaid on male voice, line 2 shows the estimate of non-correlated speech-like interference overlaid on female voice, line 3 shows the estimate of white noise type interference overlaid on male voice, line 4 shows the estimate of white noise type interference overlaid on female voice, line 5 correlated speech-like interference overlaid on male voice. It follows from this data that correlated speech-like interference is the most effective.

Thus, it can be concluded on the basis of earlier done studies that experimental data obtained specify the results of known studies and allow suggesting more efficient abstract in Arabic obtained by automatic Google translation abstract in Arabic obtained by automatic Google translation masking signal representing correlated speech-like interference for the systems of active protection of acoustic information. The measurements are taken for different signal-to-noise ratios and different types of interfering signals. Efficiency comparison of different interfering signal types is carried out comparing speech intelligibility value at given signal-to-noise ratio.

Results

Speech intelligibility value was chosen to be protection criterion as a result of literature sources analysis. Method of articulation measurements was chosen for intelligibility determination. Articulation team consisting of five men and five women was created. Test acoustic signals were recorded.

Comparative analysis of expert and objective methods of speech intelligibility determination was carried out. The algorithm of speech-like interference synthesis, which is based on compilation of the segments of speakers' utterance records according to a phonemic text taking into account language statistical peculiarities, was developed. Evaluation of speech intelligibility under the influence of interfering signal was done. White noise as well as speech-like signal generated in several ways was used as interfering signal. It was shown that speech-like correlated interference is the most efficient.

References

Alcantara, J.B., The architecture of the English lexicon, [NY], Dissertation for doctor of philosophy, 1998

Bykov, Y.S. (1959), Theory of speech intelligibility and efficiency improvement of radiotelephonic communication, [Moskva – Leningrad: Gosenergoizdat], 1959

Gavrilenko, A.V., Didkovsky, V.S., Prodeus, A.N., Comparative analysis of some methods of speech intelligibility evaluation, [Kiev, 25–27 july], Acoustic symposium “Consonance-2007”, 2007

Lobanov, B.M. Synthesis of speech on the text, [Minsk], 4th International School-Seminar on Artificial Intelligence, 2000

Mihailov, V.G., Zlatoustova, L.V., Speech parameters measurement, [Moskva: Radio isvyaz], edited by Sapozhkov, M.A., 1987

Petek B., Andersen O., Dalsgaard P. (1996), On the robust automatic segmentation of spontaneous speech, Proc. ICSLP, [Philadelphia, USA], 1996

Pokrovsky, N.B., Speech intelligibility calculation and measurement, [Moskva: Svyazizdat], 1962

Sapozhkov, M.A., Speech signal in cybernetics and communication, [Moskva: Svyazizdat], 1963

Sapozhkov, M.A., Installation of public address systems in premises, [Moskva: Svyaz], 1979

Zelmansky O.B., Davydov A.G., Speech analysis device, [Minsk], Patent 8194 Republic of Belarus, 2012

Zelmansky O.B., Ganiyev A., Kubankova A., Munster P., Synthesis of speechlike signals for masking acoustic information, [Ostrava, September 8-9],WOFEX 2011, 2011

Zelmansky O.B., Davydov A.G., Davydov G.V., Lynkov L.M., Speaker automatic recognition device,[Minsk], Patent 6229 Republic of Belarus, 2010

Zelmansky O.B., Petrov S.N., Al-Khatmi M.O., Lynkov L.M.,Active and passive methods and means of protecting information from leakage through technical channels,[Minsk: Bestprint], 2011

Zelmansky, O.B., Davydov A.G., The use of speech-like signals in active acoustic masking systems. [Minsk, September 28-30], Modern means of communication, 2010

Zhelezniak, V.K., Makarov, Y.K., Horev, A.A., Some of methodological approaches to evaluation of speech data protection efficiency.No.4., [Moskva],Special technologies, 2000