The syllable intelligibility in the system of information transmission by speech signals depending on the intensity of acoustic noise

Yu A Kropotov[1], A A Belov[1], A A Kolpakov[1] and A Yu Proskuryakov[1]

[1]Murom Institute (branch) «Vladimir State University named after Alexander and Nicholay Stoletovs», Orlovskaya street, 23, Murom, Vladimir Region, Russia, 602264

e-mail: kaf-eivt@yandex.ru

Abstract. The paper investigates the effect of the signal-to-noise ratio on syllable intelligibility under the intense influence of external acoustic interference when exchanging voice messages in telecommunication systems of public address systems. The article discusses the effect on the syllable intelligibility of the signal / external acoustic noise ratio, examines the effect of the integral articulation index, the dependence of the perception coefficient of formants on the relative level of formant intensity, the dependence of the formant parameter on the geometric mean frequency of the i-th spectrum of the speech signal. In accordance with the results of studies of the integral articulation index depending on the signal-to-noise ratio, a function of syllable intelligibility depending on the signal-to-noise ratio was obtained, using which it is possible to determine the maximum value of the output signal-to-noise ratio in the audio exchange telecommunications system to obtain a given syllable intelligibility. At the same time, experimentally determined the value of the signal-to-noise ratio in the telecommunications system of audio exchange to obtain a syllable intelligibility of at least 93% for ensure full perception of the transmitted speech information.

1. Introduction
As it is known, the main criterion of efficiency of the system of telecommunication exchange of the speech information is syllabic legibility S % or the size of an estimation of a speech signal on scale MOS (Mean Opinion Score) [1].

Telecommunication systems of audio exchange, in particular loudspeaker systems, are considered to be effective if the transmitted speech information is perceived by the object completely and without difficulties, the syllable intelligibility in this case is not less than 93% [1,2,4] or the MOS score should be not less than 3,9 points on a five-point scale [5, 6].

2. Formulation of the problem
Dependence of syllabic intelligibility in the system of telecommunication exchange of speech information on the influence of various factors has been studied in a number of works [1,3]. However, the information in the known sources [9, 10] about the influence of the signal-to-noise ratio on the syllabic legibility on the side of receiving speech messages for the case of operational-command

telecommunication systems is insufficient, so this article considers the problem of determining the influence of the signal-to-noise ratio on the syllabic legibility in telecommunication audio exchange systems.

The known results of the studies of the assessment of syllabic legibility by the instrumental-calculation method are shown in Fig. 1 [1, 3].

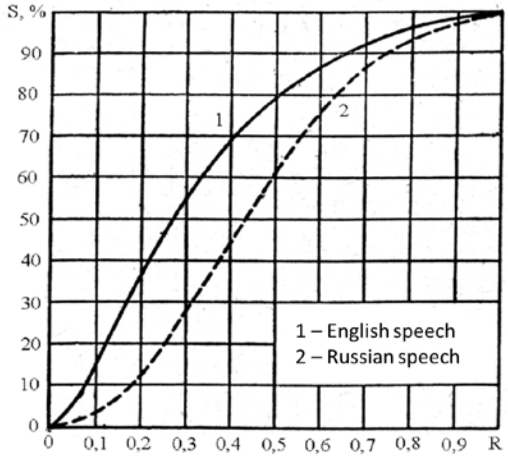

Figure 1. Dependence of syllabic legibility S on the integral articulation index *R.*

Figure 1 shows that the syllabic legibility here is represented by the dependence only on the value of the integral articulation index.

3. Instrumental-calculation method for estimating the integral articulation index and syllabic legibility

The value of the integral articulation index R depending on the value of the spectral articulation index R_i is determined by the expression

$$R = \sum_{i=1}^{N} R_i \ . \qquad (1)$$

The articulation spectral index is calculated by the expression

$$R_i = p_i \cdot k_i \ , \qquad (2)$$

where p_i is formant coefficient, k_i is weighting coefficient of the presence of formant speech in the i-th band.

The coefficient of perception of formant p_i is calculated using the expression [3]

$$p_i = \begin{cases} \dfrac{0{,}78 + 5{,}46 \cdot \exp\left[-4{,}3 \cdot 10^{-3} \cdot \left(27{,}3 - |Q_i|\right)^2\right]}{1 + 10^{0{,}1 \cdot |Q|}}, & \text{if} \quad Q_i \le 0; \\ 1 - \dfrac{0{,}78 + 5{,}46 \cdot \exp\left[-4{,}3 \cdot 10^{-3} \cdot \left(27{,}3 - |Q_i|\right)^2\right]}{1 + 10^{0{,}1 \cdot |Q|}}, & \text{if} \quad Q_i > 0, \end{cases} \qquad (3)$$

where $Q_i = q_i - \Delta A_i$ is the relative intensity level of the format.

Or the value of the perception coefficient p_i format can be determined by the graph in Figure 2.

The format parameter ΔA_i is determined by the graph in Figure 3 or by the expression

$$\Delta A(f) = \begin{cases} 200 / f^{0{,}43} - 0.37, & \text{если } f \le 1000 \text{ Гц}, \\ 1{,}37 + 1000 / f^{0{,}69}, & \text{если } f > 1000 \text{ Гц}, \end{cases} \qquad (4)$$

where $f_{\text{ср}.i} = \sqrt{f_{в_i} - f_{н_i}}$ is average geometric frequency, $f_{н_i}$ is lower frequency of the i-th bandwidth of the speech spectrum, $f_{в_i}$ is upper frequency of the i-th bandwidth of the spectrum.

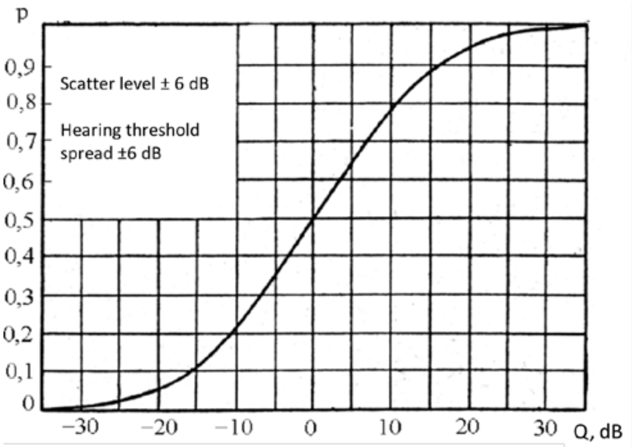

Figure 2. Dependence of the perception coefficient format p_i on the relative intensity level format Q_i.

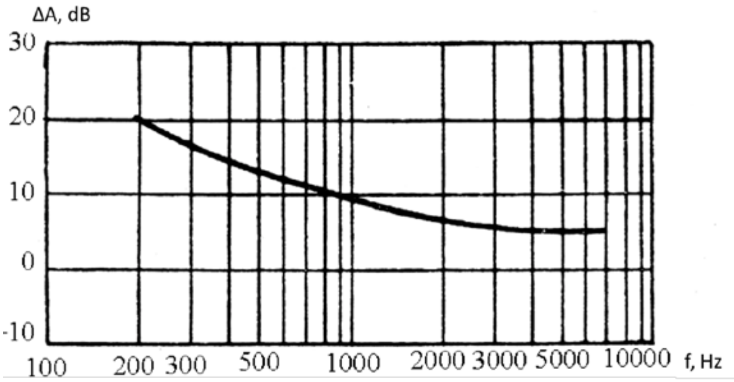

Figure 3. Dependence of the formal parameter ΔA_i on the average geometric frequency of the i-th bandwidth of the speech spectrum.

For each i-th (i=1, 2, ... N) frequency band at the average geometric frequency $f_{\text{ср}.i} = \sqrt{f_{в_i} - f_{н_i}}$, a formal parameter ΔA_i is determined, characterizing the energy redundancy of discrete components of the speech signal.

4. Results of experiments

Let's take the number of octave bands N=5. Values of the accepted limits by frequency of octave bands, values of calculated $f_{sr.i}$ and values of formal parameters ΔA_i are given in Table 1.

Table 1. Values of the accepted limits by frequency of octave bands, values of calculated $f_{cp.i}$ and values of formal parameters ΔA_i.

i	Boundary bandwidth frequency, Hz	$f_{cp.i}$, Hz	Bandwidth Δf_i	$\Delta A_i\,(f_{cp.i})$
1	180-355	250	175	18
2	355-710	500	355	13
3	710-1400	1000	690	9.5
4	1400-2800	2000	1400	7.5
5	2800-5600	4000	2800	5.5

With the help of expression $Q_i = q_i - \Delta A_i$, we determined the values of intensity levels of format Q_i depending on the signal to noise ratio q_i. The calculated values of Q_i are summarized in Table 2.

With the help of expression (3) or according to the diagram in Figure 2, the formatting factor p_i is determined depending on Q_i for i-th bands, with different values of signal-to-noise ratio, dB. The calculated p_i values for different q_i are summarized in Table 3.

Table 2. Calculated values Q_i.

$Q_i = q_i - \Delta A_i$ q_i, дБ	$Q1$	$Q2$	$Q3$	$Q4$	$Q5$
q_i = 0 дБ	-18	-13	-9.5	-7.5	-5.5
q_i = 3 дБ	-15	-10	-6.5	-4.5	-2.5
q_i = 6 дБ	-12	-7	-3.5	-1.5	+0.5
q_i = 10 дБ	-8	-3	+0.5	+2.5	+4.5
q_i = 20 дБ	+2	+7	+10.5	+12.5	+14.5
q_i = 30 дБ	+12	+17	+20.5	+22.5	+24.5

Table 3. Calculated values p_i.

q_i, dB	p_1	p_2	p_3	p_4	p_5
q_i = 0	0.07	0.18	2.2	0.29	0.34
q_i = 3	0.11	0.21	0.31	0.38	0.41
q_i = 6	0.2	0.3	0.40	0.48	0.51
q_i = 10	0.24	0.41	0.51	0.53	0.62
q_i = 20	0.47	0.6	0.79	0.81	0.88
q_i = 30	0.81	0.9	0.94	0.96	0.98

To determine the spectral articulation index R_i, the weighting coefficient of probability of presence of the format of speech in the i-th band k_i is determined by the formula

$$k_i = k(f_{vi}) - k(f_{ni}),$$

where $k(f_{vi})$ and $k(f_{ni})$ is weights for the upper f_{vi} and lower f_{ni} of the i-th bandwidth of the speech spectrum. Values $k(f_{vi})$ and $k(f_{ni})$ are determined by the formula as [3]:

$$k(f) = \begin{cases} 2{,}57 \cdot 10^{-8} \cdot f^{2,4}, & \text{если } 100 < f \le 400 \text{ Гц}; \\ 1 - 1{,}074 \cdot exp\left(-10^{-4} \cdot f^{1,18}\right), & \text{если } 400 < f \le 10000 \text{ Гц}; \end{cases} \quad (5)$$

or according to the chart in Figure 4.

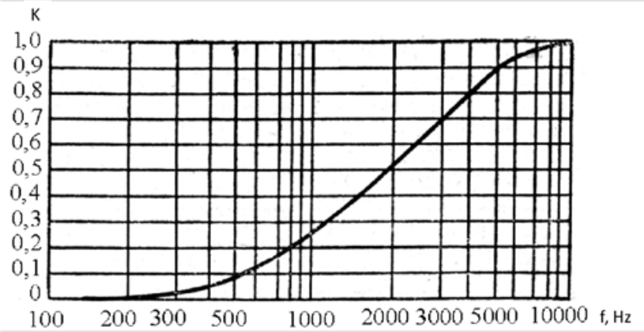

Figure 4. Weighting $k(f)$ of the formal frequency distribution.

The results of calculations of the weighting coefficients of probability of formant speech in the i-th band are presented in Table 4.

Table 4. Results of calculations of the weighted probability coefficients of formal speech in the i-th band.

k_i	$k(f_в)$-$k(f_н)$	$k(f_{в\,i})$-$k(f_{н\,i})$	k_i
k_1	k(355)-k(180)	0.04-0.01	0.03
k_2	k(710)-k(355)	0.18-0.04	0.14
k_3	k(1400)-k(710)	0.36-0.18	0.18
k_4	k(2800)-k(1400)	0.63-0.36	0.37
k_5	k(5600)-k(2800)	0.92-0.63	0.29

Calculation of the R_i articulation spectral index is performed by formula (2). Calculations of R_i, at different values of signal-to-noise ratio are summarized in Table 5.

Table 5. Articulation spectral index R_i, at different signal-to-noise ratios.

$R_i = p_i \cdot k_i$	$R_1 = p_1 \cdot k_1$	$R_2 = p_2 \cdot k_2$	$R_3 = p_3 \cdot k_3$	$R_4 = p_4 \cdot k_4$	$R_5 = p_5 \cdot k_5$
q_i = 0 dB	0,07·0,03=0,0021	0,18·0,014=0,0252	0,22·0,18=0,0396	0,29·0,37=0,1073	0,34·0,29=0,0986
q_i = 3 dB	0,11·0,03=0,0033	0,021·0,14=0,0294	0,31·0,18=0,0558	0,38·0,37=0,1406	0,41·0,29=0,1189
q_i = 6 dB	0,2·0,03=0,006	0,3·0,14=0,042	0,4·0,18=0,072	0,48·0,37=0,1776	0,51·0,29=0,1479
q_i = 10 dB	0,24·0,03=0,0072	0,41·0,14=0,0574	0,51·0,18=0,0918	0,53·0,37=0,1961	0,62·0,29=0,1798
q_i = 20 dB	0,47·0,03=0,0141	0,6·0,14=0,084	0,79·0,18=0,0918	0,81·0,37=0,2997	0,88·0,29=0,2552
q_i = 30 dB	0,81·0,03=0,0243	0,9·0,14=0,126	0,94·0,18=0,1692	0,96·0,37=0,3552	0,98·0,29=0,2842

According to the results of calculations of the spectral articulation index R_i, summarized in Table 5, it became possible to calculate the integral articulation index depending on the signal-to-noise ratio. The results of the calculation of the integral articulation index made it possible to find the values of syllabic legibility depending on the signal-to-noise ratio, which are summarized in Table 6.

Table 6. Syllable intelligibility values depending on signal-to-noise ratio.

q_i signal/noise	0 dB	3 dB	6 dB	10 dB	20 dB	30 dB
R	0.273	0.348	0.4455	0.5323	0.7952	0.9589
$S_{eng.}$	48 %	61 %	73 %	82%	96%	99%
$S_{rus.}$	25%	35 %	53 %	65 %	93%	98.5%

The graph of the syllable intelligibility function S from the signal-to-noise ratio is shown in Figure 5.

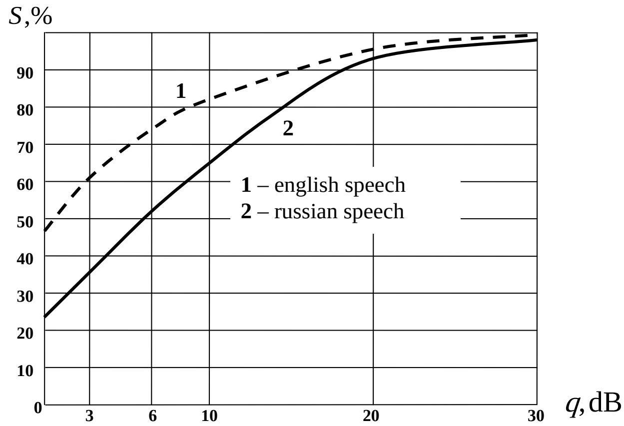

Figure 5. Graph of dependence of syllabic intelligibility on signal-to-noise ratio.

5. Conclusions

As can be seen from the graphs in Figure 5, the syllable intelligibility of the voice messaging telecommunications system is ensured by $S \geq 93\%$ for signal/noise ratio $q \geq 20$ dB [7, 8]. Thus, the dependence of syllabic intelligibility on signal-to-noise ratio, which is important for the practice of telecommunication systems, is obtained. It shows that for effective transmission of speech information by the command and control system of telecommunications, for obtaining, respectively, syllabic intelligibility of $S \geq 93\%$, in the system for transmission of speech messages, it is necessary to provide signal-to-noise ratio $q \geq 20$ dB on the receiving side of messages.

6. References

[1] Sapozhkov M A 1962 *Speech signal in cybernetics and communications* (Moscow: Svyazizdat) p 452

[2] GOST R 50840-95 *Speech transmission via communication channels. Methods to assess quality, legibility and recognizability*
[3] Zheleznyak V K, Makarov Y K and Khoreev A A 2000 Some methodical approaches toevaluation of efficiency of speech information protection *Special technique* **4** 39-45
[4] Cohen I, Benesty J and Gannot S 2010 *Speech processing in modern communication* (Berlin, Heidelberg: Springer) p 342
[5] Hansler E and Schmidt G 2006 *Topics in acoustic echo and noise control: Selected methods for the cancelation of acoustic echoes, the reduction of background noise, and speech processing* (Berlin, Heidelberg: Springer) p 642
[6] Kahrs M and Brandenburg K 2002 *Applications of digital signal processing to audio and acoustics* (New York: Kluwer Academic Publisher) p 572
[7] Kropotov Y A and Belov A A 2016 Application method of barrier functions in the problem of estimating the probability density of the parameterized approximations *13th International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering* 69-72
[8] Kolpakov A A and Kropotov Y A 2017 Advanced mixing audio streams for heterogeneous computer systems in telecommunications *CEUR Workshop Proceedings* **1902** 32-36
[9] Ryabenkyi V S 2012 Mathematical model of the external noise suppression devices in the subarea of space *Mathematical modeling* **24(8)** 3-31
[10] McAulay R and Malpass M 1980 Speech enhancement using a soft-decision noise suppression filter *IEEE Trans, on Acoustics, Speech, and Signal Processing* **28(2)** 137-145
[11] Kropotov Y A, Belov A A and Proskuryakov A Y 2018 Method for forecasting changes in time series parameters in digital information management systems *Computer Optics* **42(6)** 1093-1100 DOI: 10.18287/2412-6179-2018-42-6-1093-1100