International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I6P208
Volume 69 Issue 6, 50-55, June 2021

A Priori Assessment of The Intelligibility of Stereophonic Sound of Speech

Natalia Derkach[#1], Olena Pavelko[*2], Svitlana Luniova[#3]

[1#]*student* [2*]*student* [3#]*associate professor*

Faculty of Electronics, The Department of Acoustic and Multimedia Electronic Systems

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Kyiv, Ukraine

[1] natashadirkach@gmail.com, [3] svetlana_lunyova@yahoo.com

Abstract *— The article proposes an objective method of calculating the intelligibility of speech in the room based on the assessment of the values of the interaural correlation coefficients in the listening position. The method can serve as an a priori assessment of language intelligibility already at the design stage of the room, as well as be used for stereo and monophonic sound sources.*

Calculations are performed to emit rectangular and sawtooth video pulses. Comparison of the obtained results with the results of articulation tests in the studied hall testifies to the expediency of using a sawtooth video pulse for analysis. On the basis of the established scale of correspondence of articulatory intelligibility of language and coefficients of interaural correlation, the intelligibility of language indoors is defined.

Keywords *— speech intelligibility, methods of assessing speech intelligibility, correlation coefficient, localization of sound, stereo sound.*

I. INTRODUCTION

In language rooms (lecture halls, conference halls, drama theaters, cinemas, and other halls where the language is heard), language intelligibility is the main acoustic characteristic of the room. Assessing the legibility of language in the premises where information messages are announced or artistic language is heard today is an urgent task of architectural acoustics.

The legibility of language in the room is influenced by a number of factors, including the size, shape, and acoustic decoration of the room, the number and location of listeners, and sound sources [1]. Based on these factors, each room requires an individual assessment of language intelligibility.

The results of articulation tests in the hall are considered to be the most reliable. But such measurements require qualified speakers, a significant number of expert listeners, as well as the availability of test material in the form of standardized tables [2].

To simplify the procedure, preference is given to objective measurement methods without the involvement of listeners [3-7], which mainly use a broadband signal. However, studies performed in [8] suggest that the use of a broadband signal to assess speech intelligibility may not be sufficient. Comparison of the results of the intelligibility component of the language, obtained on the basis of measurements in the hall of the C50 accuracy index, with the results of the articulation components of the tests leads to non-compliance with the intelligibility classes of the language (according to GOST R 50840-95).

The main reason for such differences is what is called "coherence" of language. The share of pauses in the language is insignificant in time, but their duration and location provide meaningful expression. The sound during pauses is significantly influenced by the structure of speech organs, which contributes to the isolation of phonetic units [3].

As a result of the research, the authors came to the conclusion that it is necessary to use a speech signal [9] and a binaural model of perception [10]. In the article [9], the authors proposed a method of measuring speech intelligibility using an artificial head based on the interaural correlation coefficients for the speech signal.

These methods are methods for assessing the legibility of language in the built premises. The most interesting is the a priori assessment of language intelligibility at the design stage of the room when it is still possible to make adjustments.

The authors of the article propose a method of a priori assessment of speech intelligibility in listening places by calculating the interaural correlation coefficients of the sawtooth pulse signal.

II. PURPOSE AND OBJECTIVES OF THE WORK

The aim of the work is to develop an algorithm for a priori assessment of speech intelligibility in the hall by calculating the interaural correlation coefficients of the binaural signal pair at the listening positions.

The evaluation is performed on the basis of calculations of mutual correlation functions of rectangular and sawtooth video pulses at control points of the room.

Verification of the results is performed by comparing the calculated data with the data of measurements of correlation functions of pulse and speech signals obtained with the help of an artificial head [9], as well as the results of component articulation tests conducted in the hall [8].

The premises of the conference hall of the Faculty of Electronics of the National Technical University of Ukraine, "Kyiv Polytechnic Institute named after Igor Sikorsky," were selected for analysis (Fig. 1).

The dimensions of the hall are 14.3x18.5 x 6.25 m, the volume of the hall is. The lifting height from the first row to the last is 0.9 m.

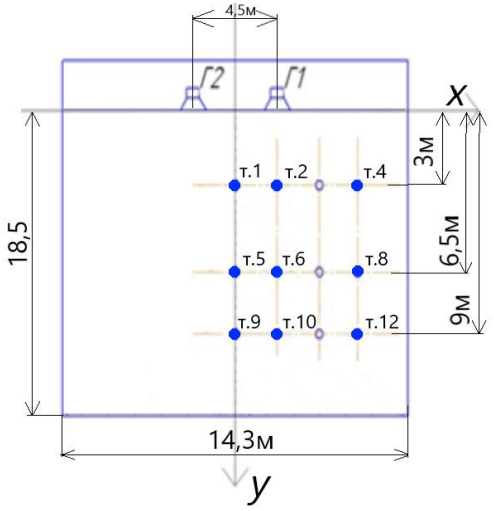


Fig. 1 Geometry of the stereo system and measurement points in the hall

Measurement points were selected over the entire area of the room (12 measurement points located in three rows of listeners are shown in Fig.1). The width of the base of the stereo system - 4.5 meters. The distance from the baseline to the first row of measurements is 3 m, to the second row is 6.5 m, to the third row is 9 m. The distances between the measurement points horizontally were: 0; 2.23; 4.5; 7 m, respectively. The height of the artificial head is 1.2 m from the floor level of the row. The stereo system is placed on the stage at the height of 1.2 m from the floor.

The room is selected as indicated in [8,9] for comparison of results.

III. METHODS OF CALCULATIONS

The method of assessing speech intelligibility using binaural measurements of interaural correlation functions using an artificial head was proposed by the authors in [8,9].

To perform an a priori assessment of language intelligibility in the hall (at the design stage or in an existing room), the following algorithm of calculations is proposed.

The functions of interaural correlation at the listener's places when the stereophonic system (Fig. 1) emits video pulses of two types are analyzed: rectangular and triangular (sawtooth). The normalized interaural correlation function and the interaural correlation coefficient are used for the analysis [6, 11].

$$IACF(\tau) = \frac{\int_{-\infty}^{\infty} p_l(t)\, p_r(t+\tau)\, dt}{\sqrt{\int_{-\infty}^{\infty} p_l^2(t)\, dt \cdot \int_{-\infty}^{\infty} p_r^2(t)\, dt}} \quad (1)$$

$$IACC = \max\left|\mathrm{IACF}(\tau)\right| \quad (2)$$

Figure 2 shows a diagram illustrating the composition of the signals S1 (t) and S2 (t), which come to the left and right ear (or to the microphones M1 and M2) from the sound sources G1 and G2. So on each ear come two signals - direct and cross:

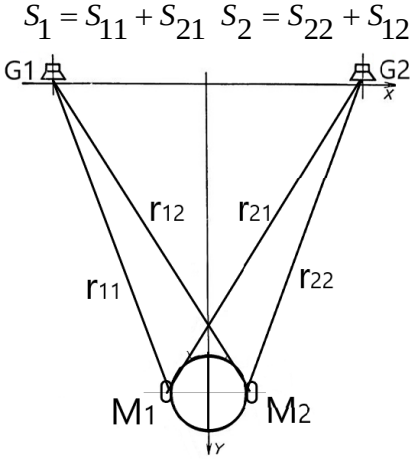


Fig. 2 Illustration to determine the functions of interaural correlation: r11, r12, r21, r22 - the distance (m) from the sources G.1 and G.2 to the corresponding ear (for the listener at a certain listening place)

The general form of the interaural correlation function is rearranged by the formulas:

$$R(\tau) = R_{11-22}(\tau) + R_{11-12}(\tau) + R_{21-22}(\tau) + R_{21-12}(\tau) \quad (4)$$

Where:

$$R_{11-22}(\tau) = \int_{-\infty}^{\infty} S_{11}(t) \cdot S_{22}(t+\tau)\, d\tau$$

$$R_{11-12}(\tau) = \int_{-\infty}^{\infty} S_{11}(t) \cdot S_{12}(t+\tau)\, d\tau$$

$$R_{21-12}(\tau) = \int_{-\infty}^{\infty} S_{21}(t) \cdot S_{22}(t+\tau)\, d\tau$$

$$R_{21-12}(\tau) = \int_{-\infty}^{\infty} S_{21}(t) \cdot S_{12}(t+\tau)\, d\tau$$

A. *Calculation of interaural correlation coefficients for a rectangular pulse*

Sound sources G.1 and G.2 emit a rectangular pulse of the form:

$$S(t) = A \cdot rect(\frac{t}{\tau}) \tag{5}$$

Or

$$S(t) = \begin{cases} 0, |t| > \frac{\tau_i}{2}; \\ \frac{A}{2}, |t| = \frac{\tau_i}{2} \\ A, |t| < \frac{\tau_i}{2} \end{cases}$$

The rectangular pulse is selected as the easiest to analyze.

The general form of the cross-correlation function of rectangular pulses is expressed by the formula:

$$R(\tau) = A_1 A_2 \tau_i \left(1 - \frac{|\tau| + \Delta t}{\tau_i} \right) \tag{6}$$

Where Δt is the time delay between signals at the receiving point? Must not be used.

Taking into account formulas (1), (2), (3) and the ratios for time delays:

$$\Delta t_{11-21} = \frac{r_{11-21}}{c_0}; \ \Delta t_{21-12} = \frac{r_{21-12}}{c_0};$$

$$\Delta t_{12-22} = \frac{r_{12-22}}{c_0}; \ \Delta t_{11-22} = \frac{r_{11-22}}{c_0};$$

We obtain the expression for the interaural correlation coefficient for a rectangular pulse:

$$R_H(0) = \frac{1}{\sqrt{N_1}} \left\{ A_{11} \cdot A_{21} \cdot \left[\left(1 - \frac{\Delta t_{11-21}^3}{\tau_i^3} \right) - \frac{3}{2} \frac{\Delta t_{11-21}^3}{\tau_i} \left(1 - \frac{\Delta t_{11-21}^2}{\tau_i^2} \right) \right] + A_{21} \cdot A_{12} \cdot \left[\left(1 - \frac{\Delta t_{21-12}^3}{\tau_i^3} \right) - \frac{3}{2} \frac{\Delta t_{21-12}^3}{\tau_i} \left(1 - \frac{\Delta t_{21-12}^2}{\tau_i^2} \right) \right] \right.$$
$$\left. + A_{12} \cdot A_{22} \cdot \left[\left(1 - \frac{\Delta t_{12-22}^3}{\tau_i^3} \right) - \frac{3}{2} \frac{\Delta t_{12-22}^3}{\tau_i} \left(1 - \frac{\Delta t_{12-22}^2}{\tau_i^2} \right) \right] + A_{11} \cdot A_{22} \cdot \left[\left(1 - \frac{\Delta t_{11-22}^3}{\tau_i^3} \right) - \frac{3}{2} \frac{\Delta t_{11-22}^3}{\tau_i} \left(1 - \frac{\Delta t_{11-22}^2}{\tau_i^2} \right) \right] \right\}$$

Where the normalization coefficient is determined by the expression:

$$N_1 = \left[A_{11}^2 + A_{21}^2 + 2A_{11}A_{11}(1 - \frac{\Delta t_{11-21}}{\tau_i}) \right] \times \left[A_{12}^2 + A_{22}^2 + 2A_{12}A_{22}(1 - \frac{\Delta t_{12-21}}{\tau_i}) \right]$$

And the amplitude of the pulses at the point of reception:

$$a_{11} = \frac{A_1}{r_{11}} \quad a_{12} = \frac{A_1}{r_{12}} \quad a_{21} = \frac{A_2}{r_{21}} \quad a_{22} = \frac{A_2}{r_{22}}$$

For calculations, the pulse length is chosen within 10ms for the reason that the speech signal localization of the imaginary source (total localization) remains possible if the time of interaural delay does not exceed 7-15 ms (level difference up to 10dB) [5].

The sawtooth pulse is selected for analysis as the closest in the signal form to the voice sound, i.e., the sound of the voice source, the radiation of which involves the vocal cords.

The envelope of the amplitude spectrum of the sequence of such pulses is described by the function $\sim 1/f^{2,}$ which allows a simulation of a speech signal with brown noise. The main frequency of the voice f_0 inversely proportional to the period of the sequence of sawtooth pulses T, c and is the fundamental frequency for the formation of discrete components of the spectrum of loud.

We present the sawtooth pulse by the expression:

$$S(t) = A \frac{t}{\tau_i} rect \frac{t - \tau_i / 2}{\tau_i} \tag{8}$$

or

$$S(t) = \begin{cases} A \frac{t}{\tau_i}, if \le t \le \tau_i; \\ \frac{A}{2} \frac{t}{\tau_i}, t = \tau_i; \\ 0, \end{cases}$$

Figure 4 presents a graphical view of the sawtooth pulse and provides an illustration for determining the function of interaural correlation of pulses with amplitude A1 and A2.

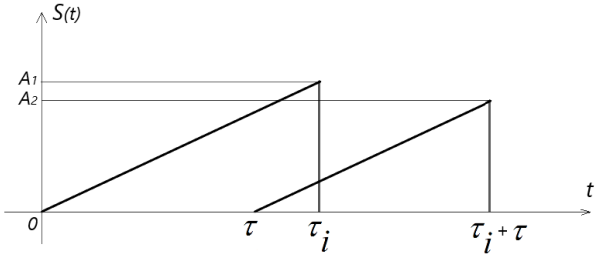


Fig. 3 Determination of the function of interaural correlation for sawtooth pulses with amplitude A1 and A2 (τi- pulse length)

The general expression for the interaural correlation function for the sawtooth pulse is described by the formula:

$$R(\tau) = A_1 A_2 \left[\frac{1}{3}\left(1 - \frac{(\tau + \Delta t)^3}{\tau_i^3} \right) - \frac{\tau + \Delta t}{2}\left(1 - \frac{(\tau + \Delta t)^2}{\tau_i^2} \right) \right] \quad (9)$$

As a result, the interaural correlation coefficient for the sawtooth pulse is calculated by the following ratio:

$$R_н(0) = \frac{1}{\sqrt{N_2}} \left\{ A_{11} \cdot A_{21} \left[\left(1 - \frac{\Delta t_{11-21}^3}{3\tau_i} \right) - \frac{3\Delta t_{11-21}}{2\tau_i} \left(1 - \frac{\Delta t_{11-21}^2}{2\tau_i^2} \right) \right] + A_{21} \cdot A_{12} \cdot \left[\left(1 - \frac{\Delta t_{21-12}^3}{3\tau_i} \right) - \frac{3\Delta t_{21-12}}{2\tau_i} \left(1 - \frac{\Delta t_{21-12}^2}{2\tau_i^2} \right) \right] \right. \quad (10)$$

$$\left. + A_{12} \cdot A_{22} \cdot \left[\left(1 - \frac{\Delta t_{12-22}^3}{3\tau_i} \right) - \frac{3\Delta t_{12-22}}{2\tau_i} \left(1 - \frac{\Delta t_{12-22}^2}{2\tau_i^2} \right) \right] + A_{11} \cdot A_{22} \cdot \left[\left(1 - \frac{\Delta t_{11-22}^3}{3\tau_i} \right) - \frac{3\Delta t_{11-22}}{2\tau_i} \left(1 - \frac{\Delta t_{11-22}^2}{2\tau_i^2} \right) \right] \right\}$$

Where the normalization factor is determined by the value:

$$N_2 = \left[A_{11}^2 + A_{21}^2 + A_{11}A_{21} \left(2 - \frac{\Delta t_{11-21}}{\tau_i} \left(3 - \frac{\Delta t^2_{11-21}}{\tau_i^2} \right) \right) \right] \times \left[A_{12}^2 + A_{22}^2 + A_{12}A_{22} \left(2 - \frac{\Delta t_{12-22}}{\tau_i} \left(3 - \frac{\Delta t^2_{12-22}}{\tau_i^2} \right) \right) \right]$$

B. Taking into account, the reverberation sound reflected from the ceiling

Preliminary calculations were performed only for direct sound coming to the listener directly from the sound source.

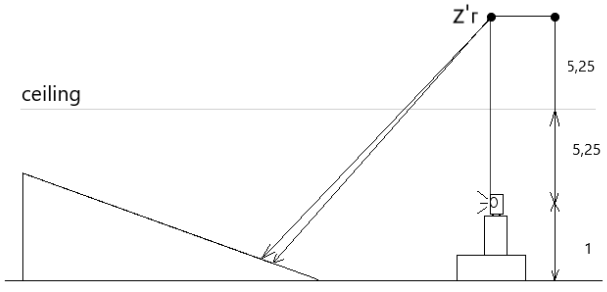
Here are the algorithms for creating a full sound field in the hall: the field of direct sound and reverberation (diffuse field) by taking into account the first reflections of sound from the ceiling.

The energy of sound waves reflected from the ceiling makes up most of the reverberation energy in the hall. And since the first reflections are the most powerful, we will limit ourselves to taking them into account to simplify analytical calculations.

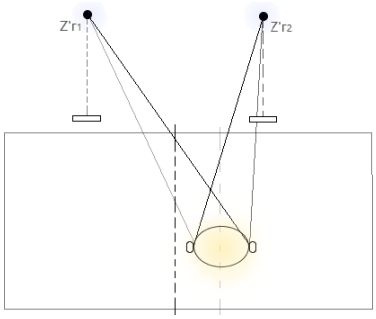
The geometry of the location of the imaginary source above the ceiling is shown in Fig.4Direct signals received by the listener are determined by the ratios (3).

Signals reflected from the ceiling, which come to the left and right ears of the listener, write as:

$$'S_1 = 'S_{11} + 'S_{21} \quad 'S_2 = 'S_{22} + 'S_{12}$$



A



B

Fig. 4 Location of imaginary sound sources above the ceiling of the room:

A – vertical diagram; B – horizontal diagram

Direct signals received by the listener are determined by the ratios (3).

Signals reflected from the ceiling, which come to the left and right ears of the listener, write as:

$$'S_1 = 'S_{11} + 'S_{21} \quad 'S_2 = 'S_{22} + 'S_{12}$$

The total signals perceived by the left and right ears of the listener will be represented by the sum:

$$S_{1\ sum} = S_1 + (1-\alpha)S_1' \quad ;$$

$$S_{2\ sum} = S_2 + (1-\alpha)S_2'$$

Where α - sound absorption coefficient of the ceiling material.

In this case, the interaural correlation function, written by analogy with formula (4), is no longer four but sixteen terms.

Next, perform the procedure described above, taking into account the formula (9) and the distances of the corresponding time delays of the signals coming to the listener.

As a result, we obtain a formula for calculating the IACC interaural correlation coefficients similar to formula (10), which due to the cumbersomeness of the article, is not given.

The calculations are performed for the pulse length τ_i=10mc and the absorption coefficient of the ceiling α=0.4.

IV. DISCUSSION OF RESULTS

The calculated values of the interaural correlation coefficients for the rectangular pulse in the listening places of the hall are given in table 1 in comparison with the measured values of IACC [6].

TABLE 1

Point number in the room	Measured IACC values		IACC values are calculated
	Pulse signal	Speech signal	Rectangular pulse
Point 1	0.88	0.49	0.89
Point 2	0.8	0.34	0.69
Point 4	0.8	0.44	0.69
Point 5	0.84	0.5	0.8
Point 6	0.78	0.35	0.7
Point 8	0.83	0.3	0.68
Point 9	0.8	0.38	0.66
Point 10	0.72	0.3	0.62
Point 12	0.78	0.27	0.68

As follows from the analysis of the above data, the calculation results are close to the measured IACC values for the pulse signal. Therefore, when modeling the speech process, the rectangular pulse will show inflated legibility values.

Table 2 shows the calculated IACC values for the sawtooth pulse compared to the measured values for the speech signal [6,8].

TABLE 2

Point number in the room	Measured IACC values	IACC values are calculated
	Speech signal	Saw-shaped impulse
Point 1	0.49	0.57
Point 2	0.34	0.41
Point 4	0.44	0.45
Point 5	0.5	0.48
Point 6	0.35	0.4
Point 8	0.3	0.38
Point 9	0.38	0.4
Point 10	0.3	0.38
Point 12	0.27	0.34

Comparison of calculated and measured data indicates their similarity. The calculated IACC values for the sawtooth pulse are significantly closer to the values measured for the speech signal.

Thus, a triangular pulse can be used to model the speech process.

Table 3 shows the calculated IACC values for direct sound and taking into account the reflection of sound from the ceiling in comparison with the measured data. A separate column shows the values of the maximum time delay of the reflected signal relative to the direct signal for the control points of the room.

TABLE 3

Point number in the room	IACC Measured values	IACC for direct sound	IACC taking into account the reflection from the ceiling	Time delay of the reflected signal relative to direct sound
Point 1	0.49	0.57	0.5	27,3
Point 2	0.34	0.41	0.37	25
Point 4	0.44	0.45	0.42	13
Point 5	0.5	0.48	0.45	16,9
Point 6	0.35	0.4	0.37	16,6
Point 8	0.3	0.38	0.39	8,4
Point 9	0.38	0.4	0.41	9,1
Point 10	0.3	0.38	0.39	8,25
Point 12	0.27	0.34	0.35	6,3

As can be seen from the above data, the time delay of the reflected sound of more than 10 ms reduces the value of IACC, i.e., impairs speech intelligibility. If the time delay is less than 10 ms, the IACC values increase as the reflections amplify the direct sound.

The obtained result is fully consistent with the known recommendations for the time delay of the first reflection, which comes after the direct sound and determines the feeling of intimacy of the room [12]. According to the recommendations [13] for language rooms, it should not exceed 10-15 ms. It turned out that this factor not only gives the impression of the size of the room but also significantly affects the intelligibility of language.

To establish the correspondence of the calculated IACC values to the articulation percentages on the language intelligibility scale, we will use the correspondence scale developed by the authors [9].

This scale is based on GOST R 50840-95, articulation tests conducted in the hall [8], and measured IASS values for the speech signal [6].

In fig. 5 presents the calculated values of IASS and the data of the subjective assessment of language intelligibility at the listening places.

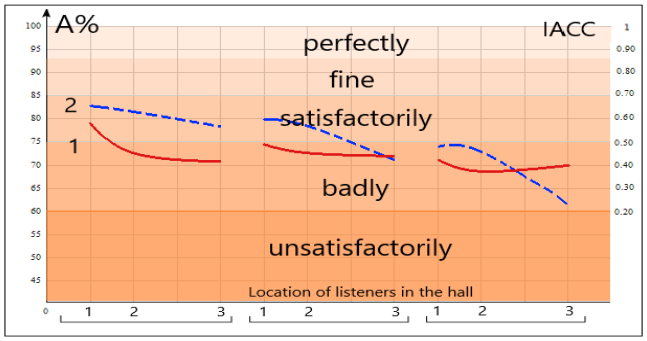


Fig. 5 Correspondence of interaural correlation coefficients of IASS speech intelligibility in the hall

Recommended font sizes are shown in Table 1.should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes.

Readability classes in Fig. 5 are marked with inscriptions ("perfectly", "good", "satisfactory", "bad", "unsatisfactory") in accordance with the recommendations of GOST R 50840-95.

As can be seen from Fig.5, the measured values of the percentage of articulation and the calculated coefficients IASS belong to one or adjacent classes of speech intelligibility.

Thus, a sawtooth pulse signal can be used to model the speech process indoors. The results of calculations of values of interaural correlation coefficients, agreed on the scale of correspondence with the percentages of articulation, allow to assess of the legibility of language in the room.

CONCLUSIONS

An objective method for assessing the intelligibility of speech in the hall based on the calculated values of the interaural correlation coefficients of the binaural pair of signals at the listening positions has been developed.

Rectangular and sawtooth video pulses with a duration of 10 ms are considered as the emitted signal. For a sawtooth pulse, in addition to the direct sound, paired reflections of sound from the ceiling are taken into account. The analysis of the reliability of the results was performed by comparing the calculated data of the interaural correlation coefficients of IASS with the results of measurements performed using an artificial head. Correspondence of the calculated values of IASS of legibility of language in percent of component of articulation A% is established on the basis of the scale of correspondence developed by authors. This makes it possible to determine the component legibility of the language in the hall by the IASS coefficients.

The results of IASS calculations obtained for a rectangular pulse approach the measured values of IASS for the pulse signal and give inflated intelligibility of speech.

The IACS values calculated for the sawtooth pulse, especially taking into account the reverberation sound, largely correspond to the IACS coefficients measured for the speech signal, which indicates the possibility of using the sawtooth pulse to model the speech process in the hall.

The advantages of this method are the relative simplicity and compliance of the calculation results with the results of data processing recorded with the help of an artificial head, and as a result - the perception of sound by a real listener at the appropriate listening location.

With the help of the proposed method, it is possible to assess the sound quality of the language in the room and make predictions about the improvement of the acoustic properties of the hall, including at the design stage.

REFERENCES

[1] W. Anert, F. Steffan. Sound amplification technique: M .: Era, (2005) 416.

[2] Pedchenko, S. Lunova. Analysis of Ukrainian Diagnostic Articulation Tables // EUREKA: Physics and Engineering, №1, (2018) 63-72.

[3] M. L. Jepsen, S. D. Ewert, T. Dau. A computational model of human auditory signal processing and perception // The Journal of the Acoustical Society of America, 124 (1) (2008) 422-438 https://doi.org/10.1121/1.2924135

[4] Lavandier and J.F. Culling. Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer // The Journal of the Acoustical Society of America, 123(4) 2237-2248, DOI: 10.1121 / 1.2871943

[5] Yu.A. Kovalgin. Stereophony. - M .: Radio and communication, (1989) 272 .

[6] M.V. Vdovenko, S.A. Luniova. Determining the area of the stereo sound of sources of speech and music signals // Microsystems, Electronics and Acoustics, .23, 58-65

[7] A. Prodeus, I. Kotvytskyi On Reliability of Log-Spectral Distortion Measure in Speech Quality Estimation // Proceedings of IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), 17-19 October 2017, Kyiv, Ukraine.; DOI https://dx.doi.org/10.1109/APUAVD.2017.8308790;

[8] N.M. Derkach, M.V. Vdovenko, Assessment of speech intelligibility by the coefficient of interaural correlation // Electronics and Acoustic Engineering, 1,.2, (2019) 50-54.

[9] N.M. Derkach, M.V. Vdovenko, S.A. Luniova. Objective Method of Speech Intelligibility With the Artificial Head // International Journal of Electronics and Communication Engineering, 7(1) (2020) 15-20.

[10] J. Blauwert - Spatial hearing. - M .: Energy, 1979 – 225.

[11] E. C.Cherry and B. Mc A. Sayers. “Human‘ Cross - Correlator ’” - A Technique for Measuring Certain Parameters of Speech Perception // The Journal of the Acoustical Society of America, 28(5) (1956) 889-895, https://doi.org/10.1121/1.1908506

[12] I. Aldoshina, R. Pritts. Musical acoustics. - C.-П .: Kompozitor, (2006) 720.

[13] V.S. Didkovsky, S.A. Луньова, O.B. Bogdanov. Architectural acoustics.-K .: NTUU "KPI", (2012) 384