SSRG International Journal of Electronics and Communication Engineering Volume 7 Issue 1, 15-19, January 2020
ISSN: 2348 – 8549 /doi:10.14445/23488549/IJECE-V7I1P103

Objective Method of Speech Intelligibility with the Artificial Head

Derkach N. [#1], Vdovenko M. [*2], Lunyova S. [#3]

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

Faculty of Electronics, the Department of Acoustics and Acoustic Electronics

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

Received Date: 01 January 2020
Revised Date: 29 January 2020
Accepted Date: 02 February 2020

Abstract

The article proposes and justifies a new objective method for determining speech intelligibility using measurements of a binaural pair of speech signals using an artificial head. According to the calculated values of the coefficients of the recorded signals interaural correlation, the class of speech intelligibility, the quality of understanding, and the assessment of the language's readability are uniquely determined.

Keywords — *unmanned aerial vehicles, speech intelligibility, interaural correlation, sound source localization.*

I. INTRODUCTION

Modern interest in the artificial intelligence application in various areas of human life requires further development of objective methods of measurement, particularly language intelligibility. Must receive a preliminary assessment of the speech intelligibility is because rooms with large dimensions or significant levels of noise such as any station, airport halls, booths and salons of airplanes, transports, dispatching stations, concert halls, and conference rooms as well as large audiences are characterized by low speech intelligibility [1], [2]. Since such premises, as a rule, information messages or artistic language have to be heard, this factor essentially influences their functionality [1], [3].

The final method for determining speech intelligibility is the subjective evaluation method, namely articulation tests [4]. But this method involves the involvement of a large number of experts and the availability of the Ukrainian language articulation tables [5].

It is easier to apply objective methods for assessing speech intelligibility [4], [6] - [9]. However, all known methods nowadays use broadband noise, which does not have the characteristic features of the language associated with articulation, pauses, accents, etc.

Since the main indicators of language quality are syllabic and phrase intelligibility [1], [4], it is difficult to explore these indicators using broadband noise since it contains neither phrases nor syllables.

Understanding the received information depends on the binaural perception quality of the language itself, not of any broadband signal [10]. For high reliability objective assessment of speech intelligibility appropriate to use speech signal [11], [12] - namely binaural auditory perception model [13] - [18].

The first attempt to evaluate the intelligibility of the language by the coefficients of intercorrelation of binaural signals was made in the article [10]. In this work, binaural pairs of speech signals were measured at the listening sites of the resonant hall with artificial head and articulatory tests.

The purpose of the article is to propose a new objective method for assessing speech intelligibility by the coefficients of the interaural correlation of the received signals binaural pair using an artificial head.

II. ANALYSIS METHOD

The method of analyzing the pitch of the detected function of the interaural correlation (IACF) speech signal on hearing signals is identified based on the most visible microphones in the piece's left and right channels.

The function of interaural correlation of the binaural pair of signals [19] - [21] is used as a function of mutual correlation [22], [23].

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

Where p_l , p_r - the values of sound pressure coming into the left and right ear; - time delay of signals.

The assessment of the similarity of the two signals is based on the correlation coefficient (IACC), which is equal to the maximum value of IACF and at the correlation interval defined by the width of the central peak of the IACF between the first zero-line crossings [19], [23].

Based on the values of the interaural correlation coefficients and the subjective assessment

of speech intelligibility in percent articulation (A%) [10], is a known similarity between the binaural pair of signals and the intelligibility of the speech at the respective listening positions.

The subjective assessment of language intelligibility is based on the classification according to GOST R 50840-95.

III. EXPERIMENTAL STUDIES

Measurements were made using an artificial head for stereo sound in the conference hall of the Faculty of Electronics of the National Technical University of Ukraine (according to ISO 3382-1: 2009). The hall's airspace is 1653 m3; size in the plan: 14,3*18,5 m2; hall height - 6.25 m. The plan of the hall is presented in Fig. 1.

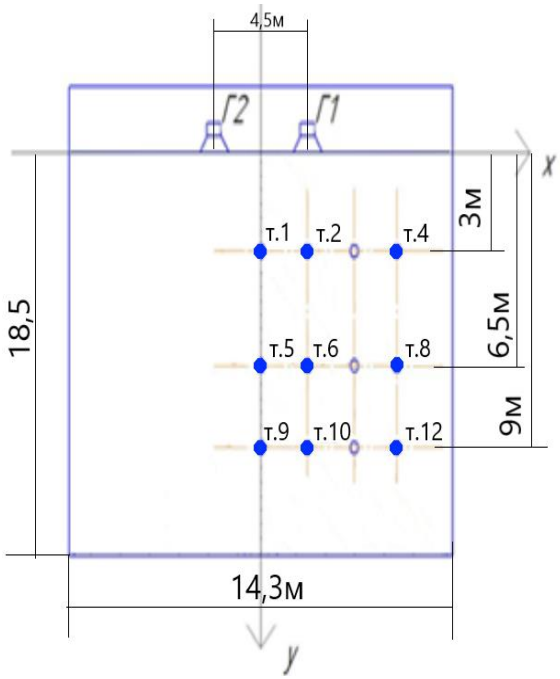

Fig. 1. Plan of location of measurement points in the hall

As the stereo pair used speakers Radiotehnika S-30B, installed at a distance of 4.5 m. Behringer ECM8000 microphones are mounted in the artificial head. The Park Audio V4-900 MkII power amplifier and M-Audio Fast Track Pro sound card are used to processing the results.

The audio signals of the artificial head's right and left channels were recorded alternately at the three points of the first, middle, and last row of the hall. The distance from the line of the loudspeaker base to the first row - 3m, to the middle - 6.5m, to the last - 9m.

The binaural characteristics of speech and pulse signals were recorded to establish the signal type recommended for measurements as a pulse signal was used a rectangular pulse of a 5 ms noise signal. The speech signal was represented by a 20 seconds passage from the novel "Master and Margarita" by M. Bulgakov.

IV. RESEARCH AND DISCUSSION OF THE RESULTS

As a result, the recorded binaural signals processing, IACF interconnection correlations and corresponding IACC values are found.

Examples of IACF charts are shown in Figure 2

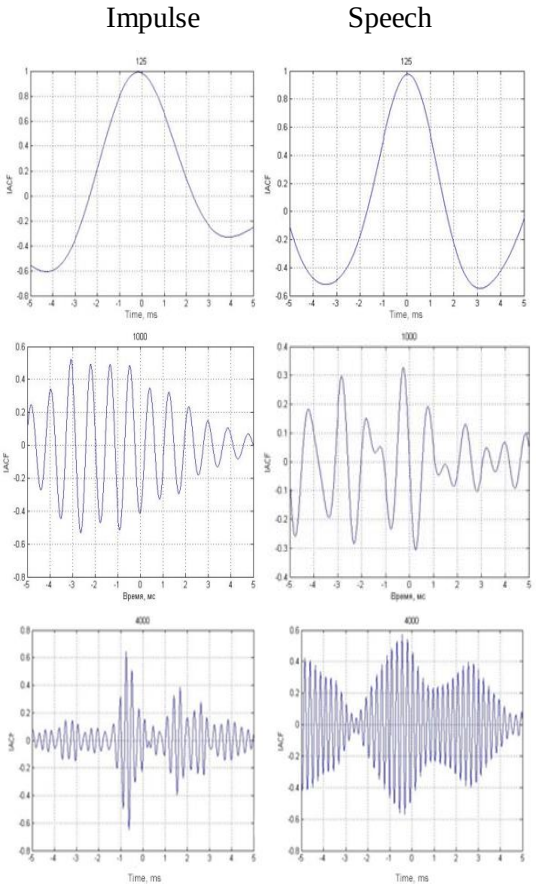

Fig. 1. Functions of interaural correlation of pulse (left) and speech (right) signals in octave frequency bands with center frequencies 125 Hz, 1000 Hz, 4000 Hz (for stereo sound)

for the central point of the first row. (item 1 in Fig. 2).

As can be seen, the IACF's shape is strongly distorted with increasing frequency, indicating a weak correlation of high-frequency components of binaural signals. This strongly affects speech intelligibility since, in the region of high frequencies, the spectra of sizzling and whistling sounds are concentrated.

The choice of the type of sound signal to assess speech intelligibility was made based on comparing the coefficients of interaural correlation of impulse and speech signals in octave bands of standard frequencies.

As a result, it was found that the IACC of the speech signal is lower than the corresponding pulse signal coefficients, indicating a worse similarity of

speech binaural signals in comparison with pulsed ones (Table 1).

Besides, according to preliminary estimates [5], it is established that the value of the composite coefficient of the speech intelligibility Kc, calculated by

TABLE I
Value of additional elements of the horn circuit

Type of signal	point 1	point 4	point 12
Impulse signal	0.88	0.8	0.78
Speech signal	0.49	0.44	0.827

measuring the index of clarity C50 (for pulse signal), does not correspond to the subjective evaluation of the speech intelligibility, do not correspond to the subjective assessment of language intelligibility (Table 1).

In assessing the intelligibility of the language in the premises, as a rule, are based on the values of the resolution index C50, which is measured by the impulse signal in the room [1]:

$$C_{50} = 10\lg\left[\int_0^{50мс} p^2(t)dt \Big/ \int_{50мс}^{\infty} p^2(t)dt\right] \quad (2)$$

With positive C50 values, satisfactory language intelligibility must be good. But such results are not always true.

According to preliminary estimates [10], it is established that the values of the component of the intelligibility coefficient of the language Kc (K_{cmax}=1), calculated according to the index of clearness C50:

$$k_C = \sqrt[5]{D} = \sqrt[5]{1/(1+10^{-C_{50}/10\partial Б})} \quad (3)$$

Do not compliance with the subjective assessment of language intelligibility (Table 2). Besides, Table 2 shows the values of the inter-correlation coefficient (IACC), which are quite high, and the percentage of the component articulation A%, which corresponds to the satisfactory and unsatisfactory values of speech intelligibility in the room.

It turns out that while the coefficient of component readability Ks (on impulse signal) the room corresponds to good speech intelligibility (86-93% of intelligibility according to GOST R 50840-95), the subjective assessment of speech intelligibility is in the range of "well-satisfactory-badly." The IACS correlation coefficient is also approaching one, indicating the identity of the impulse signals. Therefore, the impulse signal cannot be used to evaluate the intelligibility of the language.

Let us analyze the use of the language signal for these purposes. Table III shows the values of the inter-correlation coefficients of the speech signal and the corresponding percentages of the component articulation.

We can see that the stereo-pair speech signals are much weaker correlated with each other, which indicates that the room's speech is less readable than the impulse estimate. So, let's make a connection between the IACS speech signal and the subjective assessment of the percentage of composite articulation.

So, to assess speech intelligibility, you should choose a speech signal. The values of the IACC coefficients are presented in Figure 3 in the form of graphs on the listening positions of the corresponding series (for each row: p.1 is the central point of the row, p. 2 is the space opposite one of the speakers, and p.3 is the last place in the row). The same figure shows the percentage of articulation A% obtained due to articulation tests [5] for the corresponding listening sites. Also the degree of speech intelligibility is specified per the legibility class according to GOST R 50840-95.

TABLE II
SYLLABLES ASSESSMENT OF SPEECH INTELLIGIBILITY BY IMPULSE SIGNAL

Indicators of legibility	point 1	point 4	point 12
C50	4,5	3	1,1
Kc	0.93	0.92	0.89
IACC	0.88	0.8	0.78
A%	79,5	76,4	61

TABLE III
SYLLABLES ASSESSMENT OF SPEECH INTELLIGIBILITY BY SPEECH SIGNAL

Indicators of legibility	point 1	point 4	point 12
IACC	0.49	0.44	0.27
A%	79,5	76,4	61

The IACC pulse signal values shown in Figure 3 for comparison indicate that the estimation of the speech intelligibility by the pulse signal will not be correct because it does not correspond to subjective measurements by legibility class.

V. CONCLUSIONS

A new objective method for evaluating speech intelligibility is proposed based on the interaural correlation coefficients of the binaural pair of speech signals.

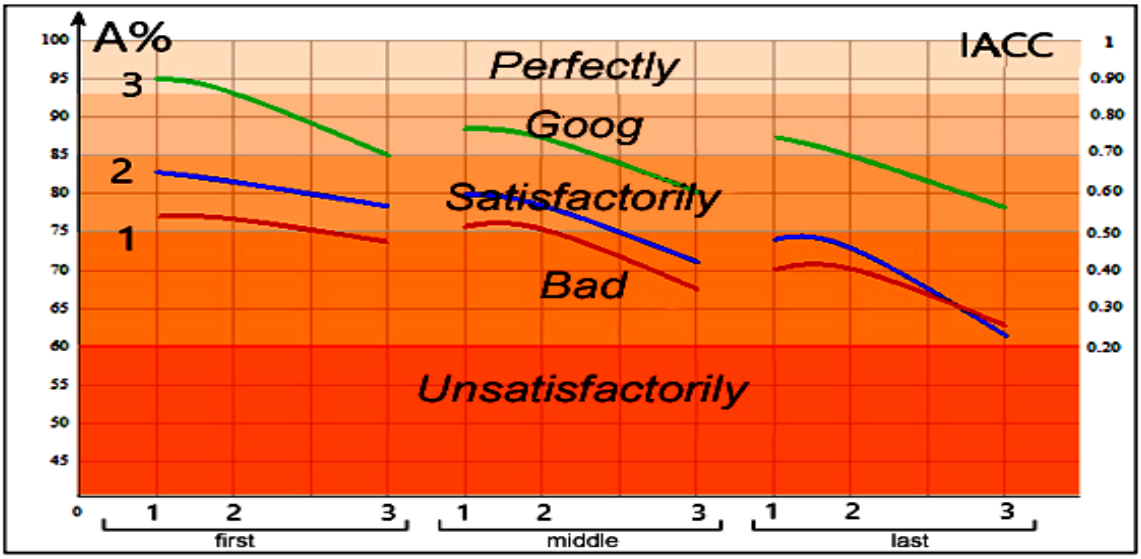

Fig. 3: Interaural correlation coefficients of the IACC (1) of the speech signal and the value of the articulation percentage A% (2), the IACC value of the pulse signal (3) on the listening positions in the room Different shades of color and lead letters indicate the legibility classes

Simultaneously, the IACC values of the speech signal and results of subjective measurements are in the same evaluative areas of speech intelligibility. This allows us to formulate an appropriate legibility scale based on the classification according to GOSTR50840-95 (Table IV).

Where ρ - medium density $\left[\frac{кг}{м^3}\right]$; γ - coefficient of propagation of the sound wave.

Created a renewed scale of speech intelligibility according to coefficients of intercorrelation (IACC) of binaural indicators of legibility pair of perceived speech signals.

Measurements are carried out with an artificial head, which does not require additional articulation tests acknowledgments in the unnumbered footnote on the first page.

Table IV
ESTIMATION OF THE SPEECH INTELLIGIBILITY

Indicators of legibility	Quality of understanding	Percentage of articulation component, A%	Interaural correlation coefficient, IACC	Class
Perfectly	Without the slightest strain of attention	>93	>0,86	Highest
Good	No complications	86-93	0,7-0,51	I
Satisfactorily	With the strain of attention, without interrogations	76-85	0,5-0,21	II
Bad	With a strain of attention, with interrogations	61-75	0,5-0,21	III
Unsatisfactorily	With a strain of attention, with interrogations and repetitions	45-60	>0,2	IV

REFERENCES

[1] S. Lunyova, V. Didkovsky, O. Bogdanov. Architectural acoustics. -Kiev.-NTUU "KPI", 2012.-p.350.

[2] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," J. Audio Eng. Soc., vol. 20, no. 2, pp. 146–159, 1972.

[3] I. Titze, Principles of voice production. Englewood Cliffs, N.J.: Prentice Hall, 1994, ISBN 978-0-13-717893-3.

[4] S. Luniova, V. Didkovs'kyy, O. Pedchenko "Akustyka movotvorennya [Acoustics of speech formation]," Lambert Academic Publishing, 2018, ISBN: 978-613-7-32891-0.

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

[6] Prodeus A., Kotvytskyi I. On Reliability of Log-Spectral Distortionalri Measure in Speech Quality Estimation// Proceeding of IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), 17-19 October 2017, Kyiv, Ukraine. DOI: https://doi.org/10.1109/AIACT.2017.8020068

[7] Prodeus A., Kotvytskyi I., Ditiashov A. Clipped Speech Signals Quality Estimation, Proceeding of 5th International Conference «Methods and Systems of Navigation and Motion Control» (MSNMS-2018), 16-18 October 2018, Kyiv, Ukraine, pp.151-155. ISBN: 978-1-5386-5869-7. DOI:10.1109/MSNMS.2018.8576308

[8] 8. I. Aldoshina, R. Prits. Musical Acoustics. - St. Petersburg: Composer, 2006.Y. A. Kovalgin, "Stereofonia [Stereophony]," Moscow, USSR: Radio i svyaz, 1989.

[9] Y. A. Kovalgin, "Stereofonia [Stereophony]," Moscow, USSR: Radio i svyaz, 1989.

[10] Derkach N.M., Lunova S.A., Vdovenko M.V., Estimation of speech intelligibility by the coefficient of inter-ear correlation of a speech signal // ELECTRONIC AND ACOUSTICAL ENGINEERING. - 2019. VOLUME 2, No. 1, pp. 50-54.

[11] G. Ye. Kedrova, V. V. Potapov, Ye. B Omel'yanova, A. M. Yegorov, and M. V. Volkova, "Russkaya fonetika: Elektronnyy uchebnik [Russian Phonetics: An Electronic Textbook]," Otraslevoy Fond algoritmov i programm M, 2002.

[12] G. Ye. Kedrova, V. V. Potapov, Ye. B Omel'yanova, A. M. Yegorov, and M. V. Volkova, "Russkaya fonetika: Elektronnyy uchebnik [Russian Phonetics: An Electronic Textbook]," Otraslevoy Fond algoritmov i programm M, 2002.

[13] T. May, S. van de Par, and A. Kohlrausch, "Binaural Localization and Detection of Speakers in Complex Acoustic Scenes," in The Technology of Binaural Listening, J. Blauert, Ed. Springer-Verlag Berlin Heidelberg, 2013, pp. 397–425.

[14] R. M. Stern and H. S. Colburn, "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," J. Acoust. Soc. Am., vol. 64, no. 1, pp. 127–140, 1978, DOI: 10.1121/1.381978.

[15] M. Lavandier and J. F. Culling, "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer," J. Acoust. Soc. Am., vol. 123, no. 4, pp. 2237–2248, 2008, DOI: 10.1121/1.2871943

[16] A. Kohlrausch, J. Braasch, D. Kolossa, and J. Blauert, "An Introduction to Binaural Processing," in The Technology of Binaural Listening, J. Blauert, Ed. Springer-Verlag Berlin Heidelberg, 2013, pp. 1–32.

[17] L. A. Jeffress, "A place theory of sound localization.," J. Comp. Physiol. Psychol., vol. 41, no. 1, pp. 35–39, 1948, DOI: 10.1037/h0061495.

[18] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech Lang. Process., vol. 14, no. 4, pp. 1462–1469, 2006, DOI: 10.1109/TSA.2005.858005.

[19] M. L. Jepsen, S. D. Ewert, and T. Dau, "A computational model of human auditory signal processing and perception," J. Acoust. Soc. Am., vol. 124, no. 1, pp. 422–438, 2008, DOI: 10.1121/1.2924135.

[20] M.Vdovenko, S.Lun`ova. Vyznachennya oblasti stereofonichnogo zvuchannya dlya dzherel movnyx ta muzychnyx sy`gnaliv// Mikrosystemy`, Elektronika ta Akusty`ka, №6, 2018.-pp. 58-65.

[21] J. Blauert and W. Cobben, "Some Consideration of Binaural Cross Correlation Analysis," Acta Acust. united with Acust., vol. 39, no. 2, pp. 96–104, 1978.

[22] J. Blauert, Prostranstvennyy slukh [Spatial hearing]. Moscow, USSR: Enrgiya, 1979.

[23] E. C. Cherry and B. M. A. Sayers, "'Human "Cross-Correlator"'—A Technique for Measuring Certain Parameters of Speech Perception," J. Acoust. Soc. Am., vol. 28, no. 5, pp. 889–895, 1956, DOI: 10.1121/1.1908506.

[24] R. Otnes and L. Enokson, Prikladnoy analiz vremennykh ryadov [Applied time series analysis]. Moscow, USSR, 1982.

[25] V.S.R Kumari , Dileep Kumar Devarakonda. "A Wavelet Based Denoising of Speech Signal". International Journal of Engineering Trends and Technology (IJETT). V5(2):107-115 Nov 2013. ISSN:2231-5381.

[26] Mr. Vijaykumar D. Shinde , Mr. C. G. Patil , Mr. Sachin D. Ruikar. "Wavelet Based Multi - Scale Principal Component Analysis for Speech Enhancement". International Journal of Engineering Trends and Technology (IJETT). V3(3):397-400 May-Jun 2012. ISSN:2231-5381.